Berlin
Universities Publishing

Harald A. Mieg
Dave Morris
(Eds.)

# The Role of Theory

With contributions by:

Andrew Abbott • André Armbruster
Jelena Brankovic • Peter Fischer
Nico Formánek • Kinga Golus
Julian Hamann • Riikka Hofmann
Moritz Klenk • Harald A. Mieg
Dave Morris • Erik J. Olsson
Maël Pégny • Leopold Ringel
Vlasta Sikimić • Marie von Heyl
Tobias Werron • Rainer E. Zimmermann

Wissenschaftsforschung Jahrbuch 2023

Harald A. Mieg | Dave Morris (Eds.)
**The Role of Theory**

The scientific series **Jahrbücher Wissenschaftsforschung** is edited by Prof. Dr. Harald A. Mieg (Humboldt-Universität zu Berlin, Geography Department).

Jahrbücher Wissenschaftsforschung Volume 2023

# The Role of Theory

Jahrbuch Wissenschaftsforschung 2023

Editors:
Harald A. Mieg
Dave Morris

Berlin
Universities Publishing

**Bibliographic information published by the Deutsche Nationalbibliothek**
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de/.

The publication of this work was supported by the Open Access Publication Fund of Humboldt-Universität zu Berlin.

**Berlin Universities Publishing, 2025**
https://berlin-universities-publishing.de/

Berlin Universities Publishing (BerlinUP) is the open access publisher from the consortium of Freie Universität Berlin, Humboldt-Universität zu Berlin, Technische Universität Berlin, and Charité – Universitätsmedizin Berlin, which together form the Berlin University Alliance (BUA).
The BerlinUP Books division publishes high-quality books across the core disciplines of Berlin's research landscape.
The publisher's name **BerlinUP** is protected by trademark law.

BerlinUP Books
Universitätsbibliothek der TU Berlin
Fasanenstr. 88, 10623 Berlin
Tel.: +49 (0)30 314 76119
Email: books@berlin-universities-publishing.de



The quality assurance of this publication was ensured by an editorial review and a single blind peer review.

Design and typesetting: Dave Morris
Copy editing: Dave Morris
Print: docupoint GmbH

ORCID iD Harald A. Mieg:
https://orcid.org/0000-0001-9272-8389

ORCID iD Dave Morris:
https://orcid.org/0009-0005-4215-060X

**ISBN 978-3-98781-034-3 (print)**
**ISBN 978-3-98781-035-0 (online)**

**ISSN 2944-6597 (print)**
**ISSN 2944-6600 (online)**

Published online on the institutional repository of the Technische Universität Berlin
DOI 10.14279/depositonce-22083
https://doi.org/10.14279/depositonce-22083

## Table of Contents

*Authors* ..... *7*

*Thanks to Reviewers* ..... *8*

HARALD A. MIEG & DAVE MORRIS
*Editorial: The role of theory today?* ..... *9*

*Acknowledgements* ..... *17*

**I The Sociological View on Theory**

ANDREW ABBOTT
*The role of theory in the social sciences* ..... *23*

JULIAN HAMANN & SINA FARZIN
*The sociological concern with theorizing, and how it could be complemented with a focus on media of theorizing* ..... *59*

TOBIAS WERRON, JELENA BRANKOVIC & LEOPOLD RINGEL
*Collaboration as a medium of theorizing* ..... *67*

PETER FISCHER
*Reflexive philosophy of science as an instrument in the social sciences* ..... *73*

MARIE VON HEYL, ANDRÉ ARMBRUSTER & MORITZ KLENK
*Theorizing through podcasts?* ..... *101*

## II The Philosophical View on Theory

ERIK J. OLSSON
*Definitions as explications and the explanatory role of knowledge.....117*

RAINER E. ZIMMERMANN
*What is, and to what end do we study, theory?.....131*

KINGA GOLUS
*Why should future philosophy teachers learn 'theory'?*
*A philosophical perspective on teaching experiences and reflections .....147*

HARALD A. MIEG
*Why theorizing should be seen as a form of research.....157*

RIIKKA HOFMANN
*From a learning sciences perspective:*
*The importance of theory for facilitating learning in universities .....165*

## III Theory and Artificial Intelligence

VLASTA SIKIMIĆ
*From data to theory and back: Why the AI era requires philosophy .....189*

NICO FORMÁNEK
*Can computer technology change physical theories?.....203*

MAËL PÉGNY
*Do LLMs contain knowledge (of anything)?.....211*

*Jahrbücher Wissenschaftsforschung.....245*

## Authors

ANDREW ABBOTT, University of Chicago, Department of Sociology

ANDRÉ ARMBRUSTER, University of Duisburg-Essen, Faculty of Social Sciences

JELENA BRANKOVIC, Bielefeld University, Faculty of Sociology

SINA FARZIN, Universität der Bundeswehr München, Faculty of Social Sciences

PETER FISCHER, Technische Universität Dresden, Institute of Sociology

NICO FORMÁNEK, Universität Stuttgart, High-Performance Computing Center

KINGA GOLUS, Bielefeld University, Department of Philosophy

JULIAN HAMANN, Humboldt-Universität zu Berlin, Faculty of Cultural, Social and Educational Sciences

MARIE VON HEYL, Universität der Künste Berlin

RIIKKA HOFMANN, University of Cambridge, Faculty of Education, Hughes Hall

MORITZ KLENK, Hochschule Mannheim, Faculty of Design

HARALD A. MIEG, Humboldt-Universität zu Berlin, Geography Department

ERIK J. OLSSON, Lund University, Department of Philosophy

MAËL PÉGNY, Sama Partners, Mannheim

LEOPOLD RINGEL, Bielefeld University, Faculty of Sociology

VLASTA SIKIMIĆ, Eindhoven University of Technology, Department of Industrial Engineering & Innovation Sciences

TOBIAS WERRON, Bielefeld University, Faculty of Sociology

RAINER E. ZIMMERMANN, Institut für Design Science, München e.V. / Clare Hall, University of Cambridge

## Thanks to Reviewers

Most of the authors have reviewed one or more of their colleagues' contributions. We would also like to thank the following external reviewers:

CLARE BROOKS, Cambridge

GUUS EELINK, Tübingen

STEPHAN GAUCH, Berlin

LUTZ HIEBER, Hannover

MARKUS KIP, Berlin

HENRIK LAGERLUND, Stockholm

HENNING LAUX, Hannover

JOHANNES LENHARD, Kaiserslautern

ANDREA RAIMONDI, Bielefeld

MARTIN REINHART, Berlin

KATHARINA SCHULZ, Göttingen

ROBERT WILLIAMSON, Tübingen

FRANCESCA VIDAL, Landau

ALEKSANDRA VUČKOVIĆ, Belgrade

DOI 10.14279/depositonce-22429

HARALD A. MIEG & DAVE MORRIS

# The Role of Theory Today?

Editorial

According to common understanding, knowledge emerges when information meets theory. But given AI's new capabilities for machine integration of mass data, do we even need theories anymore? This was the question that motivated our conference on the role of theory, which resulted in this book. The conference was co-organized by the Gesellschaft für Wissenschaftsforschung and the Robert K. Merton Center for Science Studies at the Humboldt University in Berlin, with the participation of higher education research. The joint hosting reflects the general approach of the conference and the book, namely interdisciplinary science studies. This includes philosophy, sociology, history of science, cognitive science, and, last but not least, reflection on science in higher education. Before briefly introducing the contributions from the various disciplines involved, we begin our editorial with a look at the concept of theory and will see that when we speak of theory, we should also touch on other issues such as truth and the scientific practice of peer review. Overall, the contributions show that theory is indispensable as part of what has now

Prof. Dr. Harald A. Mieg
Humboldt-Universität zu Berlin
Email: harald.mieg@hu-berlin.de

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory*. *Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

become a set of professionalized scientific practices. In this context, and on our own behalf, we will conclude with a brief discussion of the responsible role of editors.

## 1 Theories as models of the world

Understanding theory is central to the study of science. This is because theories can guide research. The word "theory" comes from the Greek and means a kind of observation. It deals with truth. This truth is about observing what is permanent and may be fundamental to natural and social phenomena. For the Greeks, this kind of observation was certainly—but not always—associated with a sense of happiness. In Greek, the word can also evoke the meaning of "theos"—a deity or god—and theory had a sense of the exclusive, as in observing the participants in a festival of the gods (see also Zimmermann, this volume), which would reveal what truth is all about. Accordingly, astronomy would be the exemplary science, since it necessarily involves observation. In this respect, astronomy, in the form of systematic observation of the stars, already played an important role in ancient Egypt and among the Babylonians, and 2000 years later even tempted a great astronomer like Johannes Kepler to include astrology in his theoretical considerations, since he believed that the stars could not but have a significant influence on human beings.

The development of science in the 20th century has shed light on the role of theory in science. Philosophers like Karl Popper and Thomas S. Kuhn were instrumental in this process. That this self-reflection itself became a science—science studies—is due to people like Derek John de Solla Price, who, among other things, founded scientometrics, the quantitative study of scientific literature. As to the role of theory, we can say: a theory is a *model of the world*—like its cognitive representation, i.e. mental models (Johnson-Laird, 1983; Thagard, 2012). A scientific theory

should be sufficiently abstract and empirically testable, and it should extend our understanding of the world. Often referred to is Robert Merton's (1968) concept of a middle-range theory that generates testable hypotheses and can be integrated with other theories. However, it is still undecided whether we should understand theories as sets of propositions, or whether there are other essential components of a theory, such as "intended applications" or components of theory which are "theoretical," i.e., theory-dependent. This was the structuralist view of theories (e.g., Sneed, 1976; Stegmüller, 1976). Therefore, testing or evaluating theories tends to be theory-dependent as well, since we may have to decide which terms to use, which may have their specific theoretical background.

## 2 Truth and peer review in professionalized science

Understanding theory as a model evokes the question, "Model of what?" and, ultimately, the question of truth. Science is unthinkable without truth. *Truth* is a concept that has been shaken up quite a bit in the last hundred years, both inside and outside of science (we speak of "science" as a unit in the sense of science as a profession). First, Tarski and Gödel proved that truth is not axiomatically derivable, as had been hoped. Now we seem to live in a "post-truth" age, where truth has become relative. This goes along with the fact that in the last 50 years science has finally become professionalized, a learnable profession, and has expanded enormously, especially with the expansion of universities (Mieg et al., 2021; Mieg, 2022). So if science does not value truth, what value can it refer to at all? To speak only of justified and peer-reviewed knowledge would be worthless without the value added by truth. The problem, however, is that the discourse of truth is too socially important to be monopolized by science. Truth is argued in court; religions and political parties also appeal to truths. Even post-truth discourses, for example around vaccination, are

not simply misinformation, but "an alternative epistemology that does not conform to conventional standards of evidence reporting" (Lewandowsky et al., 2017, p. 356). In politics and religion, we often see that *values serve as truths*. The big difference is that only in science does *truth serve as a value.*

As a profession, science must address quality assurance standards as part of the boundary work of science (Gieryn, 1983). Peer review is at the core of such standards, and yet, as a procedure used faute de mieux (for lack of a better alternative), it is itself subject to constant and fierce criticism (Tennant et al., 2017; Reinhart & Schendzielorz, 2024), theory dependence being only one issue, reflecting the schools of thought and associated networks of scientists who review each other. Overall, the system of peer review is time-consuming and therefore costly, estimated at 100 million hours worldwide in 2020 (Aczel et al., 2021) and $6 billion per year (LeBlanc et al., 2023). No wonder that we see a new phenomenon that came with the Internet and is now fueled by AI: predatory journals—or paper mills—that promise open access (of high value in science) and rapid review and make their profits from open access fees. They organize the review process through secretariats, often linked to fake (inactive) editors, and we can expect some of these journals to produce AI-based reviews. This brings us back to the role of AI in knowledge production, our starting question for our conference and this book.

## 3 This book: Contents

This current volume is part of a long-established series, The Study of Science Yearbooks (*Jahrbücher Wissenschaftsforschung*), and follows the guiding principle of all prior editions, the continued discussion. In addition to the scientific paper and discussion formats, this volume contains a new format: the *argument*. An argument presents a thesis in a

concise form and is intended to stimulate discussion. The argument is closed; the discussion takes place outside it.

Our book has three parts. The first is devoted to the sociological view of theory. One focus is on theorizing as a process. The second part is devoted to the philosophical view of theory and, in addition, is intended to focus on the role of theory in higher education. The third part takes up the initial question mentioned at the beginning: Can AI replace theory?

## 4 Discussion round 1: The sociological view on theory: From books as theories to theorizing through podcasts

This section includes five contributions, ranging from a plea for a grand social theory to a discussion of how podcasts could contribute to theory development.

1.1 "The role of theory in the social sciences" (paper). In his opening contribution, Andrew Abbott defends theory as an intellectual synthesis, defining "a book of social theory" as a "loosely deductive structure."

1.2 "The sociological concern with theorizing, and how it could be complemented with a focus on media of theorizing" (argument). This argument, by Julian Hamann and Sina Farzin, serves as an introduction to the discussion of theorizing as process.

1.3 "Collaboration as a medium of theorizing" (argument). Tobias Werron, Jelena Brankovic, and Leopold Ringel argue that theory can emerge in the cooperative practice of sociologists.

1.4. "Reflexive philosophy of science as an instrument in the social sciences" (paper). In his paper, Peter Fischer argues for reflection as a means of conducting science research in sociology. Among other things, the reviewers of this paper discuss how reflection is to be understood.

1.5 "Theorizing through podcasts?" (Discussion). Marie von Heyl, André Armbruster, and Moritz Klenk, all podcasters, break new sociological ground in their discussion of theory and podcasts. The changing relationship between producer/podcaster and audience is a specific point of discussion.

## 5 Discussion round 2: The philosophical view on theory: From epistemology to learning sciences

This section includes five contributions that range from the epistemology of definitions to theory and empirical research in the learning sciences.

2.1 "Definitions as explications and the explanatory role of knowledge" (paper). In his paper, Erik Olsson addresses the fundamental question of defining terms in the sense of Carnap's concept of explication: "transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second" (Carnap, 1962, p. 3).

2.2 "What is, and to what end do we study, theory?" (paper). In his paper, Rainer Zimmermann explains the origins of the philosophical concept of theory and its relation to attending a divine celebration. Zimmermann argues for a metaphysics of theory and calls for a psychohistory (similar to Boudieu's "cultural unconscious"?).

2.3 "Why should future philosophy teachers learn 'theory'? A philosophical perspective on teaching experiences and reflections" (paper). In her paper, Kinga Golus argues that theory, combined with inquiry-based learning, contributes to the intellectual independence that should characterize an education in philosophy.

2.4. "Why theorizing should be seen as a form of research" (argument). Harald Mieg argues that theory is a form of research that occurs in all disciplines, alongside experimentation and simulation. The

process of learning theory is often challenging, such that a theoretical basis represents the defining element of university study.

2.5 "From a learning sciences perspective: The importance of theory for facilitating learning in the university" (paper). Riikka Hofmann reflects the role of theory in three aspects: helping us start from where learners are; in changing learning cultures and conversations; and in enabling us to 'see' alternative futures for our students' learning.

## 6 Discussion round 3: Theory and artificial intelligence: Can we dispense with theory? (A discussion, to be continued...)

This section includes three contributions that take a critical view of the contribution of computers to theory-building.

3.1. "From data to theory and back: Why the AI era requires philosophy" (paper). In her paper, Vlasta Sikimić emphasizes the need for an empirically informed philosophy that provides normative guidance and ensures ethical and epistemic rigor in scientific advances based on AI. We cannot expect machine learning to lead to epistemic and moral progress, because its recommendations tend to perpetuate the values currently prevalent in society.

3.2 "Can computer technology change physical theories?" (argument). Nico Formánek argues that even in computational chemistry, an area where it is most likely that computers will be able to develop theories, this is not yet the case. For quantum chromodynamics, computer technology has brought about a lasting change within that theory, but not at the level of the defining module, that might be considered closest in spirit to laws of nature.

3.3. "Do LLMs contain knowledge (of anything)?" (paper). In his paper, Maël Pégny explores whether Large Language Models (LLMs), a subset of Machine Learning (ML), can be considered as processing theoretical knowledge, and concludes that the recent evolution of ML shows a clear trend towards task agnosticism but not towards robustness. Thus, the future of ML is uncertain. The accompanying reviews show that further discussion is needed.

## 7 Conclusion: The role of theories—the role of editors

At least three conclusions can be drawn:

1. Round 1 (starting with sociology): We can say that theory construction—or rather, theorizing—is still a central task of science as a profession.
2. Round 2 (starting with philosophy): The papers on the explicationist view of definitions (Olsson) and on the teaching of science (Hofmann) show that theory as used at the university requires both philosophy and empirical learning research.
3. Round 3 (theory and AI?): Can we do without theory? The contributed papers conclude: definitely not at present. But this is a discussion that needs to continue.

To add another conclusion concerning the challenges and opportunities presented by AI: We would like to emphasize the essential role and responsibility of scientific journal editors in upholding the principles and quality of peer review. As part of the professionalization of science and its responsibility (Mieg, 2022, 2024): Like the writing of scientific papers, the role of the editor could now become a distinct additional task in the role of a professional scientist, taught in connection with training in scientific work and endowed with a corresponding reputation. The editor also knows which discourse community a journal

serves and to what extent a submitted article is likely to contribute to knowledge that is relevant to the community served. This role in maintaining scientific integrity faces challenges, including fraudulent "paper mills" that generate plausible (but fake) academic articles/reviews and sell authorship—increasingly facilitated by AI. However, the same technology can be used to streamline genuine review processes. In this respect, we should not fear AI. AI could perform the essential but most tedious tasks (those least likely to be vigorously pursued by busy academics in their role as reviewers), such as validating statistical data sets, verifying the claimed content of cited articles, and confirming that reference sources even exist. The next GeWiF conference will look in greater depth at issues concerning the use of AI. It will be our responsibility as members of the scientific community to raise the topic of theory again—in the sense of a continued discussion.

Harald A. Mieg & Dave Morris

Berlin, November 2024

## *Acknowledgements*

First, our thanks go to Hubert Laitko, a co-founder and board member of the Gesellschaft für Wissenschaftsforschung. He came up with the idea of trying a reflective new start with a conference on a fundamental topic—e.g., theory—in the Gesellschaft für Wissenschaftsforschung. Laitko sadly passed away in 2024 at the age of 89.

We would also like to thank Martin Reinhart and Julian Hamann, who involved the Robert K. Merton Center for Science Studies as well as

higher education studies at Humboldt-Universität zu Berlin in the conference and this Yearbook on Theory.

Special thanks go to Sheena Bartscherer and Stephan Gauch, without whose commitment, including to content and organization, the conference would not have been possible.

Last but not least, we would like to thank the numerous reviewers who helped us to complete this book and continue the discussion.

## *References*

Aczel, B., Szaszi, B., & Holcombe, A. O. (2021). A billion-dollar donation: Estimating the cost of researchers' time spent on peer review. *Research Integrity and Peer Review*, *6*(14). https://doi.org/10.1186/s41073-021-00118-2

Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). The University of Chicago Press.

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, *48*(6), 781–795.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge University Press.

LeBlanc, A. G., Barnes, J. D., Saunders, T. J., Tremblay, M. S., & Chaput, J. P. (2023). Scientific sinkhole: Estimating the cost of peer review based on survey data with snowball sampling. *Research Integrity and Peer Review*, *8*(3). https://doi.org/10.1186/s41073-023-00128-2

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008

Merton, R. K. (1968). *Social theory and social structure* (enlarged ed.). Free Press.

Mieg, H. A., Schnell, C., & Zimmermann, R. E. (Eds.). (2021). *Wissenschaft als Beruf: Wissenschaftsforschung Jahrbuch 2020* (GeWiF, Gesellschaft für

Wissenschaftsforschung). Vereinigung Deutscher Wissenschaftler. Open access: Humboldt-Universität zu Berlin. https://doi.org/10.18452/23213

Mieg, H. A. (2022). Science as a profession: And its responsibility. In H. A. Mieg (Ed.), *The responsibility of science* (pp. 67–90). Springer.

Mieg, H. A. (2024). Translating values into quality: How we can use Max Weber's ethic of responsibility to rethink professional ethics. *Societies*, *14*, 183. https://doi.org/10.3390/soc14090183

Reinhart, M., & Schendzielorz, C. (2024). Peer-review procedures as practice, decision, and governance—The road to theories of peer review. *Science and Public Policy*, *51*(3), 543–552. https://doi.org/10.1093/scipol/scad089

Sneed, J. D. (1976). Philosophical problems in the empirical science of science: A formal approach. *Erkenntnis*, 115–146.

Stegmüller, W. (1976). *The structure and dynamics of theories* (translated by W. Wohlhueter). Springer.

Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., ... & Colomb, J. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research*, *6,* 1151. https://doi.org/10.12688/f1000research.12037.3

Thagard, P. (2012). *The cognitive science of science*. MIT Press.

# 1. The Sociological View on Theory

DOI 10.14279/depositonce-22430

ANDREW ABBOTT

# The Role of Theory in the Social Sciences

## Abstract

In this article, Andrew Abbott makes a case for deductive social science theory—"a book of social theory is a loosely deductive structure"—and explains why it has become so difficult to develop such a theory today. Abbott mentions historical reasons, e.g., that the greats of the social sciences (Weber, Durkheim, Adam Smith, Marx...) still had the first mover advantage as pioneers of the subject, whereas today knowledge has become so broad that a defensible synthesis (i.e., theory) is increasingly difficult. Abbott also presents an ontological argument, namely that the subject of social science (people, social groups) is fundamentally different from those of the natural sciences (people not only have cognition, but also will). The adoption of natural science methods in the social sciences leads to an alienation of empirical research from theory, which is often normatively driven. Abbott argues for a clear definition of the "knowledge ideals" (for the social sciences), especially in view of the necessary distinction from AI.

Prof. Dr. Andrew Abbott
The University of Chicago, Department of Sociology
Email: aabbott@uchicago.edu

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

There are many things that we call "theory" in the social sciences, but only one of them is social theory in the formal sense. Properly speaking, a book of social theory is a loosely deductive structure, resting on a set of simple and plausible assumptions or axioms, and deriving further concepts and results from those assumptions or axioms in an order that is logical and consistent. Within such an argument, any temporary assumptions are marked by clear annotations telling the reader where these assumptions will later be justified, and how those later justifications will rule out the possibility of having in the meantime simply assumed what one has set out to demonstrate.

Such a book is self-contained, taking no arguments from elsewhere, either from "the classical tradition," or from "authorities," or from "disciplinary consensus," or from a political position. All of these are outside arguments whose premises are not open to the reader's inspection, and might well be inconsistent with the premises of the book itself. Such external resources are therefore not employed.

Such a book involves data only for illustrative purposes. Otherwise, it risks borrowing its authority either from the assumptions made in generating that data or from the authority conferred by success with some particular explanatory problem. In either of those cases it moves towards induction, which is a laudable intellectual exercise in itself, to be sure, but one that is not ultimately capable of sustaining the cohesive argument here defined as the essence of social theory.

To be sure, social theory need not be *rigidly* deductive, like a geometry proof. Social life is not amenable to such theory, and the use of excessively precise logic—for example, the making of inferences by contradiction or by excluding alternatives with one or two counterexamples—is very dangerous for social thought, as we see occasionally in the work of

Emile Durkheim. But Durkheim's instinct was nonetheless a good one. Social theory does require a fundamental logical rigor.

To summarize, a work of social theory must:

1. Propose an abstract, systematic, and loosely deductive argument about social life.
2. Found that argument on clear and definable premises.
3. Unfold that argument with as few arbitrary assumptions as possible.
4. Avoid any reliance on external authorities.

I shall call such work "deductive theory," it being understood that the concept of deduction here denotes not mindless rigor, but systematic, cohesive, and persuasive argument from first principles. A conspicuous example of such a book is John Rawls' *A Theory of Justice* (1971).

The paper has two sections. The first will ask how this deductive conception of theory relates to the other things we call theory in the social sciences. That analysis will produce three stylized facts: first, that deductive theory is now rare in the social sciences; second, that such theory is often normative; and third, that it was formerly more connected to empirical data than it is today. The paper's second section will examine these three topics in some depth. It leads into a brief closing section on threats to our ideals for social theory.

This is a speculative essay, not a work of scholarship. It proposes interpretations and possibilities. It does not arbitrate debates. There will therefore be many loose ends.

## *Things called theory in the social sciences*

The first task is to discuss how the definition of theory as "loosely deductive argument" relates to the other things we call theory in the social sciences. There are many such things.

First, there is "theory" that consists of writing about the works of other people already identified as theorists. No one expects candidates for "theory jobs" to produce new theory in the deductive sense. Rather, such candidates are expected to teach courses about a more or less canonical list. In sociology, this would be Marx, Weber, Durkheim, Bourdieu, Habermas, and so on. In addition to teaching such courses, such "theorists" are expected to write commentary on canonical writers; to discuss historical schools of theory; and to pursue archival research upon both writers and schools. We might call this branch of scholarship "theory as commentary." Theory as commentary can be helpful, in particular by creating interpretational conventions that allow these canonical theorists to be evaluated within a common universe of discourse. But commentary does not in itself advance the cause of deductive theory.

Second, there is an applied version of theory, one that we find in empirical articles. According to this concept, "theory" consists of a large catalogue of alternative rationales that can explain empirical regularities or irregularities. This is the "mechanisms" approach to theory. Such "theories" are occasionally linked to canonical authors, but they are more often phrased in a restricted vocabulary of variables that are thought to have causal effects, variables that are usually organized into groupings that are largely conventional within particular research subfields. This world of "theory as causal mechanisms" is largely ad hoc: a given variable may be independent in one literature, dependent in another, and mediating in a third. There is thus no hope of consistent argument across studies and

subfields, and in consequence there is little convergence beyond subfields, unless it takes the form of very general, almost trivial results. As these facts about "theory as causal mechanisms" imply, such work has in practice little relevance to deductive theory.

A third concept of theory is inductive. Here the word "theory" refers to the formalization of what first appear as common properties of cases or as common patterns of relations between cases. Often these formalizations are later presented in a deductive format. For example, Ibn Khaldun's celebrated theory of dynastic cycles formalizes generalizations that he made by studying dozens of cases for his *Universal History*. Immanuel Wallerstein's World-System Theory similarly emerged from an inductive classification of countries into types. As these two examples show, inductive theory can be very powerful. But it has the problem that different inductions can have different foundations. The conclusions of an induction depend greatly on the universe of cases included as relevant and on the logical or conceptual arrangements of the induction itself. Different sets of cases and different ontological premises can easily produce incompatible inductive theories. These differences in data and presuppositions often prevent the transformation of such theories into deductive form.

There are thus three general meanings for "theory" beside the deductive one with which we began: theory as commentary, theory as mechanisms, and theory as induction. It must also be noted that the word "theory" has specific meanings in the various disciplines.

Thus, in economics, there is a theory subdiscipline. But the general acceptance of the Samuelsonian consensus has tended to restrict economic theory to the elaboration of details that follow from the premises that underlie that consensus. These premises were first set forth by Carl

Menger, then encased in marginalism by the Marshallians, then refounded on mathematical bases by Irving Fisher, and ultimately synthesized tidily by Samuelson himself. There is relatively little economic theory outside this Samuelsonian world. To be sure, there has been theoretical debate over whether economics should be predictive or explanatory, a debate that pitted Friedman against Samuelson. There have also been normative arguments about the relation of economic theory to social justice, made by writers like Anthony Atkinson and Amartya Sen. But neither of these bodies of work is as central in economic theory as is elaborating the details of the main consensus.

In political science, there is a designated subfield called "political theory," which combines the first concept of theory above—theory as commentary—with a quite active area of deductive and indeed *normatively* deductive theorizing. Writers like John Rawls, David Miller, Onora O'Neill, and Axel Honneth are examples of formal theorizing in a more or less deductive tradition, often based on prior canonical writers but often taking new and radical positions. Political science is thus unusual among the social sciences both in that it has a large subfield dedicated to deductive theory and in that this subfield is mainly dedicated to matters of justice and normative reasoning.

Sociology has no such deductive or normative subfield, nor does it have a subdiscipline like economic theory. Most theoretical writing in sociology is commentary on the canon, which has reduced over the years to Marx, Durkheim, and Weber, perhaps in recent years augmented by Goffman, Bourdieu, and various other writers. But in the daily usage of sociologists, the word "theory" most often refers to general frameworks or paradigms that are sometimes labeled by the names of the canonical writers who originally set them forth (e.g., "Bourdieusian theory"), but that are in practice much looser than deductive frameworks. Examples of

this looseness are easy to find: the number of things that have passed as "Weberian" in sociology is very large. They include idealism (as opposed to Marxian materialism); subjective action theory (as opposed to simple rational actor theories); comparative historical methodology (as opposed to contemporary causal analysis); and focus on rationalization (yet another opposite to Marx's dialectical materialism). Marx himself has similar avatars. And although Durkheim is often understood as a deductive theorist (as indeed he claimed himself to be), he is a different thinker in the eyes of different people: the sociologists cite the Durkheim of social facts and quantitative analysis while the anthropologists cite the Durkheim of aboriginal Australia and ambiguous meanings. In summary, despite loose references to canonical writers, the current sociological vernacular in the United States uses the word "theory" to denote general paradigms of research, complete with ontologies, families of specific theories, favorite types of methods and data, and conventional assumptions about various aspects of human behavior.[1]

Thus, when we step back and consider all these varying uses of the word theory—theory as commentary, theory as causal mechanisms, theory as induction, theory as differentially embodied in the different disciplines—it is evident that there is in the social sciences not much work that is theory in the sense of loosely deductive and closely wrought argument. This absence is particularly evident if we reflect about the writing that emerges from the three disciplinary theory communities just noted. Economics has the closest to a deductive system, but this was

1 In this sense, "theory" means paradigm in the sense of Thomas Kuhn. It is not "gravitational theory" as in a textbook layout of Newtonian mechanics, nor is it a simple set of equations. It is a general way of thinking that has usually emerged via a couple of major works. But it still has some kind of general order to it, if not a deductive order.

achieved largely by Fisher's borrowing of the mathematical thermodynamics of his dissertation advisor J. W. Gibbs. Moreover, consensus in economic theory has been achieved by disregarding the sources of preferences, which even economic theory itself assumes to be the ultimate determinants of human behavior. Deductive theory in economics is thus very clear but oddly insubstantial: it leaves its main causal variable untheorized.

Political science by contrast does have a number of active and quite substantial strands of "political theory." These resemble what sociologists call "paradigms," but differ from those paradigms in that they are organized not around ontologies, as are most sociological paradigms, but rather around moral or normative positions: classical liberalism, republican theory, welfare state theory, virtue ethics, and so on. Work within these various traditions is commonly deductive, and as a consequence, political theory (in the United States, at least) has the most active deductive theory community in the social sciences.

In sociology, deductive theory is rare. James Coleman's magisterial book (1990) was an exception, to be sure. But there are few others. Pierre Bourdieu's theoretical arguments are often limited by the relative undertheorization of the two central concepts of his system—domination and power. Anthony Giddens is mostly a commentator on the work of others, and he more often named problems than solved them (e.g., "structuration"). Even Parsons's famous *Structure of Social Action* was motivated as a commentary on classic writers, and the German theoretical work that has been most influential in English has been the Frankfurt School's somewhat commentative repurposing of Marx, Freud, and other classic writers to the analysis of the social problems of modernity.

To be sure, in Germany there *has* been an enduring theoretical tradition in social thought per se. But it reaches well beyond sociology and even social science. Writers like Habermas and Luhmann, Joas, and Honneth, as well as the legal theorists like the Austrian Hans Kelsen, have stayed at a fully theoretical level. These thinkers locate themselves in a longstanding tradition of social thought that originated less in social science than in German philosophy and jurisprudence. This literature's genre of writing—like that of political theory in the US and Britain—relies on a vocabulary of stylized abstractions: words like justice, value, system, communication, and action. Such a vocabulary is characteristic of the discipline of philosophy, a fact that suggests a sharp differentiation of this body of theory from the genres characteristic of the "scientific" sectors of sociology, political science, and economics. This separation is underscored by the fact that this body of German writing has extensive normative content.

Thus, when we *do* consider the few existing bodies of seemingly deductive theory in the social sciences, we find that some are insubstantial and some are not really as deductive as they seemed. Meanwhile, the strongly deductive theories that *do* exist are closely allied with philosophy and have largely normative content. In summary, looking across all types of theory in today's social sciences, we see that there is surprisingly little theory that is abstract, loosely deductive, non-normative argument about social life. Deductive theory is rare, and where it exists, it seems to some extent an isolated, independent body of work, most often connected with normative issues.

One might argue that this rarity should not be surprising. Perhaps the world doesn't need much deductive or systematic theory: "a little bit of theory goes a long way," as the American proverb puts it. But nonetheless, systematic and deductive theory does seem less common than

might be expected. For example, sociology has often developed fairly elaborate—even deductive—theories and theoretical traditions within subdisciplines. But there is little coherence across the subdisciplinary lines. Indeed, many such "sub-theories" are incompatible, and consequently cannot be brought under a common deductive structure. To be sure, one could argue that such micro-differences are beneficial, in that they enable more and more scholars to fit into a given area of study. And this support of scholarly density has been useful as academia has rapidly expanded, because it allows young people to have careers, think they are revolutionary, and so on. But these are short-term and merely instrumental benefits. They do nothing to mitigate the irreconcilability of "sub-theories" within "sub-disciplines." In effect, by allowing short-term, careerist interests to dominate sociology's intellectual life, they undercut the possibility of deductive theory.

So it is fair to think that systematic theory is indeed more rare than it ought to be, and that it is important to the future of social thought to understand why this should be true. But in the process of identifying this rarity, we have encountered two other important areas that are obviously related to deductive theory and that themselves play a role in creating this rarity.

The first of these related areas is normative writing. The exceptions to the overall rule of rarity are most often writers who are strongly normative in their point of view. This is definitionally true in so-called political theory, and it is also characteristic of the multi-disciplinary German tradition in social theory. So one wonders if the rarity of deductive theory might not be connected to some aspect of normative writing. Perhaps normative writing is inherently more difficult. Perhaps its controversial nature frightens potential theorists. Perhaps it became a casualty of the effort to create "value-free social science" in the US in the

middle of the twentieth century. Whatever the actual narratives and mechanisms, the rarity of deductive theory may somehow be related to issues about normative argument.

The second related area is empirical work. Pierre Bourdieu certainly thought of himself as a deductive theorist. And he certainly took strong normative positions. But he also stands out among contemporary theorists because he was a serious empiricist. To be sure, he had a large team of people working under him, as well as a vast army of students. Yet still he himself had much direct contact with data. A similar case is James Coleman. Indeed Coleman's theory could be read as a straightforward rationalization of the methodological preferences evident in his empirical work, borrowing rational choice from economics and deploying its ideas to solve the sociologists' problem of reconciling the individual and social levels.

But Bourdieu and Coleman were quite exceptional. American *political* theorists are not deeply embedded in empirical work. They use it only occasionally for illustrations. Similarly, none of the German social theorists mentioned earlier qualifies as a person anchored in empirical work as were Bourdieu and Coleman. To be sure, Luhmann applied his theory to many subjects, and Habermas began his career with some detailed historical research. But neither man resembled Bourdieu with his crowd of empirical acolytes nor Coleman with his long career of empirical policy analysis. Indeed, if the modern German social theorists reached out beyond the theoretical social sciences, they reached not to empirical data, but (as we shall see below) rather to jurisprudence, history, and philosophy.

This seeming disconnection between deductive theory and extensive empirical work contrasts very strongly with the situation in the past. The

founders of modern social thought often theorized with data. Marx was not himself much of an empirical researcher, but he consumed endless empirical material. Herbert Spencer personally sponsored massive comparative data collection on human societies. Weber's work was invariably based on primary historical materials. Durkheim was an avid collector and user of data. Indeed, the same connection between theory on the one hand and the world of empirics and practice on the other appears in many social science fields in the early twentieth century. Keynes was a Treasury official, Malinowski a field worker, Dewey a school reformer.

## *Sources of disconnection: Internal, normative, and empirical*

So far, the analysis has argued that deductive theory is surprisingly rare, and in the process has found deductive theory to have complex relations with both normative concern and empirical work. Indeed, the three topics—deductive theory, normative concern, and empirical work—can be seen as aspects of a single historical problem: In a brief period a little over a century ago, many of the canonical writers of modern social thought merged deductive theory, normative concern, and empirical data analysis into the work that founded our disciplines. Today this merger is gone. Moreover, deductive theory itself has become rare, and with a few exceptions, most of those rare examples pursue abstract normative argument and follow the "stylized facts" approach of philosophy. Modern deductive work largely avoids both empiricism as a project and empirical analysis as a method. Why has this dramatic change occurred?

There are three possible sets of answers. The first lie in the internal structure of theory itself— within social theory and its institutions. The second lie in the puzzling connection between normativism and deduction, and in the equally puzzling attempts of empirical social

scientists to avoid normative issues via "scientization." The third lie in the nature and methods of empirical work itself and the general absence of deductive thinking outside economics. In this second section, the paper sketches issues involved in these various explanations. It cannot produce a final explanatory account. But it can at least sketch the possible arguments that such an account might evaluate, even if this means cataloging forces with different time scales and different provenances.

*A. Internal*

To begin with internal matters, there are some general trends that might make deductive theory more rare now than heretofore. First, there is the simple penalty of success. As scholarship ballooned, the sheer growth of what was known made synthesis more difficult. Expansion required differentiation both of empirical worlds and of types of knowing. Such specialization and narrowing inevitably made deductive theory more difficult.

Other internal factors were institutional. The academicization of knowledge and the subsequent growth of universities had a large effect. Many of the founders of the social sciences in the period between 1875 and 1920 were not academics. That the social sciences moved into academia meant that they shared its development, for good or ill. Those developments included a vast increase in size of faculties, which furthered the already-mentioned differentiation in subject matters, particularly in the American universities with their egalitarian departments. Academicization also had the effect of expanding the apparatus of scholarship: footnotes, references, acknowledgements, and the like, an expansion that in many ways emphasized scholarly rectitude more than

independent thought. Both aspects of academicization raised the bar for deductive theory.

Alongside these broad forces came structural changes in the roles of knowledge. In the old-style German model, the *Ordinarius* headed a large institute of *Ausserordinarien*, graduate students, and others, all of whom could feed the professor's idiosyncratic personal enterprise. Durkheim headed such an organization in France, as did, in our own day, Pierre Bourdieu. But later social scientists less often possessed such corps of subordinates.[2] This was particularly true of the theorists, in part because of another structural change, the emergence and stabilizing of disciplines. Disciplines facilitated knowledge growth in general, but they accomplished that facilitation in part via differentiation of ways of knowing things. Disciplines were groups of people who shared the same somewhat unquestioned assumptions about theory, methods, ontology, and so on. Having set vexing issues aside, these groups could very rapidly produce new material. But the sidelined assumptions meant that these rapidly increasing piles of knowledge refused synthesis not only by their quantity but also by their qualities—their conflicting intellectual foundations. Indeed, it was precisely the limitation of the qualities that permitted the explosive growth of the quantity. The difficulty of deductive work increased because the disciplinary conventions that facilitated work within disciplines impeded work between them. In a fractal manner, the same process eventually worked at the subdisciplinary level.

2 Coleman had a considerable body of students at Hopkins, but while he did train students at Chicago his theoretical interests began to drive his habitus, and he became more of a lone scholar.

Thus, deductive theory in part fell victim to the vast increase in the amount of things to know that resulted from growth and from differentiation, but in part fell victim to a new academic role structure that differentiation produced, a role structure whose rapid increase of production rested on limiting assumptions that precluded general synthesis.

In addition to these long-run structural trends, there is also the simple fact of pioneers' advantage. Late nineteenth century scholars assembled the social sciences by grafting scientific rhetoric, progressive politics, official statistics, and jurisprudence onto the main stem of the social philosophy that they had all read in university. Each writer had his own version of these things. And each employed it to address the modernist historical moment, with its heady mix of secularism, nationalism, imperialism, and class conflict. Theory as commentary was impossible because the novel historical situation had outmoded the theory that already existed. To the extent that there was such theory, it was written by jurists, lawyers, and judges, within a context that was not only normative but also everyday and practical.[3] (Moreover, it was to some

3 Spencer was an exception. Explaining the disappearance of Spencer from the social sciences is difficult. He was a brilliant man, a voluminous (perhaps too voluminous) writer, and had sponsored a vast gathering of data on comparative social life, one that would not be superseded until the Human Relations Area Files in the mid twentieth century. Durkheim's arguments against him are often tendentious and cannot really explain his disappearance. One wonders if Spencer's actual "problem" was that Social Darwinism became the ideology of the elite capitalists, one from which they have not deviated in a century and a half. The progressive cast—and perhaps the ressentiment—of the academics have therefore kept Spencerian thought invisible, except in economics, which elaborates Spencer without reading him.

extent outside the arts and sciences university.) So the late nineteenth century generation had both the necessity and the privilege of being the first in the field. Their successors fell easily into theory as commentary, because there was now a generation of theory available for comment. Simultaneously, theory as a repertoire of mechanisms emerged very quickly out of the rapid quantification of the social sciences, begun by economics before the turn of the twentieth century, and then wafted by the new inferential statistics of the 1920s and 1930s. But as already noted, theory as mechanisms is generally inhospitable to general deductive theory, and therefore, in the event, quantification provided yet another force dividing research practice and deductive theory.

Finally, there are enduring and purely intellectual reasons for the extreme challenge that the social sciences present to deductive theory. Put simply, the ontological realities of social life make deductive theory extremely difficult in the social sciences, as compared with the natural sciences.

First, social life does not have the instantaneous, continuous space-time characteristic of the natural world. Human beings carry within their minds both memories and anticipations. The remembered past and the conditional future thus exist in the present of the social process, and the past in the present is being variously recorded and perpetually rewritten in that present, just as the future in the present is being continuously peopled with a vast and constantly changing array of contracts, options, and promises. Natural science has no such space with overlapping temporal durations, nor has mathematics developed the tools for dealing with it.

Second, the social process contains two different kinds of entities. One kind is anchored in single human organisms. It thereby has long but finite endurance, as well as extraordinary memory and mental capability.

The other kind of entity is what computer scientists would call a distributed, parallel processing system, made up of separate parts of entities of the first kind, and easily reconfigurable. Such "social groups," as they are called, tend to last for a shorter time than do the organism-based entities, but they have potentially infinite "lifetimes" as lineages of renewable structures. They also embody the confusions as well as the power of distributed parallel processing. And they are not concentric, but overlap in a crazy-quilt of patterns. Note also that because all entities of each kind are constituted of parts of entities of the other kind, social ontology is fundamentally dual. This means that causal forces affecting one kind of entity necessarily affect the other kind, whether those forces be direct determinations or determinations conditional on the various relations of entities. Again, there is no system in the natural world with these dualistic ontological properties. Nor does any elaborated mathematics exist for them.

Third, this entire system knows itself and acts through systems of purely symbolic representation that are, by definition, arbitrary and conventional. These representation systems are yet a third kind of entity, and since they are purely conventional, they are endogenous to the social process. The conventions undergirding them are perpetually at issue, and many of them have meanings that are in principle indefinite. In fact, the most powerful representations are the most indefinite—words like "democracy," "justice," and "truth." The situation is the same as if a computer program, while executing, could modify the compiler that was translating it into machine code and could change the operating system that was running the physical machine underneath. There may be experimental work on such phenomena in computer science but no system of this kind is known elsewhere.

These three properties do not necessarily refute the possibility of deductive theory in social science. And they of course were no different in the late nineteenth century than they are today. But taken together they imply that any pre-existing system of ideas to be borrowed from the natural sciences would have little potential for effective analysis of social life. Deductive theory in the social sciences must start from its own first principles, and this fact may have increased the effects of the first mover advantage that so much privileged the first and "canonical" generation in the late nineteenth century.

In summary, there have been many internal forces—both institutional and intellectual—that have militated against deductive theory in the social sciences in the last century and a half. And they all exacerbated the negative forces considered in the earlier discussion of the rarity of deductive theory: differentiation, narrowing, overproduction, and first mover advantage.

*B. Normativism*

Internal forces against deductive theory were themselves exacerbated by changes in the normative connections of theory. Those normative connections begin in the obvious difference between the objects of study of the natural and social sciences. While the objects of social science are human beings, who have normative concerns, the objects of natural science are almost without exception phenomena and structures that have no normative content in and of themselves.

They can give rise to normative debates, as have atomic power and human-induced climate change. But in themselves they involve no normative matters.[4]

More specifically, the important difference between human beings and the objects of natural science is that the former have wills, while the latter do not. To be sure, if, like Milton Friedman, one is willing to assume that humans are "rational dopes" (that they are simple maximizers of a given function of their preferences), then the fact that humans have will doesn't matter. They can exercise their wills in only one way—rational choice guided by preference schedules. Therefore, one can "reverse engineer" the choice pattern and thereby discover the preference schedules—which actually determine everything—from the behavior of individuals. Of course, the veracity of this reverse engineering is conditional not only on the rationality assumption, but also on the topology, specificity, substitutability, and accessibility of the preferences themselves, not to mention their dependence on assumptions about the knowledge of all these factors by the human deciders themselves.[5] So there are many further problems with the rational dope approach, as generations of economists have themselves argued. But more broadly, if one is *not* willing to assume that human beings are rational dopes, then one

4 This follows in part by definition. "Normative matters" is simply a name given to a certain set of motivations by human organisms. It is not clear that the term could have any meaning for other organisms or natural phenomena.

5 Another big assumption is that humans can make such calculations. Neo-classical economics defended itself against socialism by insisting that the clearing of trade in a centrally-set price system was a computationally impossible problem (this was the issue of the so-called "socialist calculation debate"). Unfortunately for the neoclassicals, precisely the same argument obtains against individuals as resolvers of their own internal calculation of tradeoffs.

must try to theorize the will, and that obligation moves one immediately into the realm of normative considerations, since normative social theory is the body of work that thinks most actively about the will, and since no society has ever existed without a normative structure embodied in some concept of "oughtness."

This being the case, it is all the more striking that despite this seemingly inexorable logic driving social theory towards normativism, the history of social science contains many attempts to avoid normativism. As the very phrase "social science" suggests, there has been since the mid-nineteenth century an impetus to "scientize" the study of society, in the specific sense of treating society as an empirical, explainable phenomenon like a planetary system. The American progressives, to be sure, believed that normativism raised no real issues: for them, social science was simply reform made more expert. But after 1920, science and reform—indeed science and normative concerns more generally—were opposed to each other, in the US at least. The 1920s became a decade of "scientizing" in all of forms of social study. Part of that scientizing was a positive move towards quantification, stimulated by the rise of inferential statistics in the late 1920s and the 1930s. But another part of it was negative—ridding the social sciences of normative content. Normative matters were to be reserved for the polity. The polity would decide what to do, even if the social scientists would advise how best to do it. This "engineering model" of applied social science had the advantage of avoiding the bitter normative and class debates that had arisen from the strong progressivism of prewar social science.[6]

6 This engineering model was made explicit in Dewey's *The Public and Its Problems* (1927).

But in the process of scientizing, many empirical students of society themselves embraced the natural science ontology and, in order to do so, assumed away not only human will but also such other human qualities as symbolization. Fisher's metamorphosis of thermodynamics into economics in the 1890s was only the first of many such direct borrowings. Robert Park borrowed plant ecology for sociology in the 1920s, George Zipf, John Stewart, and others borrowed classical physics for social science in the 1930s and 1940s, Dudley Duncan, Hubert Blalock, and others borrowed causal analysis from experimental biology in the 1960s, and a number of people borrowed the techniques of catastrophe theory in the 1970s.[7]

As these examples show, it is the *methods* of natural scientists that provide the greatest temptation, for they give the appearance and hence the status of scientificity. Borrowing methods has therefore always been a favored strategy of young and ambitious scholars, for it provides premade practices (usually with related sub-theories), and simultaneously deskills the elders who stand in the way of the young people's advancement. To be sure, the intellectual results of borrowing are underwhelming. With the exception of Fisher's borrowing of thermodynamics and, perhaps, Spencer's adaptation of Darwinism to social life, most borrowings have not transformed the theory of the social sciences that did the borrowing. Undoubtedly part of the reason for this failure is that humans *do* in fact have will and symbolization, and that only relatively trivial things about social life can be predicted without attending to such phenomena. But another reason is that borrowing from the sciences imposed on the borrowers simplifications made by the natural scientists themselves:

7 In the 1980s, I myself developed social sequence analysis on the basis of such a borrowing, from computer science and biology.

replacing irregular masses by points, disassembling complex particulars into hypothetical main effects, and so on. Thus the temptations of borrowing natural scientific methods and ways of thinking had a double effect in dividing empirical from theoretical work in the social sciences; not only did they impose the simplifications involved in treating humans as objects, they also imposed the simplifications involved in treating complex objects as simple ones.

But there is another pathway by which normative concerns divided theoretical and empirical work. As has become clear with time, every means of gathering data favors some political viewpoint, or creates the ability to ignore or to foreground this or that normative problem. But this ineradicably political aspect of data itself questions the entire separation of facts and values on which the engineering approach to social science has been built, and which is a precondition of any purely "natural scientific" (i.e., non-normative) social theory. This still continuing problem hangs over the entire project of social science qua natural science. And in the short run, it too tends to separate empirical and theoretical work, because of the great differences between the two concerning what aspects of the social process can legitimately be assumed away in order to reap the ambiguous benefits of scientization.

In summary, by several means, borrowing from the natural sciences has helped produce a split between theoretical and empirical work in social science. Eliding and ignoring normative concerns was necessary to enable such borrowing, yet such elision seems not only mistaken ontologically, but also impossible empirically.

A curiously similar process emerged even in those parts of social science that did not elide normative concerns, but rather embraced them. We see this process, for example, in much of American "political theory"

and in Germany's philosophical tradition of social theory. Where empirical social science was tempted by natural scientific methods, high social theory was tempted by a quite different external body of work—not natural science, but jurisprudence. Humans require some kind of system to adjudicate their conflicting wills, and law has been that system. It is therefore little surprising that jurisprudence—law's theoretical subdiscipline—should have produced masterpieces of social theory. Hobbes' *Leviathan* rests completely on common law theories, as do the writings of Locke. Rousseau famously invokes the legal concept of contract. In Germany, jurist Rudolf von Ihering provided the theoretical and conceptual framework that undergirds the work of Max Weber, who was himself trained as a lawyer. In the twentieth century, legal theorists like H.L.A. Hart, Hans Kelsen, and Ronald Dworkin were among the most important of social theorists. Indeed, much of social theory from the seventeenth century forward has been created within law and jurisprudence.

But the legal sphere has a fundamentally different relation of theory and practice than do the social sciences. Even the most abstract legal theories must be directly applicable to the practical task of adjudicating human disputes, just as theory in physics relates directly to experimental work. But social theory per se is not required to meet this test. Thus its relation to jurisprudence has been only a partial one, like its relation to the natural science model of theory. It can borrow the normative abstractions of jurisprudence, but it does not need to submit those abstractions to practical tests. Thus, just as social theory began to float free of empirical analysis because of the trend of scientization in empirical social science, social theory could also float free of the empirical *normative* world of legal disputes, because the conclusions of social theory were seldom applied in practice: under the engineering model, *practical*

application of social science drew *only* on the repertoire of theory as mechanisms. Put another way, while on the scientific side social theory drifted away from empirical practice because the empirical scientizers adopted the natural science ontology so ill-prepared to analyze social life, on the normative side social theory drifted away from empirical practice because it had no need to adapt its thinking to the practical matter of judging disputes.

In summary, normative concerns are built into social thought by the nature of the object of that thought, which is human activity. A number of ways of relating the normative to social theory have been tried: ignoring the normative, as in the scientistic social theory of the economists; assuming that the two are mutually reinforcing, as in progressivism; and isolating the normative from any detailed contact with empirical work (as in most philosophical social theory). The first and last positions dominate today's social sciences, resulting in a kind of mutual ignorance between serious empirical work and philosophical cum normative social theory. The problem of normativity has thus played an important role in dividing theoretical and empirical work within the social sciences, which in turn may have helped make deductive social theory less common.

### *C. Theory and practice*

We have shown the challenges raised to deductive theory by its own institutional and intellectual qualities. And we have shown the challenges raised by theory's inevitable imbrication with normativism and by the problems consequent upon avoiding normative complexities via scientization. The analysis turns now to the relation of theory and empirics, which has already emerged as drastically affected by the problem

of normativity. As noted earlier, the relation of theory to empirical work seems to have changed over the last century. The late 19th century theorists were awash in empirical data, while modern theorists largely ignore data. In part, as we have seen, this may have to do with expansion and differentiation—internal forces within the 20th century knowledge enterprise. In part, it may have to do with the place of normative concerns in a social science sometimes operating under the engineering model. In part it may have to do with scientization per se. But there are other reasons as well.

It is perhaps useful to contextualize the question by recalling the relation of theoretical and empirical work in the *natural* sciences. The *institutional* difference between theory and empirics in the natural sciences is usually quite pronounced, but their *intellectual* relation is often very close, as in physics. To be sure, other sciences have looser relations between theorists and empirical workers. Much of biology is applied science whose research puzzles are set externally by the health industry. In such research, "theory as mechanisms" dominates. But at the same time, biology does have ongoing theoretical controversies in areas like genetics and expression, and in such areas, there *are* many scholars who are pure theorists, and they are often closely tied to empirical work. Elsewhere in biology, induction plays a major role. Induction was typical of long stretches of biology prior to the twentieth century, as it was of geophysics. Both areas were eventually theoretically systematized, biology by the theory of evolution and geophysics by plate tectonics. But it is noticeable that in both cases, these theoretical syntheses took a long period of gestation, during which there was consistent interaction between theorists and empiricists. And in both cases, the theory rested on an enormous foundation of inductive empirical work.

In summary, the natural sciences tend to have clear relations between theory and empirical work. Individual scientists usually specialize in one or the other type of work (at least in the twentieth century), but the intellectual relation of the two remains intimate, although it may vary in kind.

In the social sciences, the current distance between theory and empirical work seems much larger. Compared to our late nineteenth century predecessors, our current theorists work almost completely with stylized abstractions. The major theoretical efforts of the last half century in economics have been formalizations of the great work of the 19th century, rather like Laplace's cleanup of celestial mechanics in the eighteenth century. There is no attempt to unseat Menger's limitation of economics to a theory of scarcity, for example, nor to replace the general equilibrium theory of Jevons and Walras. Game theory remains a tool, not a complete paradigm. Nor is there any area of social science where theorists and experimentalists work in close collaboration as they do in physics, with the exception of the emerging field of experimentalism in economics and parts of political science. And even there, nothing is expected to shake the foundations of the deductive economic paradigm that undergirds both literatures. Quite the contrary, that paradigm is assumed in the experimental designs.

If we look elsewhere in the social sciences, the gap between major theoretical work and everyday empirical work is almost absolute. As we have seen, most of the deductively organized theoretical work in the social sciences deals with polysemous abstractions, while most empirical work in the social sciences works with a simplified repertoire of mechanisms involving well-defined variables. Moreover, most of the purely theoretical work deals with normative issues, while, by contrast, although normative issues have certainly begun to dominate the choice of topics and data in

the empirical literature as well (in the guise of identity politics and social welfare applications), in *procedural* terms, normative issues remain invisible in our empirical work. The methodological pretense of contemporary empirical work in the social sciences remains one of impartial objectivity and engineering, and the normative and political views are smuggled in as contraband. By contrast, the theoretical literature considers normative issues quite openly and on their own terms.

But there may be another causal pathway here, one that is internal to the social sciences themselves. It too involves youth and age, but in a different way than does the age mechanism already discussed—borrowing from the natural sciences. It was noted earlier that social ontology has three properties that differ from those of natural ontology: it has a different space-time, it is dual between individuals and groups, and it is inevitably symbolic. From our daily life, we take these facts so much for granted that it takes both time and varied experience to realize that the three of them actually forbid the use of the scientific routines that we have all learned from secondary education onward. Those who recognize this difference early in career often reject the scientific approach altogether, and embark upon what they conceive to be a fundamentally different knowledge project, one closer to the arts and humanities than to natural science.

But if one remains in the no-man's land between the sciences and the humanities, one's career faces a fundamental tension because of this ontological problematic. One can handle this tension in one of two ways. Either one specializes in some particular paradigm, with its local principles and its local cumulation, or one develops multiple scholarly identities that reproduce within oneself that very ontological tension between paradigms. One is a quantitative researcher with quantitative researchers, a philosopher with the philosophers, an ethnographer with the

ethnographers, and so on. One's own work is filled with the implicit contradictions of these various selves. Now, one *can* recognize these contradictions as theoretical opportunities, but only if one has done much empirical research and engaged with many different methodologies and disciplines. This takes years of experience, which are made unpleasant by contact with specialized researchers with various kinds of epistemological blinders, who speak anonymously from the ambush of peer review.

Modern academic career structures militate against such years of diversity as much as do the unpleasant experiences themselves. In the US prior to the 1970s, expansion guaranteed employment and tenure to nearly all PhDs, and as a result, young scholars had time to think and grow. But after 1975, hiring became very competitive, and what was at first a drift towards overproduction rapidly developed into the machine-like publication of the present day. The incentives in our current system are hostile to any form of sustained, careful reflection and to any attempt at non-specialization. Quite the contrary, the incentives favor choosing both specialty and method early in one's career. Given the quite rapid turnover of quantitative methods in the post 1975 period, each new generation can revolutionize its field with some new statistical wrinkle. Nor do the qualitative scholars lack their own fads and fashions; the historical, cultural, linguistic, and other "turns" embody the qualitative equivalents to the new quantitative methodologies. In such an environment, the incentives to develop a life-long theoretical project are minimal. Such a project could be accomplished only by its concealment within a long string of intermediate results of seemingly unrelated kinds.

Even among those who specialize in theory itself, one can see that the theory of the young differs from that of their elders. When one is young, social theory seems simple. It seems a matter of clearing up the confusions of benighted predecessors and setting things straight. When one is older,

such clarity comes only with immense work. This life trajectory is evident in the comparison between the simple-minded facility of Durkheim's *Division of Labor* and the tortured complexities of the later *Elementary Forms*. In Marx, there is the contrast of the simple clarity of the 1844 manuscripts with the endlessly wrought analysis of *Kapital*. In both cases we see the difference between attributing theory's problems to simple unclarities in other people's theoretical work and attributing theory's problems to the maddening unclarity of social life itself.

Once one has faced that maddening unclarity in a variety of areas, one's theoretical ideas look different. If one has pursued a consistent intellectual project, one's ideas have met many and various challenges over a long time, particularly if one reads widely, teaches students of different levels, and uses a variety of methods. One's ideas become more abstract, but also more idiosyncratic, interlocked in a way that seems impenetrable. One discovers that one has a system, but that one doesn't fully understand it oneself. In this connection, it is a striking fact that while many great mathematicians and scientists do their work early, the great philosophers do their best work from 50 on. Hegel is an exception to be sure, although one could argue that his early work lacks the depths of his later triumphs. But Aristotle, Plato, Augustine, Hobbes, Locke, Rousseau, Kant, and Dewey all did their main work from age 50 onwards. Aquinas started his *Summa Theologica* at 40, to be sure, but left it unfinished at 49. Rawls' *Theory of Justice* came at 50. Wittgenstein's highly logical *Tractatus* came at 33, but there was nothing but his fascinating notebooks at the time of his death at 62.

This pattern suggests that the kind of knowing traditionally associated with philosophy—and with social theory as well—takes long maturation. Mature social theory reflects diverse engagement with data, and diverse engagement with data takes time. It is striking that not only do

we see this in Marx, Weber, Spencer, and Durkheim, we see it also in many of the great jurisprudential writers: Hart, Dworkin, and Ihering all produced their great works after 50. Even economics has Marshall and Keynes, both of them data analysts, and both publishing their major theoretical works at around fifty.

In summary, there are a variety of mechanisms making the connection of empirical work with deductive theory precarious. But it may well be that the main culprits in dividing the two are on the one hand the internal forces inherent in social science as a project (including the whole problem of normativism) and on the other hand the changes arising in the success and increasing size of social science, and in particular the changes arising in the changes that success and size have produced in the institutional structure of disciplines and careers.

The exact relation between all these causes must remain a matter of speculation. But between them they have created a deep division between empirical work and general, deductive theory in the social sciences of the present moment. One cannot understand the relative rarity of deductive social theory today without considering interlocking trends in the ideas and institutions of social science itself, in the project of normative as well as empirical study of society, and in the complex relation of theory and empirical work. There will be no simple story to tell, but rather a continuing swirl of processes, of varying extents and durations.

## *Ideals for knowledge*

To this point, the analysis has dealt only with the past. In closing, it seems proper to turn from past to future.

The future of theory calls not for explanations, but rather for ideals. About these, I can merely speculate. That there *are* debates over ideals for

knowledge is clear from the famous *Methodenstreit* between the Austrian and German theorists of economics. In that debate, the central question was whether there are non-trivial universal laws in a given field, or whether those laws were absolutely conditioned by time and place. As this posing of the question suggests, both sides agreed that there were laws. The difference was simply over whether the laws were universal or local. The ideals that were behind these laws remained the same in either case: the ideals of both sides were cumulation and "approximation to truth," whether in the sense of predictive power, explanatory power, or utility for policy.

These ideals of cumulation and approximation raise the possibility of alternative sets of ideals for knowledge. For there may exist forms of genuine knowledge that are not cumulative and that do not try to approximate some universal or local truth. The classic example is knowledge of beauty—aesthetics. In such a non-cumulating, non-approximating form of knowledge, the ideals are more commonly things like plenitude (defined as filling the space of possible knowledge) and limited recursion time (defined as the guarantee that no area of the knowledge space disappear too completely for too long). Such ideals are evident in some modern disciplines: philosophy is an example. Note that this second kind of knowledge involves a universal/local controversy exactly like that of the *Methodenstreit*, if we argue about the potential for differing civilizations in terms of different mixes of values, or differing "art worlds" in terms of varieties of aesthetics.

But while there are many interesting questions about the *scope* and *dimensions* of knowledge ideals, the present moment confronts us with a much more pressing choice about knowledge ideals: not so much what they are, as who shall set them—ourselves or others?

Over most of the twentieth century, knowledge communities set their ideals for themselves. Today, however, there are two external groups that want to control not only the ideals of knowledge, but also its very definition.

The first are businessmen. The removal of the universities from the control of their faculties has meant that the knowledge agendas of the universities are now set by business and government interests rather than by faculty ideas. This is a crisis because the universities had become, by the middle of the twentieth century, the main—very nearly the exclusive—location of intellectual life in modern societies. They were places where intellectuals could use large amounts of other people's money to pursue their own knowledge agendas. That era is now over. Intellectuals are unwelcome in the modern university with its neoliberal management, its excessive publication expectations, its vocationalized students, and its businessman administrators. Even many of our academic colleagues have given up on traditional knowledge ideals, usually because their own ideals are in fact not intellectual but political or—occasionally—mercenary.

There is at present no clear alternative to the university as the institutional home for intellectual life. The question raised by this new character of universities is therefore how to create such a home, in which intellectuals can again create and follow their own agendas. And in doing so, we must become much more explicit about what it is to pursue genuine knowledge and to set intellectual agendas.

That is the first issue raised for knowledge ideals by external forces. The second such outside threat to knowledge ideals is the rise of "artificial intelligence" and of the sensationalist ideologues who support it. AI is not dangerous because it replicates human thought. It cannot do that, for a variety of fairly straightforward reasons, which would require a separate

paper to set forth.[8] Rather, AI is dangerous because it will produce a massive pile of "results," and its proponents will try to redefine knowledge as being simply the limit point of that vast process of machine-produced results.

This would not matter if AI were mainly being developed by people who genuinely knew something about the varieties of ideals of knowledge. But unfortunately, AI is largely in the hands of scientists and mathematicians, many of whom view non-scientific knowledge with feelings ranging from incomprehension to contempt. In particular, many of them know little about what we might call the associational form of knowledge, which has been characteristic of the arts and humanities. What will happen, then, is that the scientists' machines will produce enormous quantities of "results," and their advocates will persistently claim that we should ultimately define the word "knowledge" simply to mean whatever it is that the machines produce. An example of this is the "next painting of Rembrandt" produced a few years ago by some people with a great deal of hokum and a simple-minded extrapolation algorithm.

In order to be able to reject this kind of specious redefinition of knowledge, we will require a generation of scholars who can define independent standards that specify the nature of knowledge without assuming any particular means of producing it. Those scholars will certainly need to know a lot about computing. They must be able to say exactly why that interpolated portrait is simply one variety of pixel-based

8 The underlying reasons are two. The first is that the algorithms are mechanical programs running on digital machines, whereas the human brain is simultaneously a digital and analog device. The second is that brains operate in bodies that have desire and will, which computers lack.

average of Rembrandt's known work and why such an average is not art. They will need to go inside all kinds of algorithms and show where the algorithms make philosophical assumptions (not just mathematical ones) that are arbitrary, idiosyncratic, or silly. At the same time, such new scholars will need to set forth the position that there are different forms of rigorous knowing, and that these different forms can have fundamentally different ideals and modes of production. Above all, they will need to develop the formal theory of associational knowledge, and perhaps other forms of knowledge that are—like associational knowledge—beyond replication by digital computers. At present, there seems to be very little formal thinking about such knowledge. But doing that thinking is the main task of intellectuals in this moment.

By 2030 the machines will probably be able to produce articles that will look like the first drafts of our graduate students. To be sure, they will merely be extrapolations and interpolations. But they will exist, all the same. And such machine products will challenge those of us who want to protect intellectual life to come up with an explicit theory of why machines are not capable of intellectual life, or, to put it another way, why extrapolation and interpolation do not exhaust or cover the possibilities of thinking. More broadly, we shall need to invent a genuine characterization of knowledge ideals—a normative theory of "good knowledge" (possibly of several kinds), against which we can measure the productions of our mechanical colleagues.

These are great tasks. And only a renewed commitment to deductive theory will enable us to undertake them.

## *References*

Coleman, J. S. 1990. *Foundations of Social Theory*. Cambridge MA: Harvard University Press.

Dewey, J. 1927. *The Public and Its Problems*. New York: Holt.

Rawls, J. 1971. *A Theory of Justice*. Cambridge MA: Harvard University Press.

DOI 10.14279/depositonce-22431

JULIAN HAMANN & SINA FARZIN

# The Sociological Concern with Theorizing, and How It Could Be Complemented with a Focus on Media of Theorizing

Argument

Theories are often conceived as self-contained, uniform sets of propositions that can be judged by their systematic coherence or comprehensiveness. In the social sciences, textbooks illustrate how theories are often taught as sterile bodies of knowledge that can be attributed to specific authors. This notion of theories is not neutral. It can be considered a form of boundary work that values a specific way of thinking and setting up arguments, and devalues alternative ways of intellectual engagement (Krause 2016). When can an intellectual effort be considered a 'theory' and not, for example, a 'perspective,' an 'assumption,' or an 'idea'? From a sociological perspective, it is highly unlikely that some

Prof. Dr. Julian Hamann
Humboldt-Universität zu Berlin, Faculty of Cultural, Social and Educational Sciences
Email: julian.hamann@hu-berlin.de

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory*. *Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

forms of intellectual engagement are inherently 'theoretical.' Accordingly, 'theory' has very different meanings in sociology (Abend 2008), and it is only fair to assume that not all sociologists—let alone all social scientists—can agree on what constitutes a theory.

Our argument introduces sociological literature that, first, mobilizes different arguments on what constitutes a theory in the first place, and, second, provides different accounts on how theorizing can be conceived of as a social practice. The debate we introduce gained pace among sociologists in the last decade and reflected mainly on the discipline's own understanding and practice of theory. In mapping out this debate, we too neglect the discussion in other fields (for example, in biology, Pigliucci 2013; or higher education research, Hamann and Kosmützky 2021) but we think many aspects are relevant for most disciplines. We conclude with the suggestion that future research on theories and theorizing should attend to different media through which theories and theorizing are accomplished.

It is not new for sociologists to challenge the mystification of theories by highlighting that theories emerge from everyday practices (cf. Bourdieu and Wacquant 1992). These more or less mundane activities may comprise thinking aloud with colleagues, reading, scribbling in a book, taking a walk, sketching on a whiteboard, or moving around sentences and paragraphs in a Word document. Sociological research has thus highlighted that theorizing is a practice (Swedberg 2016) and a craft that can be observed, taught, and learned (Werron, Brankovic, and Ringel 2023). Just like a body of knowledge or an intellectual engagement is not inherently theoretical but has to be labelled and acknowledged as a 'theory,' the abovementioned practices are not inherently practices of theorizing. It is a discursive act of theoryfication that elevates practices that are otherwise considered to be mundane to activities that are

epistemologically relevant and have the status of theorizing (cf. Ploder and Hamann 2021). Following this argument, thinking aloud with colleagues and taking a walk are two activities that are not inherently different. It needs discursive acts of theoryfication that label one as an act of 'theorizing' and the other as a mundane everyday practice. Deconstructing the scholastic notion of theories as self-contained, aloof, and exceptional bodies of knowledge, research on theories and theorizing has facilitated important insights on this contextual everyday production of scientific knowledge, on a scholarly craft and the tools that it draws on, and not least on how notions of theory can serve as boundary work (Gieryn 1983) and consecrate only very specific intellectual work (Bourdieu 1990).

The sociological discussion on theorizing can be traced back to Robert K. Merton (1945), who distinguished types of theories and assessed them according to their contributions to and limitations for the advancement of sociology. The literature gained pace in the 1980s, when Jeffrey Alexander published a five-volume book series on theoretical logic in sociology. One of Alexander's (1982) key arguments was that there is no categorical distinction between empirical observation on the one hand and theory on the other, but that both are connected along a continuum that spans from general presuppositions and models over classifications and laws to methodological assumptions and empirical observations. The more methodological concern of Alexander's work has been taken up by Andrew Abbott (2004). This strand of the discussion is interested in how we could or should use theories. It is normative in a methodological sense because it has a clear idea of what constitutes good research—for example, rigor, explanation, or the combination of a puzzle and a "clever idea" (Abbott 2004, xi) that responds to or solves the puzzle.

A second strand of the discussion on theory and theorizing is not methodological, but employs a semantic or discursive perspective. Gabriel

Abend (2008) has distinguished different ways in which the notion 'theory' is used in sociological language. This intervention is not methodological in that it does not feature a normative idea of what constitutes good research: Abend is not concerned with what should or should not be called theory. Rather, he offers a meta perspective on the different meanings that the word 'theory' can have within a specific discipline—in this case, sociology.

Complementing Alexander, Abbott, and Abend, a third important strand of the discussion is represented by Richard Swedberg and Monika Krause. While the first strand of the discussion is methodological and concerned with how theories contribute to a normative notion of good research, and the second strand entertains a semantic and lexicographic interest in what sociologists mean when they use the word 'theory,' the third strand of the discussion shifts the focus to the actual practice of theorizing. Swedberg (2012) introduced the concept of theorizing to point out that theories are not just fixed or static sets of propositions and statements. According to him, theories are considered merely an end product that has to be accomplished through practices—for example, naming, conceptualizing, or constructing typologies. Building on this insight, Monika Krause (2016) has further differentiated this argument, connecting the perspectives of Swedberg and Abend by distinguishing different meanings of the practice of theorizing. In her view, the meanings of theorizing reach from, for example, the interpretation of classical authors to the application of existing concepts to empirical phenomena.

In recent works, all three strands of the debate are used to develop an empirical perspective on the topic of theorizing. Two dimensions of the process of theorizing are investigated: a) practice-oriented approaches, and b) approaches with focus on the materiality and media-dependency of such practices.

Concerning a), practice-oriented perspectives complement the epistemological strategies in written texts, with a focus on the broader social contexts in which texts are produced. Werron, Brankovic, and Ringel (2023), for example, applied practice theoretical approaches to their own theoretical endeavors within a research project. They draw on insights from Martus and Spoerhase's (2022) case study on the importance of collaborative practices and contexts in the field of literary studies, and reflect on their own everyday practices as researchers, such as informal talks, data sessions, and note taking.

With regard to b), recently, some authors highlighted the materiality and media-dependency of theorizing. Swedberg (2016) refers to the usefulness of visual elements such as diagrams or sketches, and considers diagrams as useful for developing or communicating the actual theoretical concept, which is a written text. Guggenheim (2024) attributes more agency to visual elements. Drawing on insights from the field of science and technology studies (STS), he proposes to understand diagrams or sketches as important media in the process of translating knowledge (which in his understanding is the core principle of theorizing).

Future contributions to the literature on the practice of theorizing could even consider different media through which theories and theorizing are accomplished. These media go beyond pen and paper, or traditional outlets such as books or journal articles. The media of theorizing could include, for example, the radio as a traditional outlet of theoretical reflection for scholars such as Theodor W. Adorno, and encompass new forms such as social media platforms or podcasts. Theorizing is also accomplished through technologies such as Email, reference management systems, and through software packages like MAXQDA or Stata that guide our analysis and thinking. Reflecting on different media of theorizing could be promising for at least two reasons.

First, different media can be expected to structure attention in different ways—both the attention of the theorizer, but also that of their recipients. While the user interface and the options of different software packages structure how researchers theorize their data, the download and citation counts featured on digital platforms guide the attention of potential recipients of theory. Second, the case of podcasts as a medium of theorizing that is discussed in this yearbook suggests that different media can be expected to facilitate different forms and dynamics of communication. They range from individual thinking to monologues, and to dialogues, collaborations, or direct engagement with an audience.

## *References*

Abbott, Andrew. 2004. *Methods of Discovery. Heuristics for the Social Sciences*. New York, London: W.W. Norton & Company.

Abend, Gabriel. 2008. "The Meaning of 'Theory.'" *Sociological Theory* 26 (2): 173–199.

Alexander, Jeffrey C. 1982. *Theoretical Logic in Sociology, 4 Vol.* Berkeley: University of California Press.

Bourdieu, Pierre. 1990. "The Scholastic Point of View." *Cultural Anthropology* 5 (4): 380–391.

Bourdieu, Pierre, and Loïc D. Wacquant. 1992. *An Invitation to Reflexive Sociology*. Cambridge: Polity Press.

Gieryn, Thomas F. 1983. "Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists." *American Sociological Review* 48 (6): 781–795.

Guggenheim, Michael. 2024. "Theorizing is not Abstraction but Horizontal Translation." *Distinktion: Journal of Social Theory* 25 (2): 165–182.

Hamann, Julian, and Anna Kosmützky. 2021. "Does Higher Education Research Have a Theory Deficit? Explorations on Theory Work." *European Journal of Higher Education* 11 (5): 468–488.

Krause, Monika. 2016. "The Meanings of Theorizing." *The British Journal of Sociology* 67 (1): 23–29.

Merton, Robert K. 1945. "Sociological Theory." *American Journal of Sociology* 50 (6): 462–473.

Pigliucci, Massimo. 2013. "On the Different Ways of 'Doing Theory' in Biology." *Biological Theory* 7 (4): 287–297.

Ploder, Andrea, and Julian Hamann. 2021. "Practices of Ethnographic Research: Introduction to the Special Issue." *Journal of Contemporary Ethnography* 50 (1): 3–10.

Martus, Steffen, and Carlos Spoerhase, 2022. *Geistesarbeit. Eine Praxeologie der Geisteswissenschaften*. Frankfurt am Main: Suhrkamp.

Swedberg, Richard. 2012. "Theorizing in Sociology and Social Science: Turning to the Context of Discovery." *Theory and Society* 41 (1): 1–40.

Swedberg, Richard. 2016. "Before Theory Comes Theorizing or How to Make Social Science More Interesting." *The British Journal of Sociology* 67 (1): 5–22.

Werron, Tobias, Jelena Brankovic, and Leopold Ringel. 2023. "Theorizing Together." *Distinktion: Journal of Social Theory* 25 (2): 228–249.

DOI 10.14279/depositonce-22432

TOBIAS WERRON, JELENA BRANKOVIC & LEOPOLD RINGEL

# Collaboration as a Medium of Theorizing

Argument

This is a condensed version of ideas elaborated in our article "Theorizing Together. Distinktion: Journal of Social Theory (2023): 1–22."
This version was prepared at the request of the editors and based on the transcript of the conference.

Let us start with a few words about how we understood this concept of the media of theory, because we think it's not self-evident that collaboration could also be considered as a medium of theory. In our view, the most important implication of discussing the media of theory is that it draws our attention to the process of the production of theory rather than just the end results. This is our usual focus, if we teach theory: We present

Prof. Dr. Tobias Werron
Bielefeld University, Faculty of Sociology
Email: tobias.werron@uni-bielefeld.de

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory*. *Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

to students the books and articles that are important to us or that we consider to be important for our discipline. In sociology, this might be Marx, Weber, Durkheim, Simmel, or more recent theorists, but it's always kind of large—often intimidating—books and articles that contain very important insights that you have to read very closely. They're difficult to understand in many cases, maybe too difficult for you at this stage of your career, so it has an intimidating aspect to it. Whereas when we consider how we actually *produce* or *develop* our own theories or ideas—on how to conceptualize an article, produce an argument, make a point and so on—then we do different things. We're not just reading papers. We're not just having ideas about concepts. We have an *everyday practice* through which we actually come up with these ideas. We write them down. We test them. We write a paper—a first draft, a second draft, and a fourth draft. Sometimes we then find 28 drafts of the paper in our first folder, and then there's a second folder, and then there's the first real version, and so on.

We say theorizing is a craft, like any other *craft*. There's nothing mysterious about this craft. You can learn a craft. You may have experience of practicing a craft, and so on. It's just various practices bundled in everyday activities. They are mundane, but they are also mysterious in the sense that we often don't talk about them in our teaching. That's the problem. But we *could* talk about them, and we *should* talk about them. That's the main idea behind our paper. We should ask: what *kind* of craft is theorizing? We can observe and describe what we do on a daily basis. We could write books and articles about it, making our craft explicit for others. We can teach it. So why don't we? It's possible. It hasn't been done to a large degree, but we *could* do it. And in that sense, we should understand theorizing as a craft—consisting of taken for granted everyday practices—that can be taught, learned, studied, just as we teach theory (in terms of published books and articles) and

quantitative and qualitative research methods. In other words, theorizing in this practical sense needs a methodology just like other parts of our sociological craft.

So what are these *practices of theorizing*? If we think about these everyday practices, then there are several candidates that might be obvious to all of us working in academia, and sociology in particular:

- Conceptualizing and generalizing;
- Searching and reading;
- Defining research questions;
- Selecting, casing, and sampling data;
- Making and reviewing mistakes;
- etc.

And then finally, again, collaborating. If you accept all of these as daily practices that are important for how we produce theory, then collaboration too can be an important part of how we produce theory and should be part of a practical methodology of theorizing. In our paper (Werron, Brankovic and Ringel 2023), and we have to be very brief here now, we distinguish different styles of collaboration and focus on what we call the *synergetic* style. Here, the basic idea is that you start a collaboration without knowing the perspective that you want to develop. You develop the perspective—which you use to look at the phenomena that you're studying—*during* the course of the collaboration. So it's open-ended. Ideally, it's also on equal footing: We don't start with a hierarchical relationship between researchers, but instead with people bringing their different competences, ideas, and abilities into the collaboration and then working something out together, being loosely interested in the same thing. And then we could say theorizing together, in the sense that we argue in the paper, is a mode of theorizing adopted by two or more scholars with the purpose of developing a shared perspective on some

research topic or question in the course of everyday collaboration. So the success of such a collaboration—the practice of *theorizing together*—would be that it actually results in a common perspective. This is what we would see as the primary criteria for success of collaborative theorizing—not how many citations you get or the amount of research funding, but rather: Do you actually manage to arrive at something like a common or shared perspective?

We now arrive at a point where we ask: what could be a *methodology of theorizing together*, of doing collaborative theorizing? Based on our experience, we suggest there are five kinds of strategies or practices—bundles of practices might be the better word—that we think are important for theorizing together.

- The first is to actually *assemble a team*, a group of people who are actually working together, and are likely to be *able* to work together. If you end up with the wrong people working together or people who don't like working together, who are maybe not into the experience of writing papers together and so on, then theorizing together is not likely to work out.
- The second practice of theorizing together that we would like to single out is what we could call *'thinking aloud together.'* This is also interesting because we could say this is specifically theorizing *without* media. It's the experience that you need to be in the same room many times, discussing things together face to face, in order to come up with certain ideas, to get to know, respect and use the perspectives of your collaborators.
- Similarly, *collecting and sharing of material together*. We feel that cloud services are very helpful in that sense. So here, in our experience, media are actually very important; for enabling you to share material

instantly, and highlighting some texts, empirical material or ideas that you want your collaborators to look at.

- And then, there is the process of *writing together*—based on all kinds of media, of course; based on cloud services, for instance; using instant messages services; and even discussing and negotiating things via writing. So, basically, it's very important to write together, including in the sense of accepting, commenting on—and sometimes rejecting—what others do.
- And for us that is also one of the most interesting aspects of theorizing together: we can bring togetherness to *other* practices of theorizing. Most practices of theorizing, on this huge list mentioned earlier (and it could be even more extended) of maybe 20 or 30 practices of theorizing, you can easily do alone: You can conceptualize alone, interpret material alone, come up with research questions alone.

What actually does change in regard to these practices if you're doing them *together*? Doing practices of theorizing together draws attention to and helps you experience the standpoints of others. We'll always be different, and others bring different experiences and abilities to the table. If you're used to working on your own, it can be difficult to accept this—that your perspective on some research object isn't necessarily the most productive one or the only possible one. This is particularly important in our experience when it comes to the formulation of research questions. People have different ideas about this and pursue different strategies in order to come up with interesting questions. You may have a certain way of going about this that is productive for you. But it could become even more productive if you combine your strategies with the styles and abilities of others. And the last point (we only discovered this in the final stages of our ongoing collaboration) concerns the different knowledge of, and capabilities of using, practices of theorizing and bringing them to the

collaboration. Some people might be better in analogizing, making analogies to other phenomena. Others might have their strength in abstracting, generalizing from something. All of these abilities can come together in a team.

These methodological points are, of course, based on our own limited experience with collaborative theorizing—this is just one paper of (we can guess) many research papers that could report on the experiences of other research teams and come up with additional methodological ideas. In fact, this would be very much in the spirit of our paper: we see it as *a collective task to come up with a methodology of theorizing together*.

## *Reference*

Werron, Tobias, Jelena Brankovic, and Leopold Ringel. "Theorizing together." Distinktion: Journal of Social Theory (2023): 1–22. https://doi.org/10.1080/1600910X.2023.2259288

DOI 10.14279/depositonce-22433

PETER FISCHER

# Reflexive Philosophy of Science as an Instrument in the Social Sciences

## Abstract

Based on a double tension, this essay reconstructs the relationship between the social sciences and the natural sciences and philosophy. The result is a continuing interaction, but also an increasing autonomy, which is also reflected in efforts to develop an independent methodology. Numerous phenomena are currently challenging the social sciences: Skepticism towards science, as well as the pluralization and closure of lifeworlds, are changing the knowledge-order of modern society. In a final section, an attempt is made to provide an answer to these challenges; the self-reflection of social science practice is outlined as a reflexive methodology.

PD Dr. Peter Fischer
Technische Universität Dresden, Institute of Sociology
Email: peter.fischer1@tu-dresden.de

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023.* Berlin Universities Publishing.

## *Introduction*

Two impulses were decisive for the following essay. Firstly, a response to my textbook on the philosophy of science in the social sciences published in 2023 (Fischer 2023b), which prompted me to ask whether a unified philosophy of science in the social sciences based on the philosophical tradition is possible. Secondly, an invitation to the theory conference of the Robert Merton Center and the Berlin Society for Science Studies at Humboldt University in October 2023, at which I gave a short lecture on the benefits of the philosophy of science for the social sciences.

In the following, these impulses are taken up and continued as given below. Starting from a dual relationship of tension between the social sciences, natural sciences, and philosophy that is still effective today, I present a brief historical-genetic reconstruction of the relationship between the two. In the case of the natural sciences, which are the subject of the first section, a lively import of metaphors and models into the social sciences is evident; nevertheless, there is no recognizable methodological compatibility between the two types of science. However, the different logics of the natural and social sciences are not a new finding, but can already be recognized in the discussion of natural philosophy and its application by Thomas Hobbes. In the case of philosophy, which is dealt with in section 2, a protracted process of detachment of the social sciences from the bosom of philosophy can be recognized. The autonomy of the social sciences ultimately leads to a differentiation of several coexisting paradigms. One consequence of this development is that the claim of a general methodology valid for all sciences is increasingly being questioned. In addition, current trends in the theory of science identify several indicators for the distinction between scientific and everyday knowledge and also take scientific practice ("doing social science") as the starting point for a theory of science.

In section 3, I outline several phenomena based on the knowledge society that pose a challenge to the sciences, but above all to the social sciences. However, scientific skepticism and shifts in the knowledge order of modernity should not be answered with a politicization of the social sciences, but with a reflection on their own activities.

This ultimately leads to the fourth and concluding section, using philosophy of science as a reflexive instrument that accompanies the research process and raises questions about the scope, significance, epistemological interest, and position of the researchers.

## *Natural and social sciences: Metaphor import, but no method compatibility*

The methodological relationship between the natural and social sciences has been the subject of extensive thought and writing since the beginnings of the modern social sciences in the early modern period. Thomas Hobbes (1588–1679) already spoke of an advanced natural philosophy and a much less developed social philosophy (Hobbes 1912). In Hobbes' time, it was above all the new mathematics of astronomy by Copernicus, Tycho, and numerous others that provided more accurate statements than the somewhat outdated theory of state and society (cf. Fischer 2023a: 102). However, the sciences played by different rules in the early modern period than they do today; for example, in the tradition of ancient philosophy, it was quite natural to embed society and the social sphere in a common cosmos with natural phenomena.

The consequences of such a disproportion in the structure of the sciences are still obvious in retrospect and can also be read in Leviathan (Hobbes 1982); not only do mathematical–mechanical methods permeate Hobbes' work, but the "exact" methods of natural philosophy were also

generally regarded as a model in the social philosophy of his time. Hobbes' explanatory power, despite his insistence on mechanics, is, like his popularity, remarkable. The import of mechanistic explanations and theorems finally became fashionable in the 17th century and a position on the new mechanical or the old Ptolemaic world view became the duty of every (natural) philosopher. A certain irony of the post-Early Modern history of science lies in the fact that it was only through the conscious separation from the natural scientific model that the social sciences succeeded in becoming independent. One consequence of early modern natural philosophy is that mathematical and scientific methods force social phenomena into a narrow corset that does not do justice to their phenomena. With regard to Thomas Hobbes, who can be seen here as a prime example of mechanical philosophy, this means that although he distinguishes artificial (social) bodies from natural bodies, he only has the methodological instruments of natural research at hand for their investigation.

The fact that the social sciences not only claim their own subject matter, but also insist on independent methods, does not mean that there will be no interaction with the natural sciences from now on—on the contrary. To date, however, there has been no conclusive clarification of the methodological relationships between the two scientific fields. On the one hand, attempts to base social science theories on the natural sciences continue[1]—with limited success—while on the other hand, the perspective is currently being reversed and the humanistic foundations of the natural sciences are being examined (see D'Avis 2019). However, it is not necessary to go so far as to look for social science foundations based on natural science principles, or even social science theories based on

1 For the example of sociobiology, see Richter 2005.

natural science assumptions. The method of adopting metaphors from scientific models and using them to describe and explain social phenomena is more successful because it has proven itself in scientific practice. This practice is still present, even if the adoption of metaphors and models is accompanied by a reflection on their usefulness and limitations as well as modifications with regard to their explanatory content. Suffice it to briefly illustrate this with three examples.

As early as 1908, Georg Simmel used the term "interactions" (Wechselwirkungen) to explain the "problem of sociology" (cf. Simmel 1908: 2). Simmel thus uses a term that gained importance in the physics of electromagnetism in the 19th century, but does so quite broadly and generally. He states that countless social phenomena bring about "that man enters into a togetherness, into an action for one another, with one another, against one another, into a correlation of states with others, i.e. exerts effects on them and receives effects from them. These interactions mean that the individual carriers of those instigating drives and purposes become a unity, a society." (ibid. 5).

Niklas Luhmann's adoption of a scientific model is much more consistent than that of Simmel. Strictly speaking, however, Luhmann does not adopt a metaphor or a model, but rather builds his systems theory on a general systems theory. This metatheory is interdisciplinary, so that it is not a matter of a theoretical foundation based on the natural sciences, but of participation in a common theoretical program. Luhmann recognizes that general systems theory offers explanations and theories that are useful for the further development of sociological theory. Below the level of general systems theory, Luhmann mentions machines, organisms, social systems, and psychological systems (Luhmann 1984: 16), which have common characteristics as objects of research.

Analytical sociology, which is currently emerging as a research program and is also receiving broad support in the newly founded "German Academy of Sociology," places explanation by means of social mechanisms at the center of its theory (cf. Schmid 2010: 31). This recurs to a concept that refers to the paradigm of mechanistics described above, which began with Hobbes and has undergone numerous manifestations in the natural sciences. As far as I can see, "analytical sociology" is not interested in the scientific history of the term and its effect as a metaphor that provokes a certain way of thinking, but with the reference to "causal inference" (Abell 2010: 207) it incorporates another scientific term into its theoretical program.

The list could be continued, but it is neither intended to criticize individual positions nor to be exhaustive. In fact, the three examples illustrate a close link between the natural and social sciences in the form of a lively exchange. In addition to this exchange—in the form of metaphors, terms, and concepts—questions about the production of scientific knowledge and the social and cultural embedding of researchers and scientific institutions have emerged in recent years. These questions are linked to the research interests of Science & Technology Studies (STS), as well as the sociology of knowledge and the sociology of science.

Undoubtedly, the methodology of the natural sciences has changed since the heyday of the mechanical paradigm, which can best be studied in the philosophical discussion of quantum mechanics (see Hieber 2023: 238). On the other hand, the social sciences have never really managed to catch up with new developments in the methodology of the natural sciences. It is not possible to deal with these issues in any greater detail here. This brief and fragmentary excursus on the methodological relationship between the natural and social sciences is important for at least two reasons. Firstly, the natural and social sciences are connected with each

other in many ways. A historical look at the relationship shows that the discussion about social science methodology cannot avoid dealing with the natural sciences. The promise of exact methods in the natural sciences is still regarded as a model; the scientific nature of the "strong sciences" differs from the social sciences not least in the exactness and reproducibility of their results. Despite extensive deconstruction efforts, the natural sciences continue to occupy a prominent position in the public and social consciousness. It is therefore not surprising that efforts to view the social as an object of the natural sciences are also becoming more frequent at present. The realization of increasing social and societal complexity leads to the scientific simulation and modulation of social phenomena and comes quite close to Auguste Comte's (1798–1857) idea of "social physics". Secondly, a compatibility problem persists, because even the new sociophysics does not succeed in solving the old problem that social phenomena do not have the same computability and predictability as natural phenomena. The social world is characterized by change. Unlike Hobbes, Baron de Montesquieu (1689–1755), for example, was critical of the mechanistic paradigm of early modern science. For him, there was no question that the social world was more than a game of atoms (cf. Weigand 1976: 4) and therefore had its own logic that could not be explained by mechanical laws. If we follow the explanations of the sociologist Dirk Baecker, 300 years after Montesquieu, then a similar conclusion can be drawn. According to Baecker, contemporary sociophysics fails to recognize the inherent logic of the social because it is interested in contagion and imitation, whereas sociology, as an expert on the social world, is interested in difference and ambivalence (cf. Baecker 2012: 113). In other words: Change, diversity, and difference stand in the way of explanations that aim for assimilation, approximation, and imitation.

The complicated and multi-layered relationship between the natural and social sciences is contrasted by the independence of the social sciences, which have undergone a process of differentiation into disciplines and sub-disciplines since their beginnings in the early modern period, but yet are based on common methodological premises. The long history of the modern social sciences, whose starting point can be found in the renaissance of political and social thought in Europe, not only promotes the self-confidence of the disciplines, but also raises the question of whether natural sciences such as physics can be used as a comparison at all. To put it bluntly, it could be said that just because the development in one field of the natural sciences took place in a certain way, this development does not necessarily have to be a point of reference for the social sciences. Perhaps this is why a different path is currently being taken in the social sciences, that of an empirical theory of science that builds on the research practice of the social sciences. This aspect is discussed in more detail below.

## *Philosophy and social sciences: Critical methodological relationships*

The relationship between the social sciences and philosophy is just as complicated and multi-layered as the methodological relationship between the natural and social sciences. The starting point here is the fact that numerous social sciences broke away from philosophy and strove for scientific autonomy. With regard to the above example of natural philosophy in the early modern period, this means that some of what used to be part of natural philosophy became independent social sciences after a lengthy and complex process of separation in the modern era, such as political science and sociology, which Auguste Comte (1798–1857) classified as "social physics" within natural philosophy.

According to tradition, the philosophy of science (Wissenschaftstheorie) is initially epistemology and therefore a matter for philosophy. With the scientific revolution at the latest, the question of the methodology, the "what" and "how" of the natural sciences moved to the center, resulting in numerous drafts on the logic, the possibilities of knowledge, and the cognitive purposes of natural research (cf. Fischer 2023a: 251f.). This focus within philosophy remains, because even today its methodological discussion is still primarily aimed at the natural sciences, with consequences for the relationship to the social sciences. On the one hand, this starting point of philosophical philosophy of science in the natural sciences and the simultaneous claim to provide an epistemological framework for all sciences is one reason for the comparison of the natural and social sciences. This comparison is flawed, as already noted above: social phenomena alone have different characteristics from natural phenomena, and there are also numerous different perspectives on the social world. Of course, the social sciences are also essentially based on rational and logically conclusive statements, but attempts to prescribe a certain deductive approach as the key to all sciences—as in the case of critical rationalism—are associated with similar problems as the attempt to subordinate social philosophy to the mechanical paradigm. On the other hand, the social sciences, which strive for autonomy, often see philosophical foundations and methodological drafts as interference. As established scientific disciplines sui generis, they have produced an independent methodological tradition that differs from the philosophical variety.

One consequence of comparing the social sciences with the natural sciences is the discussion about multi-paradigms. In fact, there is a historically evolved difference here, as the natural sciences are not familiar with the situation of several more or less equal paradigms existing side by

side, as is the case in the social sciences. This difference also has an impact on the philosophy of science. The philosopher of science Thomas S. Kuhn attributed the orientation towards a single paradigm to the nature of the "normal sciences". He states in this regard (1976: 38): "In no way is it the aim of normal science to find new phenomena; and indeed those that do not fit into a pigeonhole are often not seen at all. Nor do the sciences normally claim to find new theories, and often enough they are intolerant of those found by others. Instead, normal scientific research is geared towards clarifying the phenomena and theories already represented by the paradigm." Irrespective of the empirical verification of this assumption, the consequences are obvious: If normal sciences, i.e. natural sciences, are oriented solely towards one paradigm, then the social sciences are not a normal science, because they have more than one paradigm; in addition, there is a recurring need for new theories that react to social change or new empirical findings. The popularity of Kuhn's writing, which is unusual for a theory of science, therefore tends to cement the comparison between the two types of science and thus emphasize the subordination and "unscientific nature" of the social sciences. This, however, is not intended to say anything about the future development of social sciences and their status as normal science. Concurrent research in the philosophy of science has shown that the idea of a single paradigmatic science is somewhat misleading. The problem lies in the comparison of two varieties of science, both of which follow a specific logic.

In contrast to this is the realization, which has been spreading since the beginning of the 20th century, that the scientific system is becoming increasingly differentiated and that major methodological differences are establishing themselves in practice. A common bracket, such as the empirical sciences (empirische Wirklichkeitswissenschaften), is no longer sufficient for understanding. The distinction made by Wilhelm

Windelband (1894) and Heinrich Rickert (1926) between nomothetic and ideographic sciences, i.e. those sciences that are interested in legal knowledge and those that are more interested in describing (and explaining) a case, has become well known. Even if social science methodology is only significant in this discussion insofar as it is historically oriented, the limits of a methodology that encompasses all sciences are already apparent. Meanwhile, the state of multi-paradigmatism in the social sciences is a fact that cannot be ignored, which is due to different perspectives, traditions, and schools. Ultimately, it is the interest in knowledge that determines the methodological approach.

The sociologist Robert Merton has pointed out an essential reason that has determined the development of the social sciences and, in particular, the subject of sociology. While European philosophy in the 20th century was primarily characterized by large systems and blueprints, sociology was unable to find any guidance for empirical research in all-encompassing systems, or in the large blueprints that were created in sociology, such as those of Comte or Lester F. Ward (1841–1913) (cf. Merton 1968: 46). As a consequence of the orientation towards empirical sciences, however, the social sciences are dependent on forming a research design that can be derived from theories. In this regard, Merton proposes so-called "middle-range theories" (ibid.), which are located between working hypotheses on the one hand and grand theories on the other. The interest in knowledge thus differs not only from the classical philosophy of the 19th and 20th centuries, but also from the natural sciences, which aim for universal knowledge of laws.

It is therefore not surprising that an independent "Philosophy of Social Science" (POSS), which emerged late as a philosophical variation, receives little attention in the social sciences themselves (cf. Lohse 2015).

In contrast, an independent methodology is establishing itself at the latest with the academization[2] of social science subjects.

In fact, ontological assumptions, such as those discussed in philosophy, do not play a direct role in social science research, although social science explanations are often based on ontological assumptions. In other words, sociological theories also have general content-related suppositions. The Frankfurt sociologist Ritsert has explained in detail, using the examples of Émile Durkheim and Theodor Adorno, how sociological discourses essentially revolve around social ontological foundations (cf. Ritsert 2022: 17). However, this is where the overlaps between the practice of the two disciplines end. The methodology of the social sciences, for example, is little interested in questions of an a priori standpoint and instead focuses on the practice of research in relation to the underlying theoretical constructs. Thus, questions about the scope, knowledge gain, and limits of knowledge or the empirical application of theoretical approaches and paradigms can, but de facto, remain a case for specialists.

The development of an independent social science methodology has its limits. The increasingly pragmatic orientation of most social science

2 The so-called classics of the social sciences in particular were forced to answer methodological questions about the "what" and the "how," about the epistemological interest and the scope of their approach. Establishing a new discipline as an academic subject means not only differentiating it from established subjects, but also saying how the new appears and how it can be methodologically managed. However, such a methodological significance of the classics is hardly considered in the current discussion about a canon. For sociology, see the discussion on the online portal Soziopolis (2024). From a methodological perspective, shelving the classics of sociology would require these questions to be answered by other representatives. However, more theory of science remains a desideratum at present.

subjects has marginalized the philosophy of science, so that its content is barely visible in the subjects or in the university curriculum. While in the 1960s philosophy of science was understood—and often misunderstood—as a critique of the sciences, in recent decades it has been institutionally outsourced to university-wide centers. What can be interpreted as a bundling of competencies, in combination with the widespread tendency to streamline study programs, has led to a dwarfing of the teaching of science. In most degree-level social sciences curricula, the theory of science only has a place in the introductions to the subjects, then once again (for the last time) in methods training, where unfortunately methodology is often equated solely with methods issues.

The previous development in the relationship between philosophical and social science methodology is currently being counteracted by a change in perspective. Both in the field of philosophy and in the social sciences, there is an effort to take scientific practice as the basis for methodology. Take, for example, the philosopher of science Hoyningen-Huene, who takes a descriptive and comparative starting point for determining the question "What is science?". Here, too, the distinction between scientific knowledge and other forms of knowledge such as everyday knowledge is at the heart of his interest. The difference, according to Hoyningen-Huene, is the higher degree of systematicity inherent in scientific knowledge. There is no uniform structure of the sciences or scientific disciplines, but there is a "complex network of family resemblances" (2015: 227). The author identifies nine dimensions that can exhibit characteristics of a higher degree of systematicity (ibid.), such as descriptions, explanations, predictions, but also critical discourse and the presentation or publication of knowledge. Without going into the details of the approach, it is already clear that this focus can be understood as a further development of earlier or classical positions in the theory of

science, which, for example, relied solely on the logic of research methods. Hoyningen-Huene speaks of the abolition of such one-sidedness (cf. ibid. 228) in order to do justice to newer sciences, such as the social sciences and engineering (cf. Hoyningen-Huene 2008). The scientific nature of a discipline is not determined by methodology alone.

The sociologist Hubert Knoblauch has recently advocated a reflexive methodology, or in other words: an empirical theory of the social sciences in which actions and interactions, social and cultural practices, and communication are at the center of research interest. Knoblauch draws on the traditional line of ethnomethodology and the recent development of science and technology studies (STS), and designs an "ethnography of research" (2020: 252) that explores the question of the production and appropriation of scientific knowledge—the "social construction of scientific facts," so to speak (2021: 6). This is also associated with a normative claim, which is reflected in the search for the right or good research.

It is not necessary to speak already of a practical turn in the philosophy of science, but these two examples show that there are reasons to focus not only on the normative requirements, but also on the practice of science.

### *Knowledge society and competing knowledge systems*

Social sciences not only have society as their object of research; their practice (doing social sciences) also takes place in society. This fact forces us to take a closer look at the framework conditions under which the social sciences are practiced. Philosophy of science is required to consider not only the numerous social and societal phenomena that impose themselves on research, but also the direct or indirect social, political, economic, and cultural influences on research. Changing framework conditions for

research inevitably also change the opportunities and interests of research. Without going into these interactions in more detail, I would like to briefly discuss two key phenomena that have been preoccupying European societies for the past few decades. Firstly, the knowledge society and secondly, but related to this, competing or conflicting knowledge systems.

On the first point: knowledge society. Despite some criticism, the term knowledge society has become established as a description of the present (see Engelhardt/Kajetzke 2015). If we look at the phenomena associated with this diagnosis, several interconnected processes become apparent. As early as 20 years ago, Schulz-Schaeffer and Böschen identified three topics that shaped the discussion about the knowledge society (cf. 2003: 10f.) and are still relevant today. Firstly, the question of the possibilities and limits of the usability of scientific knowledge as a resource in non-scientific fields of action. Secondly, the question of the relationship between scientific knowledge and other forms of knowledge. And finally, there is the question of the transformation of internal scientific knowledge risks into risks of social modernization. While this point addresses the application of new technologies and scientific practices in society, the first two points are aimed at changes in the knowledge system of modern societies. This can be followed up.

Knowledge society can initially be interpreted in economic terms, namely as a shift in economic forces towards a knowledge-based guiding principle. In this understanding, the economy is no longer characterized by industry or services, but by knowledge work. It is precisely from this perspective that the first point mentioned above can be understood: scientifically produced knowledge is becoming increasingly important for economic purposes. Certainly, the effort to make scientific knowledge usable for other purposes is not a new phenomenon, but the striving for

the practical use of scientific knowledge is pushing its way into the university itself, so that one can speak of an economization of the scientific enterprise, in which the usefulness of knowledge is emphasized above all.

The question of the relationship between scientific knowledge and other forms of knowledge describes a process of de-differentiation between science and society that runs counter to the first finding. This means that science is increasingly losing its authority to interpret knowledge, and that other forms of knowledge are competing with it. One reason for this shift in knowledge systems can certainly be found in a recent structural change in the public sphere (see Seeliger/Sevignani 2021) and in the reconfiguration of the media landscape. At present, more and more voices can be heard criticizing the position of science as the sole producer of true, fact-based knowledge or questioning its credibility. Hostility towards science is primarily directed at research that deals with topics that are close to people's lives and can be applied to their own lifestyles. In other words, the criticism is directed primarily, but not only, at the social sciences.

Two examples from the recent past and present make this clear. Both the coronavirus pandemic and climate protection efforts have a strong social dimension, even if their findings are determined in medicine or the natural sciences. The social sciences also contribute to gaining an understanding of these phenomena and ultimately generate scientific knowledge. Both the pandemic and climate protection are social phenomena, as they only arise through the global networking of people, transport, states, organizations, and institutions. In both cases, the consequences of political and social measures informed by social science elicited not only justified and lively discussion, but in some cases also provoked a strict rejection and radical criticism (of the sciences) that extended beyond extremist circles. The ensuing debate focused on trust in

science, its credibility, and its role as an enlightener (cf. Bartels/Lehmkuhl 2022).

There is no need to go so far as to completely reject the knowledge order of modernity, but a gradual restructuring can hardly be overlooked. This becomes particularly apparent when one considers the pluralization of information media and the processes of knowledge appropriation. In this sense, science is being joined by knowledge that is anchored in everyday life and largely unreflected upon, which is reproduced in the media and shared by people with the same world view. This flip side of the knowledge society describes a process in which the importance of scientific knowledge is being pushed aside by the primacy of the lifeworld. This is not a completely new process either, but today the possibilities for disseminating and networking knowledge are of a completely new kind, so that in the digital world, for example, scientific knowledge and knowledge anchored in the real world appear to be on an equal footing. This also refers to the conflicting nature of knowledge systems, with the "rational man" of science facing the "emotional man" of the adventurized society (Erlebnisgesellschaft). In this context, the concept of the "neotribe" as a postmodern mechanism of group formation through similarities in attitude, opinion, and lifestyle proves to be appropriate (cf. Maffesoli 1996: 72). Information or claims that align with these shared views are circulated and amplified within the tribus, whereas opposing or inconvenient perspectives are excluded.

A few decades after its emergence, the knowledge society has thus created a dynamic of the knowledge order, but which is accompanied by an inner conflict—most clearly evident in the area of tension between the fields of scientific knowledge and everyday knowledge or worldviews (Weltanschauungen). This situation is challenging for the sciences in general and for the social sciences in particular. On the one hand, there is

hostility towards science, the politicization of knowledge and targeted disinformation, as well as increasing educational inequality (cf. Druckmann 2022), while on the other hand there are limited opportunities for education and influence. How can the social sciences respond to this problem? Politicizing the social sciences, as can currently be observed in the USA and Europe, cannot be the solution. Political partisanship not only makes the social sciences methodologically vulnerable, but also inherently limits their broader acceptance. In contrast, the path I briefly outlined in the previous section sees the need for self-reflection on one's own research activities as one way of responding to such challenges.

## *Reflexive philosophy of science*

So, what could be a reaction to the new conflictual knowledge order of the present and the associated problems of criticism, legitimization, and acceptance of social science knowledge? One possibility that avoids politicization and criticism of addressees is to reflect on one's own activity as a social scientist. Reflexivity therefore focuses on the research process as a whole.

What do we do when we conduct social sciences? A methodological reference to "doing social sciences" understood in this way takes a turn towards a reflective instrument with the aim of examining the quality of research—meaning, above all, its validity, scope, and the position of the researcher in the process. In this sense, the philosophy of science is not seen as an annoying and outdated legacy of philosophy, but is used as a reflexive instrument of one's own research practice. In this respect, the prejudice—that philosophy of science or methodology is nothing more than a theory of methods aimed at further developing highly specialized methods and applying them as precisely as possible—can also be eliminated. In contrast,

reflexive methodology in the sense of scientific theory aims to reflect on the individual, interlinked steps of the research process and thus not only on the methods, but also on the underlying theoretical assumptions, their operationalizations, and the empirically produced results.

This is based on the assumption that, from a sociological point of view, the appearance of the timelessness and general validity of the theory of science that has been created in philosophy is hardly tenable. In contrast, it has now been recognized that theories of science—like other theories—are dynamic. Perspectives on the theory of science are always determined by historical and social events and conditions, which forces research to keep even established methodological perspectives open to reflection. A twist in history shows that the very school that most strictly advocated formal logic and the claim of universal validity (keyword: unified science) is hardly an issue in large parts of the philosophy of science, but also in the social sciences. The influence of logical empiricism and the Vienna Circle (cf. Hahn/Carnap/Neurath 1929) is itself an example of the overstretching of a paradigm against the backdrop of a scientific success story.

On the other hand, a reflexive theory of science can also be integrated into the curriculum to accompany research activities, e.g. in student research projects or final theses. In fact, the research process always involves numerous methodological questions and problems, but these are rarely explicitly addressed and discussed.

If philosophy of science is understood as such an instrument and thus as a medium of quality assurance, then there are only a few role models that can be referred to. One of the few examples in which philosophy of science is used to examine the logical stringency of the research process can be found in the reflections on the "craft of sociology,"

which were developed at the end of the 1960s by Pierre Bourdieu, Jean-Claude Chameberodon, and Jean-Claude Passeron (1991) for teaching purposes, but also to establish an independent epistemology of the social sciences. I will conclude by explaining the idea of a reflexive theory of science in the social sciences in more detail with reference to Bourdieu et al. (1991):

> Reconstructing the process of research is one of the ways of checking the logical rigour of a piece of research, but it can have the opposite consequences when it is represented as a reflection of the real processes. It then helps to consecrate the dichotomy between the real operations, which are subject to intuition and change, and the ideal rigour that can more easily be actualized in formal exercises or the replication of surveys. (Bourdieu et al. 1991: 90)

The authors first make a distinction between a methodologically and formally correct application of empirical methods and a more comprehensive methodological reflection. In this sense, Bourdieu et. al., following Bachelard, speak of an epistemological break (cf. ibid. 69), which is helpful for sociology in order to ensure the differentiation of scientific knowledge from other forms of knowledge. In this sense, there is a contradiction between the knowledge order of science and the knowledge order of everyday life. Science must have different criteria, problems, and questions as well as a different interest in knowledge compared to those that arise in everyday life. What initially sounds like a matter of course becomes more relevant in the social sciences, which are not only integrated into the lifeworld of researchers, but also investigate social phenomena.

Doing science means more than just applying methods correctly; it combines several interrelated and interdependent levels in a research process. These levels include at least the theory or theoretical assumptions, their operationalization into researchable statements, the connection between theory and empirical investigation, the development of a research design, the empirical research process, as well as the result and the role of the researcher, who must be considered in all these levels. Bourdieu et al. speak of a dialectic of the scientific process, which is by no means completed with the formation of hypotheses, but includes empirical research and also takes into account an adequate understanding of its results (cf. ibid. 61). Routine and "automated thought processes" (ibid. 62) are particularly obstructive for social science research because they ignore phenomena that lie outside the measuring instrument and do not allow reflection on the methods themselves.

Reflexivity is not to be seen as a general way of thinking that focuses on everyday practice, but as a step-by-step direction for the whole research process in the social sciences. It involves the idea of making understandable and contestable the assumptions that guide research. The individual levels of the research process can be assigned to questions of scientific theory, which can ultimately help to make the idea of a reflexive methodology tangible. The following questions are exemplary and deliberately kept simple; the presentation follows the logic of the research process, and a later level can be related back to the previous one.

- For theory, or rather the theory paradigm:

What traditions underlie the theory or paradigm? What is the scope of the theory and where are its blind spots? What is its epistemological interest, what is the criticism based on it and how can it be countered?

- For operationalization:

How does the construction of a research object or modeling succeed? How can the connection between theory and research questions be ensured? How is a research question possible against this background?

- For empirical study:

What are the limits of the method, what can be seen with it, and what cannot? What is the connection between the theory and the chosen methods? How can coherent application be ensured without following automatisms?

- For results:

What are relevant data and what are not? How can the results be related to the question or to the research interest? What is the scope of the statements?

- For the role of the researcher:

What are the assumptions, ideologies, and expectations that flow into the research process and how should these be dealt with?

The benefit of a reflective theory of science lies in making transparent what is otherwise hidden and largely unsystematic. The possibility of understanding the research process as a whole helps both to improve quality by identifying limitations and potentials and also to present it to the outside world. Introducing greater reflexivity into the research process is certainly not a panacea for all the problems currently facing the social sciences, but it is a starting point and is best done in a group process of doing and thinking about research. Bourdieu et al. have exemplified such a reflexive theory of science in the context of curriculum development on the basis of classical social science studies. It is desirable to take this idea further and to methodologically review additional relevant studies that have emerged since the 1970s.

## Open review

The two reviewers, Lutz Hieber and Harald A. Mieg, provided written comments on the chapter. An online discussion took place on 23 August 2024. It was agreed to publish the main points of the review.

*Lutz Hieber*

The argument of the text focuses on an extraordinarily important aspect of sociological research by addressing the indispensability of reflexivity. It is useful to take as a starting point those sociological approaches that attempt to adopt methods from the 'exact sciences' in order to import them into sociology.

It should be noted that many of the sociological approaches mentioned by Fischer follow an approach that is long outdated in the natural sciences. Mechanistic thinking, the pursuit of 'objectivity' in the sense of Descartes' distinction between the cognizing subject (res cogitans) and the object to be cognized (res extensa), lost its validity in physics a century ago. The concepts of quantum mechanics take into account the mediation of subject and object by giving theoretical relevance to the chosen method of measurement.

For the list of requirements to be made, it should be taken into account that the imperative of reflexivity is, as Bourdieu states, tremendously difficult to put into practice. The reasons for this lie in the imprinting of educational processes. Bourdieu also assumes socialization processes. People who have been shaped by comparable educational processes have similarities in their "habitus." According to Bourdieu, the habitus can be understood as a system of internalized patterns that make it possible to produce all the typical thoughts, perceptions, and actions of

a culture—and only these (Pierre Bourdieu, Zur Soziologie der symbolischen Formen, Frankfurt/M 1974, p. 143). Pierre Bourdieu coined the term "cultural unconscious." The routines created by socialization in our educational institutions form the core of the cultural unconscious. For every person who has undergone such socialization stands in relation to the education he has acquired and ultimately consolidated in daily practice in a relationship that can be described as that of 'bearing' and 'being borne', because he is not aware that the education he possesses—possesses him (Bourdieu 1974, p. 120).

Here's a kind of academic anecdote: When I came to sociology from physics, I asked a group of colleagues a technical question the way a physicist asks a question, in one sentence. The group then spent 20 minutes discussing how the question was formulated. In physics, a question has to be short and clear. Only then can it be clearly defined and answered. In sociology it is different, precisely because we ourselves are part of the object of study. I would like to talk more about reflexivity and habitus.

Bourdieu demands reflexivity, which—in the later Manet lecture—he succinctly expresses by saying that the sociologist must carry out a double historicization: the historicization of the object, of the texts, documents, or objects he studies; and at the same time the historicization of his own concepts, instruments of thought, etc. (Pierre Bourdieu, Manet, Frankfurt/M 2015, p. 371). Such demands can hardly be met by one person or one research team. Bourdieu himself repeatedly brings his cultural unconscious, shaped by French high culture, into his studies. For two of Fischer's points in particular, it therefore seems necessary to point to possibilities of realization, without which Bourdieu's demand would only formulate a lofty goal that remains in the realm of ideas. These are, above all, the first point, which calls for a critical examination of existing

paradigms and the recognition of their blind spots; and the last point, which addresses the tacit assumptions and ideologies of researchers. Neither of these issues can be adequately addressed "off the cuff" by those conducting a project, as this would require something akin to self-analysis.

*Harald A. Mieg*

In your current presentation, the term "reflection" is still largely under-defined. When you talk about reflection in sociological research, are you talking about:

(i) an ongoing process of reflection during research?

(ii) observation and analysis of other researchers' sociological research?

(iii) a retrospective sociological analysis of one's own sociological research?

(iv) a philosophy of sociological research?

(v) discussing research and science with students?

This is probably not what you mean, but it also means reflection on research:

(vi) Reflection is a requirement for the discussion section in standard scientific papers.

(vii) Reflection—in the form of critical discussion—is an important element of scientific work and of quality assurance in the presentation of research results, e.g. at conferences.

In my opinion, your approach of a "reflexive philosophy of science" can be reduced to an ethnomethodology of research. Interesting research, but not a new approach.

One more remark. Heidegger once said: Science does not think. I would add: Science does not think, it works. Science today is professionalized; It must offer the possibility of earning a living. For the normal functioning of science, therefore, verifiable and evaluable routines are necessary, or: a professional socialization, so to speak, and—to paraphrase Hieber and Bourdieu—a scientific habitus. Too much reflection—*thinking*, in the Heideggerian sense—is rather disruptive, although reflection as a sociological habitus seems to already exist.

## *References*

Abell, Peter (2010). Singuläre Mechanismen und Bayessche Narrative. In: Thomas Kron / Thomas Grund (Hg.), Die Analytische Soziologie in der Diskussion. Wiesbaden: Springer VS.

Bartels, Andreas / Lehmkuhl, Dennis (2022) (Hg.). Weshalb auf die Wissenschaft hören? Antworten aus Philosophie und wissenschaftlicher Praxis. Wiesbaden: Springer VS.

Baecker, Dirk (2012). Ansteckung, und was man gegen sie tun kann. In: S. A. Jansen et al. (Hg.), Positive Distanz? Wiesbaden: Springer.

Bourdieu, Pierre / Chameberodon, Jean-Claude / Passeron, Jean-Claude (1991). The Craft of Sociology. Epistemological Preliminaries. Berlin / New York: De Gruyter.

D'Avis, Winfried (2019). Geisteswissenschaftliche Grundlagen der Naturwissenschaften: Eine Kritik des Szientismus. Weinheim: Beltz / Juventa.

Druckmann, James N. (2022). Threats to Science. Politicization, Misinformation and Inequalities. The Annals of the American Academy of Political and Social Science, 700(1), 8-24.

Engelhardt, Anina / Kajetzke, Laura (2015) (Hg.). Handbuch Wissensgesellschaft. Bielefeld: Transcript.

Fischer, Peter (2023a). Kosmos und Gesellschaft. Wissenssoziologische Untersuchungen zur Frühen Moderne. Weilerswist: Velbrück.

Fischer, Peter (2023b). Wissenschaftstheorie der Sozialwissenschaften. Bielefeld: UTB.

Hahn, Hans / Carnap, Rudolf / Neurath, Otto (1929). Wissenschaftliche Weltauffassung – Der Wiener Kreis. Wien: Artur Wolf Verlag.

Hieber, Lutz (2023). Eine Entwicklungslinie der Kritischen Theorie in Hannover: Oskar Negt. In: Martin Endreß / Stephan Moebius (Hg.), Zyklos 7 – Jahrbuch für Theorie und Geschichte der Soziologie. Wiesbaden: Springer VS.

Hobbes, Thomas (1912). The Metaphysical System. Elements of Philosophy Concerning the Body. Chicago: Open Court.

Hobbes, Thomas (1982). Leviathan. Stuttgart: Reclam.

Hoyningen-Huene, Paul (2008). Systematicity: The Nature of Science. Philosophia, 36(2), 167-180.

Hoyningen-Huene, Paul (2015). Précis zu Systematicity. The Nature of Science. Zeitschrift für philosophische Forschung, 69(2), 225-229.

Knoblauch, Hubert (2020). Relationale Phänomenologie, reflexive Methodologie und empirische Wissenschaftstheorie. Zur Kritik und Weiterführung der „Kommunikativen Konstruktion der Wirklichkeit". Zeitschrift für Qualitative Forschung, 21(2), 245-257.

Knoblauch, Hubert (2021). Reflexive Methodology and the Empirical Theory of Science. Historical Social Research, 46(2), 59-79.

Kuhn, Thomas (1976). The Structure of Scientific Revolutions. Chicago: The University of Chicago Press.

Lohse, Simon (2015). Pragmatism, Ontology and Philosophy of the Social Sciences in Practice. Philosophy of the Social Sciences, 47, 3-27.

Luhmann, Niklas (1984). Soziale Systeme. Grundriß einer allgemeinen Theorie. Frankfurt am Main: Suhrkamp.

Maffesoli, Michel (1996). The Time of the Tribes. London: Sage.

Merton, Robert K. (1968). Social Theory and Social Structure. New York: Free Press.

Richter, Dirk (2005). Das Scheitern der Biologisierung der Soziologie. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 57(3), 523-542.

Rickert, Heinrich (1926). Kulturwissenschaft und Naturwissenschaft. Tübingen: Mohr.

Ritsert, Jürgen (2022). Philosophische Erkenntnistheorie und die Grundlagen der Soziologie. Weinheim: Beltz / Juventa.

Schmid, Michael (2010). Mechanismische Erklärungen und die „Anatomie des Sozialen". Bemerkungen zum Forschungsprogramm der Analytischen Soziologie. In: Thomas Kron / Thomas Grund (Hg.), Die Analytische Soziologie in der Diskussion. Wiesbaden: Springer VS.

Schulz-Schaeffer, Ingo / Böschen, Stefan (2003). Wissenschaft in der Wissensgesellschaft. Wiesbaden: Westdeutscher Verlag.

Seeliger, Martin / Sevignani, Sebastian (2021) (Hg.). Ein neuer Strukturwandel der Öffentlichkeit? Leviathan Sonderband 37. Baden-Baden: Nomos.

Simmel, Georg (1908/2005). Soziologie. Untersuchungen über die Formen der Vergesellschaftung. Frankfurt am Main: Suhrkamp.

Soziopolis (2024). Ein (neuer) Kanon für die Soziologie? https://www.soziopolis.de/dossier/ein-neuer-kanon-fuer-die-soziologie.html (30.04.2024)

Weigand, Kurt (1976). Einleitung. In: Charles Louis de Secondat de Montesquieu, Vom Geist der Gesetze. Stuttgart: Reclam.

Windelband, Wilhelm (1894). Geschichte und Naturwissenschaften. Strassburg: Heitz und Mündel.

DOI 10.14279/depositonce-22434

Marie von Heyl, André Armbruster & Moritz Klenk

# Theorizing Through Podcasts?

Discussion

## Abstract

This chapter contains a discussion between three podcasters who started their podcasts to reflect on (theoretical) debates in science or to provide a science-oriented reflection. The approaches are very different, ranging from a theoretical dispute between two sociologists and self-reflexive, soliloquized thinking about scientific work, to the psychoanalytical concept of transference onto the audience as a mode of production. The changing relationship between producer/podcaster and audience is a specific point of discussion.

Marie von Heyl
Universität der Künste Berlin
Email: mail@marievonheyl.de

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory*. *Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

*Julian Hamann*: We now move to the panel discussion, which is concerned with podcasts as a specific medium of theorizing. It's my pleasure to introduce the three panelists. In alphabetical order:

*André Ambruster* is a sociologist from the University of Duisburg-Essen. Among other things, he works on religion, not least concentrating on various forms of violence and deviant behavior in religion or in the name of religion. Together with Robert Seyfert, André is co-editor of a book series called *Neue Soziologische Theorie* [New Sociological Theory], and also co-host of the podcast *Der Streit* [The Dispute]. In their podcast, André and Robert discuss recent sociological research, both books and papers, both empirical and theoretical, although leaning towards theoretical work. I will leave it to André to explain the antagonistic idea behind the podcast that explains its name, *Der Streit*. One last thing I'd like to emphasize is that your podcast always sticks very closely to one specific text; sometimes, texts are even discussed paragraph by paragraph. André will tell us more about the podcast later.

Next on the panel is *Marie von Heyl.* She's a Berlin-based theorist and artist who has studied art and philosophy in Stuttgart, Berlin, and London. Since 2020, Marie has produced the podcast *Eclectic Engineering*, in which she reflects on art, philosophy, psychoanalysis, and feminist theory. Not least, she reflects on dialogue and on the performativity of media. Some episodes include a guest, whereas others are more of a monologue. What I find striking about *Eclectic Engineering* is not only its scope and the different topics that are covered, but also the fact that classical references, authors, and texts are sometimes the implicit backdrop to the conversation, and sometimes (to greater or lesser extents) actively brought into the conversation and the discussion. I'm looking forward to Marie telling us more about her podcast.

I'd also like to welcome *Moritz Klenk*. He's a cultural scientist at the Mannheim University of Applied Sciences. Moritz has worked and studied in different disciplinary contexts, covering sociology, religious studies, and cultural theory. Moritz started to record a diary of his daily reflections on academic work, specifically on academic writing, speaking, and reading. And over time, the very form of these recordings—roughly translated as thinking through speaking, *sprechendes Denken*—became his object of research. So what started as an experiment became a podcast, an ongoing research interest; and not least, the podcast became his dissertation. Rest assured: In the end, he still had to publish the book. But it's surely an interesting switch of media between podcast and book, which I'm looking forward to hearing more about.

As I'm sure you'll agree, we could not have wished for better experts on podcasting as a medium of theorizing. And I'm especially happy that each of you brings their own perspective on podcasting, but probably also their own notion of how theory is related to podcasting. I would like to start with a short round, with each you telling us everything that I forgot or got wrong, and secondly, telling us why you decided to launch a podcast covering your scientific work and/or your scientific field in the first place. Who would like to start?

*von Heyl*: So why did I decide to launch my podcast? My background is in the arts, and there we have an ethics of production: we just start. I noticed that, in terms of thinking, I was most productive in conversations, when I was talking to people. The podcast was launched as an attempt to keep some of that, some of the insights produced in those conversations—albeit knowing that once you record something or document it, you do change it.

But I noticed that there was something else at work in these dialogues, something I today would call the *eros* of conversations. I noticed that sometimes both parties are amazed because they feel something is happening here. It's almost like ... I don't know ... you're perplexed, you're intoxicated, you feel almost high on this conversation. And aren't these the most productive ones? I wanted to describe this exact dynamic more precisely, so the podcast turned into a PhD and now I'm looking at what is actually happening here. My research questions for my PhD came to me through the podcast, as it were.

*Armbruster*: The starting point was the COVID pandemic. My colleague Robert Seyfert (who can't be here today) had adjacent offices. But when we each had to work from home during the pandemic, the debate stopped. So we were looking for something to start to talk again. And the second starting point was: When you're reading reviews in sociology, they are all quite boring. Because the text, the book is always okay, but not citing enough work and there's no passion in reviewing or debating text. So we thought we would start a podcast to talk about recent literature in sociology, meaning not older than four or five years.

We then looked for books to review; the texts that we use as a basis are more or less theory-driven. We wondered how the debate could be made more passionate, more lively. And we came up with the idea of *different but opposing roles*. One role is the supporter, advocate, or promoter, who loves the text; He's a big fan; It's the best text ever written. The other role is antagonistic, the critic. He hates the text, he hates the argument, he hates the structure, the conclusion, everything. So through these roles, this opposition, we hope to have a dynamic argument, to have some passion. And we discuss the text from these respective roles, both trying to convince each other. So when I'm the critic, Robert is the supporter, I have to try to convince him that this is the worst text that is

ever written, whereas he takes the opposing position. So we have a debate on the pro and cons.

Beforehand, when Julian asked us how the podcast relates to theorizing, a quote by Richard Swedberg came to me: "Theory is the end product and theorizing is a process to theory." And of course this is right, but on the other hand I think Swedberg is wrong because theory is not an end product. You have to debate on theory, you have to think about theory. And the podcast, we hope, is something that we want to engage with theory. It's some kind of ongoing theory work or theorizing that we criticize, evaluate, see what are the upsides of the book or the text, what are the blind spots or dead ends. And so we just mainly talk about books for sociologists.

*Hamann*: Marie mentioned the notion of trying to capture the *eros* or the energy of a conversation. André, is this antagonistic setup of *Der Streit* maybe a mechanism to enforce a certain kind of energy in the conversation?

*Armbruster*: I think so, because when you start the conversation, your role specifies whether you either hate the text or love it. And then you have to try to convince the other. And I think it's a kind of mechanism to have a dynamic, to have some lively debate. I'm not sure if it's *eros*, maybe *illusio*, you're a fan. It's some kind of energy.

*Hamann*: And how do you decide who takes which role?

*Armbruster*: Who's the opponent, who's the fan of the text? We flip coins.

*Klenk*: I started doing podcasts in 2015, I think, and podcasting became a bigger and bigger part and a passion for me. So it took too much time, and kind of got in the way of my PhD, which at the time was in sociology. And so I faced the prospect of quitting one or the other. But then I decided:

No, wait a minute, I could just start a podcast talking about my work, and using it for doing *it* and getting through this—all the conferences, and the work, and everyday practices, and the meetings, and all of that. And I thought: OK, I will do this as a daily podcast called *Podlog*, Podcast Logbook, so to speak.

And then, as was already mentioned: A few weeks in, I realized that something fascinating is happening here. At least it was fascinating to me, because I had never experienced this kind of very frequent talking to myself. I was not famous for talking to myself ... as some may be, but I was not at the time. But when you do this every day—and you have nothing but the things you come up with when you press [*Record*] (since I don't prepare anything)—then you start realizing that you are in a conversation with yourself. It's not a monologue—or at least I would not think of it as a monologue—but rather a dialogue or a *soliloquized* thinking.

And what you find is that it is this conversation, or the dialogue itself, that brings in rather different and surprising elements of life into the work; Because you have to talk about it, you have to make sense of it. And you find—or at least this is what fascinated me—that it works differently than writing, or taking notes, or anything else.

- You have to, for example, reiterate everything.
- You have to re-mention it.
- You have to speak it out loud again.
- You cannot rely on notes.
- You cannot just flip through the pages of what you've already said.
- You have to repeat it, actually say it out loud.

And this involves a different way or style of thinking, which is connected or related to this. And that was what fascinated me. So the different aesthetics, in a way, how you think, or actually *do* "thinking" in a

conversation. And, well, I think of the widespread disregard for *conversations* in academia. Because you don't mention them, besides obscure papers that are now published. But besides them, no-one mentions that there *are* conversations. But that is mainly the thing we are *doing* in science, and in arts and humanities in particular. And this must find a formal medium to express, to be expressed, or to even be developed—if that might be something that some people are interested in.

More recently—like the last four years or something—I just used podcasting for teaching, since I became professor and had to—especially during the pandemic, for obvious reasons (of enforced distance learning). But I had to, and I was *prepared* for it. I was lucky in that I was prepared through the experience of podcasting. So I think it's about the *aesthetics of talking*; of the conversation—even with yourself (in which you become someone else, in a way): You have to treat yourself as a partner or someone, and you cannot pretend roles, for example, or get in conflict with yourself. You have to find a dance with yourself, in a way, which would be a different model (of discourse), I think.

*von Heyl*: I think it's interesting to see what feels different when you record a conversation for the podcast, versus when you have a conversation with yourself. I do script and record monologues as well ... I hesitantly call them monologues. On the one hand, it's just me who's talking, but on the other hand I'm still talking to *someone*: the audience. And I think that this productive dynamic—call it *eros*, or attention, or energy, or whatever—is that you're actually always talking to *someone*. What I like about podcasting is that it's open-ended. It's casual. Whom do I produce for? I would say *to whom it may concern*. I'm just putting it out there. And I have no idea who's listening. I see numbers, statistics ... how many are listening. But I have no idea *who's* listening. I wonder who am I addressing when I'm talking, when I'm putting these things out there. Especially in my

monologues. It's difficult to describe because it's not a specific person I'm talking to, but it's very important that I'm talking to *someone*.

In psychoanalysis, we have the concept of transference, *Übertragung* in German, which means that you believe the analyst knows more than you do. Lacan would describe such a person as "the subject that is supposed to know". In doing so, you overestimate their knowledge and their expertise. I think this concept holds outside of the psychoanalytic setting as well. A transference to the audience, as it were. You suppose they always know more than you do, even in your field of expertise. Of course, this is a fiction. So you have to produce the knowledge that you think they have, they allegedly have. You have to produce it yourself. And I think that's why this open address "to whom it may concern" can be so incredibly productive. Transference can be at work in an actual conversation, or a conversation with yourself, or it can be the audience.

*Hamann*: What I find interesting is that two of you have turned their podcasts into a PhD or are in the process of doing so. Does this imagined audience change once you decide that the podcast is also going to be a dissertation? Does the supervisor become an imagined audience or does a part of the audience become the supervisor? And how does this affect the theorizing or your thinking more generally?

*von Heyl*: As for me, and I think this is true for Moritz as well, I did not launch the podcast thinking that this project would turn into a PhD. I just casually started, and only later, in the middle of it, realized that I had created a form that actually turned out to be quite productive. And some of this innocent liberty, I could maintain. It is not that academic standards did not hold any more, but the imaginary nagging voices were tuned down. I have this incredible freedom; I incorporate stuff, anecdotes,

material that I wouldn't bring into a PhD. You know, material from the fringes; the everyday life becomes part of my theory.

*Klenk*: Just to add to this: For me, it was already under way when someone else asked me: Why isn't this your PhD? So I thought: Right, why not? I could at least try. But, as you said: Rest assured, I still had to write a book. Yet the main part was the podcast itself, about 150 hours of conversation with myself. And it was just the appendix to the PhD dissertation that I published later as a book. So it can be the other way around, but it was late in the project that it became this. And if you listen back to the episodes of these daily recordings, then you couldn't imagine in a million years that this might be a PhD project, because sometimes I was ranting about, I don't know, people that annoyed me or theories that annoyed me.

One of the most downloaded episodes is where I criticize Armin Nassehi[1] because I got really angry about his conversation with this Nazi fellow. And so this is not something you can normally imagine as part of a PhD project, but it became one. And I think this is important because I consider it as a methodological work of how to study, how to ask how thinking is done in this way, and how theorizing is done in this way, and how the life comes into this whole context, and what else is relevant and what is not, and where you draw the line. You don't have to draw the line the usual way when you do it like this.

*Hamann*: Speaking of audiences: André, would you say that your podcast is more directly oriented towards an audience, however vague or fuzzy it may be? And how do you imagine your audience to be and what role does it play for theorizing in your podcast?

1 Armin Nassehi is Chair of General Sociology and Theory of Society at Ludwig Maximilian University, Munich.

*Armbruster*: While recording the podcast, we don't think about the audience. There's so much trouble to convince the other one, in order to win the argument, that you don't have capacity to think about who might listen to this. And then when the podcast is finished and published, then it's done. I don't know how to describe this, but we never listen to it again. Because then, you think (or I might think): Why didn't I raise a particular argument at this time? So I don't know ... maybe it's strange, but the podcast is a finished product or every episode is a finished product. And then we let it go.

So I just wanted to know, how is it to listen to yourself again? I think you've done that. You too?

*von Heyl*: I edit, but never later.

*Armbruster*: Okay, because I think I couldn't stand to hear it again. How is it to read it again? Do you edit it again?

*Klenk*: This is also an aspect I forgot to mention. Actually, one of the most important technological settings of this whole project was that I was always talking to myself using a microphone and headphones with direct monitoring. So I was *listening to myself,* which kind of preserved the dialogue situation because I was listening to my own voice. And this is a very relevant part of it, because then you stumble over your own words, not only the mouth feel, but the corporeality of the work you're doing, of talking, of language itself, and you stumble and you get irritated just because of the sound of the words ... it doesn't *fit* right. And you ask why, and you wonder, and you stutter. And that's the most important aspect, I think, for doing that kind of work. And listening to myself was just, in this sense, like an ongoing process all the time.

And after that, I just edited very minimalistically; only if there was any noise or something, I edited it out. But I had to listen to it again. And

then there was a third step, which was finding chapter marks if necessary, which I did. And a fourth step was using a picture that I took on that day. That was the condition. Any picture. Sometimes it was a picture of my cat at the time, sitting on a book, if it fits. And then put it online. And then sometimes you even get reactions or comments or emails. But I think for the dialogue, experimental system, working with dialogues as a medium of doing theoretical work: For this, it is really important—at least it was for me—to listen to oneself.

*von Heyl*: I listen too, when I edit the episode. In the beginning, it was very strange. I edit an episode and then I put it online and then I never listen to it again. And I think that's quite a relief. You know it's out there. And I get what you're saying: That you think about what you could have said. Or worse: Later you find out you made a mistake. I mean, this happens. Oh my God, I got it wrong. It might be embarrassing, but it's all part of it. And maybe that's the beauty of the podcast as well, because, you know, it's a time-based medium. You present a *glimpse of your thoughts at that very moment*. It's fine to make mistakes. And since the format suggests that more is coming, you're sort of relieved of that burden of settling it once and for all. You know, it's not set in stone. For me, this is good to know: I did say this, I did think this, it's out there, and I don't have to listen to it again. I always work like this, in my artistic practice as well. I have a very strange relationship to my previous work. But I think maybe that's why you have to produce more.

*Hamann*: Marie, André, Moritz, thank you very much for your insights and the interesting discussion!

## Follow-up discussion with Markus Kip (sociologist, Berlin), co-editor of the "Urban Political" podcast

*Markus Kip*: What struck me most was that none of the podcasts paid much attention to the question of the audience. I got the impression that podcasting is seen primarily as an end in itself and as a way of understanding oneself, rather than as a way of communicating with an audience that is explicitly taken into account during production. As a result, the feedback loops from the audience were not given much importance.

*von Heyl*: To what extent is podcasting about communicating with an audience? In my understanding, we have to distinguish between the empirical other as a concrete counterpart and the abstract other as the condition of possibility for production. In my work, I am working from an extended concept of conversation that has addressing the other as a basic premise. The psychoanalytic concept of transference, as mentioned earlier, assumes that I am talking to *someone*, even if I am not communicating with a specific person. In this respect, the relationship to the audience in podcasting is similar to that between an author and a reader. My production is not a monologue. Without an audience, I would not simply *publish* nothing, I would not *produce* at all.

*Klenk*: I would respond to the question on the supposed lack of audience focus (in our discussion as well as in the podcasts in question) in two ways.

In regard to the *Podlog* project, I think I have to disagree: The podcast places the question of feedback loops and the relationship between speaking and listening at its very core. However, for a podcast explicitly designed as a soliloquy, it functions differently. In fact, the entire project could also be understood as an exploration of the feedback loops between speaking and listening, implicitly raising the question of the audience of a *soliloquy*.

Furthermore, and by this I try to address a more general issue, the question itself seems to express a rather conventional understanding of production—experts on one side and their passive audience on the other—essentially, a traditional mass-media perspective that compares podcast production to radio. Yet, Brecht's radio theory already challenged this, arguing for the transformation of radio into a communication apparatus. In my opinion, those who still see podcasters as merely broadcasters, and listeners as merely passive audiences, do not grasp the essential characteristics of podcasting. Often, podcasting itself serves as a response to other podcasts, a reaction, provocation, question, or answer in an asynchronous conversation—whether or not there is an explicit audience or direct listener engagement. The counter question would perhaps be: Would the lack of audience reference also be missed in written scientific texts; And if not, why not?

In short, the traditional relationship between media professionals and audiences has not yet been fully dissolved in the podcasting medium (if only because professional media producers, particularly from broadcasting, have heavily appropriated the medium and benefit from networks that grant them interpretive authority—down to the definition of 'podcast' itself. They remain existentially dependent on their status as experts, and thus on an audience as passive consumers rather than self-producing listeners). At the same time, however, the standard media expectations do not simply continue to function unchallenged in podcasting. The medium inherently *questions the sender/receiver logic* and the division between producers and their audiences. The fact that this issue was not explicitly addressed in this context is, I would argue, the strongest indication of this underlying shift, regardless of whether the distinction between producers and audiences will ever completely dissolve.

# 2. The Philosophical View on Theory

DOI 10.14279/depositonce-22435

ERIK J. OLSSON

# Definitions as Explications and the Explanatory Role of Knowledge

## Abstract

What is a good definition of a philosophical or scientific concept? In this paper I will argue that definitions can fruitfully be seen as explications, essentially in the sense of the logical empiricism of Rudolf Carnap, whose mature view on the matter was detailed in a 1950 book on the logical foundations of probability (Carnap, 1950). Being an epistemologist, my example will be the concept of knowledge, but I believe that much of what I say about definitions applies to other concepts as well. The second part of this paper is more specifically epistemological: it concerns the prospects of subsuming various approaches to epistemology under a methodology which relies on Carnap's method of explication—in particular, on approaches that emphasize the explanatory role of knowledge. Hence, in a sense, this paper develops a metatheory of epistemology—a theory of the theory of knowledge.

Prof. Erik J. Olsson
Lund University, Department of Philosophy
Email: erik_j.olsson@fil.lu.se

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023.* Berlin Universities Publishing.

## *Introduction*

The question of the nature of knowledge has in the analytical tradition been understood as the question of how the concept of knowledge can be defined. There are a number of different types of definitions in circulation. A lexicon definition tries to capture what people mean. An ostensive definition attempts to explain a concept by pointing to a representative example. A stipulative definition decides the matter by simply ruling that the term in question should henceforth mean this or that. However, the dominant view in analytical philosophy has been, and probably still is, that the kind of definition we are looking for is conceptual analysis. Just as the concept of bachelor can be analyzed as "unmarried man," so too can the concept of knowledge be analyzed, it is thought, in terms of other, simpler concepts.

We seek, then, a definition of knowledge that has the following form:

*S knows that p if and only if ... (add conditions here!)*

We can test the adequacy of this attempt at conceptual analysis by asking ourselves the following questions: Is every case where all conditions are satisfied also a case of knowledge? Is every case of knowledge also such that all conditions are satisfied? If both questions can be answered in the affirmative, we have found our sought-after analysis.

So, what are the conditions? In contemporary philosophy, the most common conceptual analysis of knowledge equates the latter with "true, well-founded conviction" or "justified, true belief"—the so-called JTB analysis. If you know something, you must believe it, or be convinced that it is true. Moreover, if it isn't true, you don't know it. Finally, someone may have a true belief purely by luck, in which case the person doesn't know it. Hence, in a sense, a person needs to have some reason for believing the thing in question to be true. More rigorously:

*S knows that p if and only if (i) p is true, (ii) S believes that p is true, and (iii) S's belief is justified.*

A variation, called reliabilism, emphasizes reliable belief-production rather than justification:

*S knows that p if and only if (i) p is true, (ii) S believes that p is true, and (iii) S's belief was acquired through a reliable process.*

Providing an analysis of the concept of knowledge, along the lines of JTB or reliabilism, is a separate task from arguing that there *is* knowledge, that we can actually know—contrary to what skeptics have claimed since antiquity. Relatedly, giving an analysis of knowledge is a separate task from giving a method for making sure that the conditions for knowledge are satisfied. Hence, having "truth" as part of the concept of knowledge does not imply that we have provided a method for finding the truth. To take an analogy: defining gold as an element with a certain atomic number is one thing; providing a useful method for making sure that a given piece of metal is gold is quite another.

Nevertheless, it might seem provocative from certain perspectives that have gained popularity to define knowledge in a way that makes knowledge imply truth, in an objective sense of "truth." There is a long tradition in sociology of regarding such claims with suspicion. In the sociology of knowledge research program, knowledge is thought of as socially determined or even constructed, for instance by one's social position. Claims of objectivity are regarded as instruments employed by the ruling class to secure its power. Influential thinkers in this tradition include Karl Mannheim and Michel Foucault. However, there is not necessarily any deep opposition here, as sociology of knowledge can be thought of as studying—not knowledge *per se*—but what is *regarded* as knowledge by a particular social group. Moreover, few people manage to

go about their daily life without using knowledge in the objective sense. Surely, even sociologists of knowledge think they know where they put the car keys, how much their house is worth on the market, what kinds of food they are allergic to, and so on. They think that they know these things objectively, not merely as social constructions. Harry Frankfurt famously made similar points about our concern for truth (Frankfurt, 2006; cf. Olsson, 2008). Finally, in the scientific realm, few people regard *all* such knowledge as merely socially constructed. If they did, they would consider climate scientists' claims about global warming to be mere social constructions entirely unconnected to an independent reality. Surely, knowledge of climate change is widely considered to be objective knowledge, otherwise we wouldn't care about it to the extent that we do.

Even though it is difficult to question the role played by the concept of knowledge in our daily and scientific endeavors, there is a powerful argument against the JTB definition—which, arguably, hits the reliabilist conception as well: they fall prey to the Gettier problem, named after its discoverer, Edmund Gettier (Gettier, 1963). Consider the following propositions:

(i) Jones owns a Ford (justifiably believed)

(ii) Brown is in Barcelona (only imagined, neither believed nor justified)

(iii) Jones owns a Ford *or* Brown is in Barcelona (justifiably deduced from (i) and therefore believed)

But as the story continues, (i) is false—Jones does not own a Ford after all; all evidence to the contrary. Yet (ii) turns out to be true, whence (iii) is still true. Since (iii) is true, believed and justified it constitutes knowledge on the JTB analysis. But it seems that (iii) is true "only by coincidence." Most epistemologists have concluded, therefore, that JTB wrongly classifies (iii) as knowledge.

There are many similar examples intended to show that justified, true belief is not enough for knowledge to be present. A common theme is that an (epistemically) unfortunate circumstance jeopardizes the justification that the person has for the main claim, but the situation is saved because the claim turns out to be true nonetheless. Such examples can be constructed if the justification the person has does not necessitate the truth of the thus justified proposition. Thus, what is required is a fallibilist concept of justification according to which justification does necessitate the truth of the thus justified proposition. The problem with an infallibilist conception of justification is that preciously little comes out as justified—and known.

The famous Gettier problem can be viewed as the problem of finding necessary and sufficient conditions for knowledge that preclude cases like the above from becoming knowledge on a fallibilist account of justification. Perhaps the first thought that comes to mind is that the main claim in the Brown in Barcelona case is based on a false premise, namely (i) that Brown owns a Ford. Why not simply add as an extra condition in the definition of knowledge that knowledge must not be based on a false premise? Unfortunately, one can construct "Gettier examples" that do not involve reasoning from a false premise, the most famous being the so-called Barn facade example of Alvin Goldman (1986).

A number of alternative definitions of knowledge have been presented as solutions to the Gettier problem. According to the indefeasibility theory, it is not enough that a person's true belief is justified; his or her justification must also be robust in the face of true information (such as the information that Jones does not own a Ford). Less conservative departures from the JTB analyses require consideration of relevant alternatives to the present situation or of contrafactual situations—situations that could have materialized but didn't. Even more

radically, some have argued that we should forget about justification as a condition for knowledge, and instead view knowledge as simply "true belief." What knowing is all about, in this view, is "getting things right."

Another alternative is to question standard methodology. An example is Hilary Kornblith's suggestion that epistemology should focus not on the *concept* of knowledge, but on knowledge itself, as a phenomenon existing in the natural world (Kornblith, 2002). Some have even questioned the very idea of defining knowledge in terms of other—presumably more basic—concepts, such as truth and belief. Timothy Williamson has argued that knowledge should rather be seen as an undefined, "primitive" concept (Williamson, 2000). Knowledge can, he maintains, be used to define or elucidate other concepts, such as evidence and assertion, but it cannot itself be defined in terms of something else. Both Kornblith and Williamson highlight the explanatory value of knowledge. I will return to their theories in a moment.

A common deficit in these different reactions to the Gettier problem is their lack of an independent justification: they were typically devised with reference and in response to the Gettier problem. I now turn to a general methodology of definitions, against which no similar accusation can be made—its formulation precedes Gettier's 1963 paper by more than a decade. I will argue that this methodology, nonetheless, when applied to the Gettier problem, suffices to take the sting out of it.

## *Carnap on explication*

By explication "we mean the transformation of an inexact, prescientific concept, the explicandum, into a new exact concept, the explicatum," Rudolf Carnap wrote (Carnap, 1950, p. 3). He identified two steps in the process of providing an explication. The first step amounts to properly elucidating the explicandum. What is the more specific intuitive concept

to be explicated? Answering this question does not necessarily involve producing an outright definition or analysis; often, a rough characterization or representative example is sufficient. The specification of the explicatum is the second step. What new concept—the explicatum—is to replace the explicandum in relevant contexts?

An explicatum should satisfy the following conditions as well as possible (Carnap, 1950, p. 7): It should be similar to the explicandum, exact, fruitful, and simple. Specifically:

1. The explicatum [the thing that explicates] is to be *similar to the explicandum* [the thing that is explicated] in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted.
2. The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an *exact* form, so as to introduce the explicatum into a well-connected system of scientific concepts.
3. The explicatum is to be a fruitful concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept; logical theorems in the case of a logical concept).
4. The explicatum should be as simple as possible; this means as simple as the more important requirements (1), (2), and (3) permit.

One of Carnap's illustrations is the replacement of the traditional concept of fish by the artificial concept of *Piscis* in zoology (Carnap, 1950). The latter, which defines fish as a certain kind of gill-bearing vertebrate, excludes several kinds of animal that were subsumed under the concept of fish, e.g., whales and seals. Nevertheless, most kinds of animal that were previously classified as fish are also *Piscis*, indicating that much

of the original meaning has been preserved. The new concept is not only exact, but also more fruitful: it allows for the formulation of a greater number of interesting general truths, such as that that all *Piscis* are cold-blooded.

A recent example along similar lines is the 2006 redefinition of the concept of planet by the International Astronomical Union (IAU) (Cordes & Siegwart, 2019). Within our Solar System, a planet is "a celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighbourhood around its orbit" (IAU, 2006). The new definition excludes Pluto, but incorporates key aspects of the earlier use patterns, while being clearer and more fruitful (cf. Murzi, 2007). Frege's definition of a natural number and Tarski's definition of truth are some examples of explications in the realm of logic.

## *Explicationist epistemology*

Let us refer to a philosophical method based on Carnap's methodology of explication as *explicationist philosophy*. Explicationist philosophy should be understood as implying that all four requirements on an explicatum be given substantial positive weight (but simplicity less so). By *explicationist epistemology* we will mean the corresponding methodology applied to the problems of epistemology. Among the well-known proponents of this approach to philosophy and epistemology we find Hempel (1952), Quine (1960), Lehrer (1990), Maher (2007), and Baumann (2016).

Carnap's account of fruitfulness seems rather restrictive. On a broader account, any improvement of a theory occasioned by the addition of a concept would count in favor of the fruitfulness of the latter (not only improvement in the system of laws). Thus, we may distinguish between

narrow ("nomological," "theorem-oriented") and broad ("holistic") fruitfulness and, correspondingly, between a narrow or broad explicationist methodology

Similarly, we may recognize other kinds of concept than just logical or empirical ones, e.g. legal or ethical concepts.

Crucially, for our purposes, explicationist epistemology is essentially immune to the Gettier problem (Olsson, 2015). Suppose that someone proposes an explication of knowledge along the following lines:

*(K) S knows that p if and only if C*

where C is some possibly complex condition. Suppose C entails, presumably wrongly, that people do know in Gettier cases. To see that it does not follow that (K) is not a good explication of knowledge, consider again the first Carnapian desideratum:

> The explicatum [the thing that explicates] is to be similar to the explicandum [the thing that is explicated] in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted.

For the Gettier problem to be a threat to the claim that (K) satisfies this desideratum, it would have to imply that it is not true that, in most cases in which the ordinary concept of knowledge has so far been used, the condition C in (K) can be used instead (Olsson, 2015).

But it doesn't show this; Gettier cases are too rare: they involve the consecutive occurrence of two improbable events. It follows that the Gettier problem is at least not a knock-down argument against a given explication of knowledge; much depends on how well the other requirements on an explication are satisfied. The interested reader may

want to check out Olsson (2015), where I argue that the reliabilist analysis of knowledge satisfies all four conditions on an explication of knowledge.

I have argued (in Olsson, 2017) that explicationist epistemology promises to have unificatory benefits, since many other methodologies can be seen as sub-methodologies within an explicationist framework. For instance, conceptual analysis, ordinary language philosophy Oxford-style, and experimental epistemology can all be useful in the first step of the explication process: the elucidation of the explicandum. But what about Kornblith's natural kind and Williamson's "knowledge first" approaches? Can they, too, be seen viewed as part of a more general, and arguably less idiosyncratic, explicationist approach, rather than as competing methodological perspectives? If so, they may add valuable insights regarding the explanatory role of knowledge, but do so not separately—as stand-alone theoretical vehicles—but as embedded ingredients in a broader explicationist outlook. In the following I rely on the exposition in Olsson (2022), which the reader is referred to for more details.

Kornblith (2002, pp. 61–62) explains his natural kind theory of knowledge as follows:

> I want to claim that knowledge is, in fact, a natural kind. [...] I take natural kinds to be homeostatically clustered properties, properties that are mutually supporting and reinforcing in the face of external change. [...] The knowledge that members of a species embody is the locus of a homeostatic cluster of properties; true beliefs that are reliably produced, that are instrumental in the production of behavior successful in meeting biological needs and thereby implicated in the Darwinian explanation of the selective retention of traits.

For Kornblith, we recall, what is related to biologically interesting properties, and therefore important, is knowledge itself, as a phenomenon in the empirical world, not the concept of knowledge. For Carnap, by contrast, fruitfulness is a property of our concepts: it is the property a concept has if it figures in many true lawlike generalizations. But this looks like a mere verbal difference. Surely, everything Kornblith says about the nomological importance of (reliabilist) knowledge in the natural world can be translated into claims about the fruitfulness of the concept of (reliabilist) knowledge in biological theory. Generally (Olsson, 2022):

> (Phenomenon–Concept Bridge Principle) A phenomenon X is important in the sense of being an important part of the portion of reality belonging to scientific domain Y just in case the concept of X is fruitful (in Carnap's sense) in the true account of Y.

Given the Bridge Principle, Kornblith's many insights about the importance of reliabilist knowledge as a phenomenon occurring in the natural world can be translated into statements about the fruitfulness of the concept of reliabilist knowledge. Incidentally, the translation of natural kind epistemology into explicationist epistemology also solves the "problem of access" for Kornblith, a problem raised by Alvin Goldman (2005): How do we identify what natural kind to pick out as knowledge without paying attention to, or indeed analyzing, the concept of knowledge? As we saw, the method of explication does not require the provision of a conceptual analysis of the intuitive concept to be explicated; a crude characterization of suitable examples will often do. Kornblith makes a similar remark in his reply to Goldman (Kornblith, 2005), but the lack of an independently justified account from which this response follows makes it less convincing.

Turning to Williamson, he concludes (based on the Gettier problem and other considerations that I won't go into here) that knowledge cannot be analyzed in terms of other concepts. He then proposes to take knowledge as a primitive, unanalyzed concept and to explore its explanatory role. In fact, knowledge, according to Williamson, plays two explanatory roles. First, knowledge figures in nomological laws, relating it (statistically) to things like stability of true belief. Second, knowledge explains, non-nomologically, what our evidence is and when we are allowed to make assertions. To be specific, the evidence we possess at a given point in time amounts to the knowledge we have at that time and, furthermore, in order to be in a position to assert something, we must know that thing to be true. Throughout his discussion, Williamson predictably refuses to give a conceptual analysis of knowledge in terms of necessary and sufficient conditions. The closest he gets to providing such an analysis is his "characterization" of knowledge as a "most general factive mental state". Examples of factive mental states include John's observation that the clock is ticking or Jane's recollection that she had scrambled eggs for breakfast. These mental states are factive in the sense that, if they obtain, the propositional content is true. Knowledge is a most general factive mental state in the sense that it is a factive mental state that is necessarily implied by all other factive mental states. Thus, John's observing that the clock is ticking implies John's knowing that the clock is ticking.

Now, even if Williamson's characterization of knowledge as a most general factive mental state is not an outright definition of knowledge, it can still be an explication. Consider once more the second Carnapian desideratum (my italics):

> The characterization of the explicatum, that is, the rules of its use (*for instance, in the form of a definition*), is to be given in an

> exact form, so as to introduce the explicatum into a well-connected system of scientific concepts.

Carnap's choice of words suggests that an explication does not need to take the form of a definition so long as the result is exact.

Williamson's many valuable insights concerning the explanatory role of knowledge can also be accommodated within explicationism. His remarks about knowledge figuring in nomological laws falls under the narrow concept of fruitfulness. The rest plausibly fall under the broad conception of fruitfulness. However, there is a complication. As we saw, on explicationism, the Gettier problem itself is not a knock-down argument against any analysis or account of knowledge. Hence, Williamson's Gettier argument for his approach amounts, from an explicationist standpoint, to an overreaction to the Gettier problem. The upshot is that while much of Williamson's theory can be appreciated by an explicationist, the net results are not fully coherent under an explicationist translation. But this need not bother the explicationist, who can cheerfully decide to view Williamson's Gettier argument as an anomaly of no particular consequence in an otherwise useful account.

## *References*

Baumann, P. (2016). *Epistemic contextualism: A defense*. Oxford University Press.

Carnap, R. (1950). *Logical foundations of probability*. Chicago University Press.

Cordes, M., & Siegwart, G. (2019). Explication. In *The internet encyclopedia of philosophy*. Retrieved August 20, 2019, from https://www.iep.utm.edu/explicat/

Frankfurt, H. (2006). *On truth*. Alfred A. Knopf.

Gettier, E. L., (1963). Is justified true belief knowledge? *Analysis, 23*(6), 121–123. https://doi.org/10.1093/analys/23.6.121

Goldman, A. I. (1986). Discrimination and perceptual knowledge. *The Journal of Philosophy, 73*(2), 771–791.

Goldman, A. I. (2005). Kornblith's naturalistic epistemology. *Philosophy and Phenomenological Research, 71*(2), 403–410. https://doi.org/10.1111/j.1933-1592.2005.tb00457.x

Hempel, C. (1952). *Fundamentals of concept formation in empirical science*. Foundations of the Unity of Science series, vol. 2, no. 7 (pp. 1–88). The University of Chicago Press.

IAU. (2006, August 24). *IAU 2006 General Assembly: Result of the IAU resolution votes*. International Astronomical Union. https://www.iau.org/news/pressreleases/detail/iau0603/

Kornblith, H. (2002). *Knowledge and its place in nature*. Oxford University Press. https://doi.org/10.1093/0199246319.001.0001

Kornblith, H. (2005). Replies to Alvin Goldman, Martin Kusch and William Talbott. *Philosophy and Phenomenological Research, 71*(2), 427–441. https://doi.org/10.1111/j.1933-1592.2005.tb00460.x

Lehrer, K. (1990). *Theory of knowledge*. Routledge.

Maher, P. (2007). Explication defended. *Studia Logica, 86*(2), 331–341. https://doi.org/10.1007/s11225-007-9063-8

Murzi, M. (2007). *Changes in a scientific concept: What is a planet?* [PhilSci Preprint]. https://philsci-archive.pitt.edu/3418/

Olsson, E. J. (2008). Knowledge, truth and bullshit: Reflections on Frankfurt. *Midwest Studies in Philosophy, 32*(1), 94–110. https://doi.org/10.1111/j.1475-4975.2008.00167.x

Olsson, E. J. (2015). Gettier and the method of explication: A 60 year old solution to a 50 year old problem. *Philosophical Studies, 172*, 57–72. https://doi.org/10.1007/s11098-014-0383-z

Olsson, E. J. (2017). Explicationist epistemology and epistemic pluralism. In A. Coliva & N. Jang Lee Linding Pederson (Eds.), *Epistemic pluralism*. Palgrave Macmillan.

Olsson, E. J. (2022). Explicationist epistemology and the explanatory value of knowledge. *Journal for the General Philosophy of Science, 53*, 41–60. https://doi.org/10.1007/s10838-020-09520-8

Quine, W. V. O. (1960). *Word and object*. The MIT Press.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press. https://doi.org/10.1111/j.1933-1592.2005.tb00537.x

DOI 10.14279/depositonce-22436

RAINER E. ZIMMERMANN

# What Is, and to What End do We Study, Theory?

## Abstract

The presently available ideas on the concept of theory are frequently characterized by various misunderstandings that tend to obstruct a significant insight. This implies obvious consequences for both research and teaching, respectively. Hence, it is necessary to re-establish sufficient clarity by returning to the origins of the concept itself.

So what I will do is the following: First of all, in section 1, I will give a brief overview on the origin and development of the concept of theory. In section 2, I will come to the common view taken today. In section 3 then, I will present metaphysics as a fundamental theory of theories. And finally, in section 4, I will give a currently topical example to enable some conclusions.

Prof. Dr. Rainer E. Zimmermann
Institut für Design Science, München e.V. /
Clare Hall, University of Cambridge, UK
Email: rainer.zimmermann@hm.edu

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

As Cicero and Diogenes Laërtius report, the concept of theory (θεωρία) was already in use at the time of Pythagoras. They tell us that he compares life with a religious festival whose participants come with different attitudes: Some participate on account of honour and glory, others because of material benefit; but only a few come in order to look, to consider, and to understand (Pythagoras leaves no doubt that those are the best).[1] Following the presentation by Hannelore Rausch, Pythagoras assumes that humans actually come into the world as entering a festival. And the philosopher is the person that tries to grasp, through theory, the meaning of this festival. Hence, this meaning of his achieved free cognition is in the celebrating of the festival itself.[2]

1 Cicero: Tusculanae disputationes V 8-9 (According to the 6th edition of Artemis and Winkler, München, Zürich, 1992, 322/323. The passage in Cicero is: "[...] raros esse quosquam qui ceteris omnibus pro nihilo habitis rerum naturam studiose intuerentur [...]" (ibid.). Hence: "There are some rare [people] who disdained everything else and attentively regarded the nature of things." See also Diogenes Laërtius: Vitae philosophorum VIII 8. It is he who compares life to the Great Games (ἐοικέναι πανηγύρει), where the best would behave like mere spectators, comparable with the philosophers who searched for the truth (We quote here according to the Loeb Classical Library, Harvard University Press 1931, 326 sq.; See also the German edition of Klaus Reich, Otto Apelt and Hans Günther Zekl, Meiner, Hamburg, 2015, 440.) [English translations always mine here.]

2 We follow here in principle the organization of Andreas Kirchner: „Alles strebt nach Theorie." Bemerkungen zu Plotins Konzept der Theoria. (2017, 66). https://freidok.uni-freiburg.de/fedora/objects/freidok:151374/datastreams/FILE1/content (16.09.23) See also Hannelore Rausch: Theoria. Von ihrer sakralen zur philosophischen Bedeutung. Fink, München, 1982. Here: 71 sq. (see the digitized text by the Bayerische Staatsbibliothek München, available at: https://digi20.digitalesammlungen.de/de/fs1/object/display/bsb00041869_00001.html (16.09.23))

In Plato, this festival is further relocated into the *eros*-driven origin of a mystery. Hence, in his "Symposion" (210 E), Plato calls what the philosopher sees an "amazement-provoking essential beauty" (τὶ θαυμαστὸν τὴν φύσιν καλόν).[3] This amazement is simultaneously both the beginning and end of philosophy. Plato writes in the "Theaetetos" that: "The philosopher's *pathos* is very much the amazement (τὸ θαυμάζειν). That is to say, there is no other origin of philosophy than this. And who said that Iris be the daughter of Thaumas, does not seem to be a bad genealogist." (155 D).[4]

Hannelore Rausch counts such conceptual determinations as "pre-philosophical" and refers to the earlier results of Koller, who mentions the function of the official festival delegate (θεωρός), which was (as the delegate of his city) to expose himself to the exhibition of a festival (and actually also to judge and evaluate its ritual adequacy).[5] Nevertheless, the terminology is more complex: Rausch discusses different variants which we will not explicate in detail here. But although Plato follows the common usage of the word "to practise theory" (θεωρείν) when referring in the "Nomoi" to the tasks of travellers,[6] the central meaning focuses on the visitor of a festival: "The *theoreín* of these visitors points to human praxis. They need leisure, i.e. freedom from all other purposes, in order to focus their undivided attention onto it. [...] The knowledge of human nature, the awareness of the difference between the humanly Good and Bad springing from this, is visualized as a foundation of the state."[7]

3 Karl Albert: Platon und die Philosophie des Altertums (Part 1). Röll, Dettelbach, 1998, 119.

4 Ibid.

5 Rausch, op. cit., 9.

6 Ibid., 48 sqq.

7 Ibid., 51.

Hence, it is from the traveller whose official assignment is to explore alien states that the concept of theory points to the interior of his own city and then increasingly to a general notion of unveiling the principles of a state (connoting the two-fold meaning of condition and political constitution, respectively). In this sense, the theoretician of the future will strive for knowledge, a process that Plato defines as seeking something that remains the same within a world subject to permanent change, the objective of which is to structure the world of phenomena.[8] Returning to the traveller, the concept of theory refers essentially to analysing an anthropological structure, very much in the sense of Lévi-Strauss and others, and much earlier than we would have expected in the first place (We will come back to this).[9]

Later, in Aristotle, when the philosophical concepts are fused into a unified terminology, *humans actualize themselves by practising theory*. Obviously, in this sense, theory is still a form of praxis, because it is also the *maximal* actualization of humans (ἐνέργεια). The starting point for Aristotle is here the classification of the three modes of living. As he writes in the "Nicomachean Ethics" (EN 1,3; 1095b 15 sqq.): "When taking the three different modes of living into consideration, it is not without reason that the raw crowd localize the highest Good and true Happiness in desire and practise a life of indulgence. Because there are three modes of living in particular that dominate the others: the life we have just mentioned (βίος

8 André Tosel: [keyword] Theorie-Praxis-Verhältnis. In: Hans Jörg Sandkühler (Ed.), Europäische Enzyklopädie zu Philosophie und Wissenschaften. 4 Bde. Meiner, Hamburg, 1990, vol. 4, 585–592, here: 586. (par.)

9 See also Rainer E. Zimmermann: Ungewohntes als Vertrautes. Zu einer stringenten Philosophie der Differenz. In: Francesca Vidal (Ed.), Bloch-Jahrbuch 2018/19. Königshausen & Neumann, Würzburg, 2019, 49–64.

ἀπολαυστικός), the political life (βίος πρακτικός) and the life of philosophical reflexion (βίος θεωρητικός)."[10]

It is only the latter that enables a secure knowledge about the world, based on eternal and necessary principles. Theory shall conceptualize what has been recognized. But what has been recognized is to be left within its form of being.[11] The difference with respect to the concept of theory which is prevalent today becomes obvious. We quote Zekl here: "Knowledge must not be degraded to a mere duplication of the things of the world by means of thinking. As far as it does not reduce multitude and change to rules in a unifying manner, it does not donate any theoretical benefit. [...] The conspicuously rich inventory of things and their apparently chaotic change crave their synthetic and regulative simplification within the horizon of knowledge—hence, their alteration. This alteration by means of science [...] is for Aristotle (as it is in each philosophy) a theoretical act of fixation and reduction. [...] Because being [of beings] is experienced as one that is beautiful, ordered, admirable—πάντα γὰρ φύσει ἔχει τι θεῖον—by nature, everything possesses something divine in itself! [EN 1153b 32 sq.]"[12] And Kirchner adds: "[...] the striving for secured knowledge does not submit itself to any practical benefit, but is chosen for its own sake."[13]—because otherwise this striving would again be practical not theoretical. Essentially, this position of theory derives from the fact that practising theory refers basically to nothing else other

10 Translation according to Eugen Rolfes, reproduced in Kirchner, op. cit., 69 n. 18.; See more recently the edition Gernot Krapinger. Reclam, Stuttgart, 2017 (2020), 16. [Here always abbreviated as EN].

11 Cf. Kirchner, op. cit., 69 (par.).

12 Hans Günther Zekl: Topos. Die Aristotelische Lehre vom Raum. Meiner, Hamburg, 1990, 270. (Quoted here according to Kirchner, op. cit., 70.).

13 Kirchner, op. cit., 70.

than that kind of leisure (σχολή), within which the highest blessedness of humans (εὐδαιμονία) expresses itself being the human prime objective (τέλος). And at the same time, this is the best work of humans (ἔργον).[14] Hence, Aristotle once more: "As far as this contemplating (θεωρία) reaches out, also the blessedness reaches out, and who is granted this contemplating (θεωρία) to some notable degree is also granted this blessedness to some notable degree, not accidentally, but instead due to the contemplating (κατὰ τὴν θεωρίαν), which is estimable in itself. Hence, blessedness will be a kind of contemplation (θεωρία)."[15]

So on the one hand, the close relationship between contemplating (looking-observing) and amazement is still conserved in Aristotle. As Hannah Arendt writes: "In this philosophy, θεωρία is actually nothing but a different, more modest, and preliminary word for θαυμάζειν; the viewing of truth that is eventually achieved by philosophy is the conceptually and philosophically clarified amazement with which it began in the first place."[16]

On the other hand, Aristotle, as compared with Plato, dislodges the original motive of the mystery, though without abandoning this reference completely. For him, that science which is theoretical in the strict sense (σοφία) refers to what is free from the necessities of mortal life.[17]

In the end, Plotinus will later follow the Aristotelian approach and develop it further, and by doing so, the concept of theory becomes the

14 Cf. Kirchner, op. cit., 70 (par.), with reference to EN 1177b, 19–28.

15 EN 1178b 28–32. (Here, I have slightly modified Krapinger's translation.).

16 Hannah Arendt: Vita activa. Oder vom tätigen Leben. Piper, München, Berlin, 18. Auflage 2016 (1967), 385.

17 Karl Albert: Platon und die Philosophie des Altertums, op. cit., 325–328. (par.).

nucleus of his own metaphysical considerations. According to Plotinus, everything (no matter what) strives for theory (πάντα θεωρίας ἐφίεσθαι),[18] which however is organized hierarchically in various degrees according to the means of those who strive. Praxis, on the other hand, especially exterior praxis, becomes a mere shadow of theory. Plotinus precisely founds these means (the potential to strive) on a deficiency of the soul such that the encountered discrepancy is a measure for the exteriorization of praxis.

Hence, within Greek and Roman antiquity, theory was predominantly associated with a contemplative activity that advanced the insight into the true essence of things, whereas present-day opinion opposes this idea in a two-fold manner: On the one hand, by visualizing theory in terms of everyday language as an untechnical, speculative kind of thinking; and in scientific terms as a set of interconnected propositions following self-consistent rules. The transition from the ancient to the modern view is mainly owed to the centuries-spanning transformation of the *enkyklios paideia* (ἐγκύκλιος παιδεία) into the *artes liberales* of the early universities and subsequently into the whole spectrum of isolated academic disciplines.[19]

Looking especially at the sciences now, already in Plotinus a hierarchy has been developed based on differing degrees of separation between theory and praxis. Further development consolidates this viewpoint such that it leads—some time during the 15th century—to an ongoing competition that nowadays gives cause for various linguistic delimitations—even within a sober and austere field like mathematics,

18 Cf. Kirchner, op. cit., 74 (par.), with reference to Enneade III 8, 5, 29–31.

19 Cf. Rainer E. Zimmermann: Enkyklios paidea. In: Gert Ueding, Francesca Vidal (Eds.), Handbuch Rhetorik und Pädagogik. De Gruyter, Berlin, Boston, 2023, 243–261.

when e.g. the denomination "pure mathematics" implies the existence of an "impure mathematics" or in a reverse conclusion "applied mathematics" (angewandt) implies the existence of a "misapplied mathematics" (abgewandt = averted)—referring without doubt to the other discipline, respectively.

Hannah Arendt calls this an "eversion" of theory and praxis or the reversal of the traditional order of *vita contemplativa* and *vita activa*.[20] Comparatively recently, the classical perspective has been vindicated somewhat, e.g. in the theory concept discussed by Pierre Bourdieu, admittedly restricted to the ethno-sociological context: "Indeed, the indication that the theory of praxis which appears as a strict science of the forms of praxis and practical actions is not less theoretical [...] as that theory of praxis which implicitly enters the objectivistic models, is quite correct, but this does not mean that the question would be invalid to ask whether the social pre-conditions that are actually given in order to hold a particular category of individuals ready for the execution of a theoretical activity, are not *per se* an unconscious adoption of a special type of a theory of praxis."[21] In fact, it is anthropology in particular, or ethno-sociology as to that, where the analysis of participating observations is in the topical centre, that a theoretical construction can be understood as one which is based on a preliminary contemplation that shows up as a theory of theories, actually ahead of any theory of praxis. In this case, theory (as analysis and construction) and praxis (as lived experience) are mediated with each other in a complementary manner. Hence, Bourdieu continues:

20 Arendt, op. cit., 367.

21 Pierre Bourdieu: Entwurf einer Theorie der Praxis. Suhrkamp, Frankfurt a.M., 1979 (1976). (Translation into German by Cordula Pialoux and Bernd Schwibs) Originally: Droz, Geneva, 1972. Here: 139.

"The immediate 'understanding' presupposes an unconscious technique of decipherment which is only completely successful where the competence of both: of those who objectively actualize it by means of their action or works, and of those who do so by means of *perceiving* these actions and works, become congruent. In other words, this is the case when the codification as transformation of a meaning into a praxis or into a work coincides with the symmetric technique of decipherment."[22]

It goes without saying that academic teaching clearly reflects this viewpoint in some detail. Unfortunately, Bourdieu's viewpoint has only reluctantly been introduced into lecture courses and seminars. And specialization is still increasing, such that disciplines are further isolated rather than being merged by interdisciplinary efforts. This is especially obvious when we observe that courses for general studies have widely vanished from the curricula or exist only in rare, fragmented residua. The idea that, originally, theory had something to do with contemplation is more or less lost by now. And even the theoretical activity itself degenerates, very often becoming a mechanized routine process (when e.g. in theoretical physics all possible solutions to Einstein's equations are produced for given variables by some sort of algorithm rather than deducing results from interpretational details in the first place).

The fundamental theory, however, is *metaphysics* in the end.[23] In trying to describe the observable world and to orientate oneself within this world, from the beginning on, humans are confronted with the

22 Ibid., 152. (my emphasis).

23 I am following here the structure of my essay: Emergenz und Evolution aus dem Geist der Indifferenz. Systemtheorie zwischen Ethik und Politik. In: Beatrice Voigt (Ed.), Vom Werden. Entwicklungsdynamik in Natur und Gesellschaft. Voigt Edition, München, 2019, 170–177.

circumstance that the world *is not as we observe it*. Instead, there is a generic discrepancy between the true world as it is—independent of humans—versus the observed world, which is the foundation of theoretically and practically grasping what is existing. But the speculative determination of this fundamental discrepancy always acts upon ethics: This is because orientation entails adequate behaviour within the world, and ethics controls for adequacy in the first place.

The origin of this idea dates back however to the philosophy of nature discussed by the pre-Socratic philosophers. They dealt mainly with the question of which among the many forms of observable substances possessed the qualities of fundamental "elements" (i.e. types of matter). It is Heraclitus who for the first time introduces an abstract (non-observable) principle as foundation of the observable, namely the principle of dichotomy within all phenomena. The approach of Parmenides is different, because he equates the concept of *phýsis* with the actual and true essence of things. He asserts that the beings (not the being of beings!) have not become and are everlasting, total, unique, as well as imperturbable and consummated. He declines the becoming of something, likewise out of being or non-being, respectively. It is not before Aristotle that the difference between substance (οὐσία) and subject (ὑποκείμενον) is systematically reflected. The latter gains then the connotation of a primary matter (Urstoff).

Hence, metaphysics deals primarily with the exploration of the foundation (Grund) of observable actuality, obviously to be found in the exterior of the latter. This exploration is essentially a contemplative activity and can thus be called *practising theory*. This type of theory, which is literally basic, can also be called *fundamental theory*. And because its categories are only available in a sufficiently heuristic sense, it can be also be called *fundamental heuristics* (Hogrebe).

Formally, metaphysics is confronted with ethics that deals with adequate practical behaviour within the observable world. However, already in Aristotle both these domains are brought together in a conceptual manner. This is because the criteria for adequacy are based on knowledge that is adequate in the first place, in the sense that it is gained in an adequate way (and codified on a regular basis). Moreover, it is uncovering what is adequate, such that who possesses it cannot do otherwise than to unfold a space of adequate possibilities for the intended practical action aimed. And to permanently seek this space of possibilities is what Aristotle calls supreme blessedness. The latter is thus an action, the actualization of the soul according to virtue. (τὸ ἀνθρώπινον ἀγαθὸν ψυχῆς ἐνέργεια γίνεται κατ' ἀρετήν)[24]

As Wolfgang Schneider has shown in detail, in Aristotle ethics is not only conceptually referred back onto metaphysics, but specific problems of metaphysics can only be clarified and illustrated after referring them forward to ethics in the first place.[25] Schneider continues: "The decision is the moving principle, the *archè* of praxis [...][that] renders an action ethically relevant. [This] is what it becomes due to the prevenient consideration that it is originally based on *lógos*."[26] Visualized in this manner, *eudaimonía* in Aristotle is simply what he himself calls *chrýsis*, i.e. utilization of virtue.[27] Strictly speaking, theory becomes the consummate form of praxis.[28] This is why we can formulate: "Contrary to *poíesis*

24 EN 1098a 16–17.

25 Wolfgang Schneider: Ousía und Eudaimonía. Die Verflechtung von Metaphysik und Ethik bei Aristoteles. De Gruyter, Berlin, New York, 2001, 7. (par.).

26 Ibid., 37.

27 Cf. ibid., 80.

28 Ibid., 106 (caption).

(producing), praxis is consummate. It is, as long as it lasts, always already at its end. In so far, praxis is not within time."[29] And Schneider concludes: "Since Plato, philosophy and truth [...] are in the reversal performed by the psychic action of perceiving and thinking, in the *metastrophé* from becoming to being (Politeia VII 525c 5 sq.). It is necessary however that this is also performed the other way round, as a *katábasis*, as descent from the world of beings that is transcendental to becoming to the world of becoming and change, as descent from philosophy and the *bíos theoretikós* to the *pólis* and to the *bíos politikós*."[30]

We come now to the conclusion, and briefly consider a theory that is not yet completely developed but whose principles have by now been drafted for a number of decades. In fact, from time to time, it is also practically applied. Namely, *psychohistory*. This deals with modelling the structure and evolution of social systems, and resulting prognoses for their future development over long periods of time.

The idea originates from the science fiction literature, specifically what is called Isaac Asimov's *Foundation* trilogy.[31] A small set of elementary rules are collected for this enterprise: The population to be analysed by what is called the First Foundation must have a sufficiently large size (implying the law of large numbers and the intrinsic reduction of variables), its members must remain in ignorance of the analysis, rare events (such as mutations) have to be supervised by a Second Foundation,

29 Ibid., 114.

30 Ibid., 313.

31 Originally a series of short stories (1942–1950), subsequently republished (1951–1953) as three collected volumes. For a recent German edition see Heyne, München, 8th edition, 2012. The trilogy has among the highest circulations ever.

and all of the results including the further development of the methodological basics are summarized in a central knowledge repository called the Prime Radiant. It is not difficult to notice that the elements of psychohistory can be interpreted in the sense of those recent results that are dominating the theories of self-organization and the formation of structure introduced by Thom, Prigogine and others, admittedly not earlier than three decades after the publication of this trilogy.[32]

Of course, Asimov's perspective, dealing with a millennium of projected evolution and recovery following the decline of a galactic empire, is not one that comes to our attention today in the first place (Asimov credited as inspiration Edward Gibbon's *Decline and Fall of the Roman Empire*). But the important point is that many aspects of this psycho-historical layout (e.g., big data, AI, etc.) have been actualized in the meantime without a public noting this in detail. But it is not the prognosis concerning individual historical events or the global development of widespread empires that is at issue. Instead, the more interesting issue is the contextualization of local (social) systems whose variables and parameters can be held within the range of a suitably stable set of interactions among the participating groups, such that a long-term state of dynamical equilibrium can be secured. Hence, it is possible to build bridges from the first formal works on this topic by Nicolas Rashevsky[33]

32 A detailed survey is given by Michael F. Flynn: Einführung in die Psychohistorik. Epilogue to the limited special edition. Heyne, München, 1991, 831–908.

33 See e.g. in: Mathematical Biology of Social Behaviour (1951), Looking at History Through Mathematics (1968).

to the more recent works of the Haken school.[34] Presently, the publications of Peter Turchin are particularly interesting within this context. Turchin himself refers directly to the psychohistory of Asimov.[35]

It is not the appropriate place here for discussing this theory in every detail. A more elaborate work on this topic is currently in preparation.[36] We will only mention that the mathematical nucleus of such an enterprise consists of modelling suitable differential equations that are able to represent stochastic dynamical systems.[37] The result allows for the interpretation of a set of possible transitions of the system from one state to another. It is necessary then to select suitable transitions according to given criteria and actualize them in praxis. In other words: The mathematical nucleus of the procedure serves the analysis of a space of possibilities. But results require further methods of interpretation, those of hermeneutic kind in particular.

It is not a coincidence that the propagation of scientific knowledge is closely related to aspects of science fiction literature. Dietmar Dath has discussed this in a comprehensive work that presents the influence of

34 Wolfgang Weidlich: Sociodynamics. A Systematic Approach to Mathematical Modelling in the Social Sciences. Dover Publications, Mineola, New York, 2000.

35 Peter Turchin: War & Peace & War. The Life Cycles of Imperial Nations. Pi Press, New York, 2006.

36 Rainer E. Zimmermann: Asimov's Legacy. Reconstructing Psychohistory. In preparation. (2024).

37 As I have shown recently on another occasion (at the 6th autumn conference of the Institute for Design Science Munich e.V. at the Leucorea in Wittenberg, 8th September 2023), we can utilize an equation of the Fokker–Planck type for this, including drift and diffusion terms. This is equivalent to a path integral as per Feynman's formulation, because it possesses the structure of a Schrödinger equation.

devising and inventing possibilities onto scientific progress.[38] (It goes without saying that Asimov's *Foundation* trilogy is necessarily mentioned here.[39]) At the very beginning, Dath formulates a motivation that is valid for theories in general, especially for metaphysics understood as a theory of theories: "The later ones will misunderstand the earlier ones. This is the course of the world. If the later ones would like to understand what the earlier ones thought of the later ones, they have to look for traces that indicate whether the earlier ones were able to desist from themselves while speculating in order to recognize other, that is: later, accesses to the world, different from their own familiar ones."[40] Indeed: To develop theories means to establish sufficient distance to what is being given, in particular to desist from oneself. This is essentially one of the theorems of the ancient Stoá: The main point is to study theory in order to take pause within the systematic and methodological diversity of the various disciplines and to look around (in the Stoic manner of what is called *asygkatathetein* [ἀσυγκαταθετεῖν]).[41] It is not completely free of an intrinsic irony that the state of *ataraxía* (ἀταραξία) aspired to in the Stoá, in particular in the later Roman version of the Pyrrhonic scepticism, is equivalent to the contemplative insight into the results of psychohistory's Prime Radiant (Significant science fiction authors often display a tendency towards a spiritual, if secular, prospect in their otherwise sober and rational main ideas. Frank Herbert's Dune cycle is another prominent example). In fact, as far as I can see, there is presently almost no ongoing research following

38 Dietmar Dath: Niegeschichte. Matthes & Seitz, Berlin, 2019.

39 Cf. ibid., 308 sqq.

40 Ibid., 13.

41 Cf. e.g. Sextus Empiricus: Grundriß der pyrrhonischen Skepsis (Ed. Malte Hossenfelder). Suhrkamp, Frankfurt a.M., 10. Auflage 1985.

the direction (close to the original concept of theory) as indicated here. But a conceptual return would be promising, after all.[42]

42 One of the rare examples is the 8-volume series "Einstein meets Magritte," in which the sciences and the arts shall be merged by means of a conceptual synthesis. Cf. Diederik Aerts, Jan Broekaert, Ernest Mathijs (Eds.), Einstein meets Magritte, An Interdisciplinary Reflection, The White Book. Vrije Universiteit Brussel, Springer-Science+Business Media, Dordrecht, 1999.

DOI 10.14279/depositonce-22437

KINGA GOLUS

# Why Should Future Philosophy Teachers Learn 'Theory'? A Philosophical Perspective on Teaching Experiences and Reflections

## Abstract

This article starts from the basic assumption that philosophy is the epitome of theory. When prospective teachers of philosophy study philosophy, they form both their professional identity and the associated ability to philosophize through the theories of others. A corresponding mindset is required in order to be able to do this. This article attempts to describe how philosophy can succeed in training students to become good thinkers in an epistemic sense.

Dr. Kinga Golus
Bielefeld University, Department of Philosophy
Email: kinga.golus@uni-bielefeld.de

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

## *Introduction*

The group of future teachers who study at universities and will later teach at schools is special in many respects. By this, I mean that these students are particularly addressed by the transfer of theory/practice in their professionalization. In recent years, a greater emphasis on practice in university education has been interpreted as a key indicator for practice-oriented teacher training. Particularly with the introduction of university didactic seminars and long practical phases such as the 'practical semester,' future teachers seem to be more interested in teaching practice than in the discipline-specific theory to be taught. This cannot be blamed on them, as discipline-specific theoretical studies have become less important than 20 years ago in comparison to discipline-specific didactics and school internships.

Following from these observations, this article has three aims. Firstly, to give a definition of theory from a philosophical perspective. Secondly, to argue why a strict distinction between theory and practice in philosophy is highly problematic and—for teaching philosophy—counterproductive. Thirdly, to explain why independent thinking is essential for enabling philosophy students to philosophize independently during their studies in order to generate 'theories'. This should be done with reference to intellectual character formation, regarding the theses of Virtue Epistemology.[1]

1 The following observations on the ambivalent relationship between theory and practice in teacher training are based on individual observations made by the author over the last 10 years as a teacher trainer in philosophy.

## *What does it mean to teach theory in philosophy?*

Philosophy is undoubtedly one of the few disciplines at universities in which empirical research is simply not carried out; Rather, philosophy is virtually the epitome of theory.[2] Qualitative or quantitative research methods, as they are relevant in educational science, for example, have no significance for philosophical cognitive processes, although philosophical thinking is often connected to practice (especially applied ethics). Traditional philosophical research is intrinsically theoretical and, at first glance, highly individual. If one attempts to carefully formulate a professional understanding of philosophy, then theory—in a very broad sense—can certainly be equated with philosophy.[3] In academic contexts, it would be more precise to define philosophy as a discipline; and philosophical *theories* not as a discipline, but rather its 'products.' After all, during their first semesters, philosophy students learn the 'theories' or 'products' of philosophers through basic lectures, then memorize and reproduce these in exams. This helps to gain both a historical overview of the discipline and an impression of how the various philosophical theories are systematized. What students learn in those first semesters are theories developed by other philosophers, but this is only a basic building block of the course. Knowing philosophical theories and dealing with them is part of professionalization in philosophy. In my opinion, a first argument can

2 Recently, some philosophers have engaged with empirical research, combining philosophy, psychology, and cognitive science. This 'experimental philosophy' is part of research on, e.g., free will, moral judgment, or responsibility. In the context of this article, it would be interesting to determine whether this 'experimental philosophy' is applicable to research on teaching and learning philosophy.

3 At this point, it should be briefly noted that the whole setting cannot be applied the other way around, because not every theory is philosophical.

be identified here as to why students should deal with theories: Because this is how they develop a discipline identity. This professional orientation is particularly important for prospective teachers of philosophy.

In addition to knowing philosophical theories, the more important learning objective is to be able to philosophize independently at the end of a degree course. This means being able to deal critically with historical philosophical theories as well as being able to philosophize independently—to *theorize*, in the context of this article. This independent philosophizing does not primarily refer to the history of philosophy, as this makes up a relatively small part of current philosophical discourse. Being able to theorize independently—and, in a further step, to discuss the ideas with colleagues—is the highest level of professionalization for philosophers. In my opinion, intellectual independence is a form of thinking that can be systematically taught and also learned, particularly at university. Motivating students to take this path, however, requires more specific guidance from educators and therefore a way of learning or studying (cf. Euler, 2005, 254) that enables this.

However, types of teaching that promote intellectual independence are less likely to be found in the current university education system. As Euler (2005, 258) states: "Answers, not questions and problems, dominate the teaching process. There is a flight from thinking to knowledge. Within this framework, knowledge is accepted by the students, but less scrutinized in terms of its origin, value premises or consequences." Even if the description of this type of learning is quite dramatically exaggerated here, it can certainly be observed among students. I don't think that students lack interest in intellectual independence, but the teaching system gives them few opportunities to develop this ability, especially at the beginning of their studies. This is evident in, for example, examination

formats that are designed for pure reproduction, such as multiple-choice exams. The extent to which this diagnosis can actually be applied to certain university disciplines will not be discussed further. However, it should be noted that all disciplines should provide their students with opportunities for intellectual independence. In my opinion, the capacity for independent theorizing is a central component of intellectual professionalism. Therefore, a second argument can be made here as to why theory should be taught: Because it promotes intellectual independence and thus increases the likelihood of training future scientists.

## *Theory and practice—A counterproductive distinction in philosophy*

A distinction between theory and practice is a model of thought that is by no means established in philosophy. In order to be able to understand this thesis, the understanding of the discipline must first be explained. As mentioned above, philosophy can be interpreted as being purely methodological. The contents are interchangeable. Philosophizing means thinking independently about systematic questions. The independence of thought that is trained here is to be understood as a learning process that is guided during the course and, in the best case, manifests itself at the end of a higher-level degree in professional, independent thinking. This ability is the central educational goal for all students of philosophy, regardless of whether they become schoolteachers or decide on a different profession (cf. Golus, 2021).

This interpretation of philosophy makes it possible to think of theory and practice not as two different areas that influence each other in the teaching profession, but as being fundamentally different. For this reason, the classification of different places of learning—which interprets universities as places of theory, and schools as places of practice—should

be rejected. This distinction alone is problematic for philosophy, as the act of philosophizing is practiced in both places—university and school. The ability to philosophize can be interpreted as an act of practice, as this is where consistent and systematic work takes place. A differentiation between theory and practice contradicts the professional understanding of a methodically interpreted philosophy. Furthermore, a distinction between theory and practice in teacher training can be confusing for the professional self-image of prospective philosophy teachers. While an antagonistic distinction is common (especially in comparison with other teaching disciplines), two different areas are implicitly constructed in which students and teachers operate. Theory is assigned to the university, and should leave this place of learning in order to be taught in schools—i.e., a transfer from one place of learning to the other. Such an antagonistic model cannot be applied to philosophy interpreted as a method, because although philosophical practice changes location, from the university to the school, the act of philosophizing remains identical; It is the same activity at both university and school. It is therefore not appropriate to speak of a theory–practice transfer in philosophy, but rather of a practice–practice transfer. This way of thinking can help prospective philosophy teachers to not perceive the two places of learning—university and school—as genuinely different (cf. Golus, 2021).

However, practice–practice transfer would make philosophy an exception in teacher training, as the prevailing model continues to differentiate between theory and practice, which is significantly influenced by an educational science perspective. This distinction forms the basis for self-positioning in the respective discipline that student teachers will enter. If this dichotomous understanding is projected onto philosophy, this can lead to confusion or even resentment in the self-perception of prospective philosophy teachers. A paradoxical

phenomenon can be observed in this context: On the one hand, student teachers 'only' want to go into practice, i.e. into schools. Practice seems to be hierarchically subordinated to theory in their professional self-image. On the other hand, theory is devalued because it is classified as 'too difficult' and therefore partly irrelevant for school practice. A deprofessionalization of prospective philosophy teachers occurs when they are not aware that philosophical thinking practice at university already represents the *practice* of a philosophy teacher (cf. Golus, 2021). Consequently, if philosophy is interpreted methodically, including the ideas of other or historical philosophers, then the contradiction that arises, for example, from the logic of educational science, cannot be resolved. How can intellectual independence be taught and learned?

Although not entirely new, a teaching and learning concept that has often been cited in recent years is that of inquiry-based learning. According to Dieter Euler, this concept can help students to overcome an intellectual lack of independence. Euler is concerned with nothing less than a restructuring of the foundations of educational theory, so that universities establish inquiry-based forms of learning that teach and promote the required intellectual independence (cf. Euler, 2005, 258f.). In terms of how students become as intellectually independent as possible, philosophy has the opportunity to argue from within itself. This means addressing the theses of a current international debate that shows a way for students to become good thinkers in an epistemic sense. I refer here to the theses of Virtue Epistemology, which are explained below.

## *Virtue Epistemology as the basis for an inquiring attitude*

A central purpose of Virtue Epistemology is to shape people's intellectual dispositions in such a way that they are enabled to think well in an epistemic sense, i.e., such that their thinking leads to knowledge or justified beliefs. Enabling students to think in this way is primarily about

helping them to develop their intellectual personality, which is characterized by corresponding character dispositions. Certain intellectual dispositions with regard to the facilitation of inquiry-based learning are central, and precede the actual act of inquiry-based learning, which can be seen as an epistemic end. Publications on virtue theory deal with the development and cultivation of intellectual dispositions that enable the practice of inquiry-based learning. In this context, virtues are defined as traits of character, regardless of whether they are intellectual or moral. They are characterized, among other things, by the fact that they can be acquired independently by each person (cf. Golus, 2024). Since acquisition is determined by a process, a certain amount of time and work must be invested. "This means that typically a virtue is acquired through a process of habituation [...]" (Zagzebski, 1996, 135f.). However, it does not mean that a virtue, once acquired, is a permanent part of the character; It is a stable part, but may also be lost. Characteristic of an intellectually virtuous person is a fundamental motivation to achieve so-called epistemic ends, which can include inquiry-based learning (cf. Baehr, 2011, 208). Jason Baehr expresses this in a more concrete and differentiated way:

> Intellectual virtues aim at deep understanding. Put another way, an intellectually virtuous person is one who thinks and inquires in ways that are open, honest, fair, careful, and courageous out of a desire for an understanding of important subject matters. She is not content with simply memorizing what others (including her teachers) have to say; nor is she satisfied with a superficial or cursory grasp of important topics. She wants to know why things are the way they are, how they have come about, how they work and relate to each other, and so on. She desires deep understanding. It follows that educating for intellectual character growth requires educating for deep understanding. (Baehr, 2015, 7)

In order to facilitate deep understanding among learners, an appropriate mindset is required. This mindset consists of a set of intellectual virtues that predispose researchers to acquire knowledge, to probe it deeply and, furthermore, to generate reliable knowledge themselves. Therefore, intellectual virtues play a central role in enabling students to practice inquiry-based learning. It is about establishing character traits that help students develop the disposition not only to acquire knowledge actively and responsibly, but also to generate it. This process is highly individual and is the responsibility of each person. Whether a person forms a true or false opinion is fundamentally dependent on whether they choose character virtues or vices as the basis for their decisions. Character virtues, like open-mindedness, inquisitiveness, or thoroughness, enable a person to successfully and reliably gain knowledge—for example, in research. This also increases the probability of achieving epistemically good results (cf. Kindley, 2021, 95–96).

Applied to the process of inquiry-based learning, it is not only a matter of learning and practicing intellectual virtues, but also of avoiding intellectual vices. Intellectual vices, which are seen as defects or deficits of the mind, are also learned like virtues, are habitualized, and can be unlearned. So which intellectual virtues should we embrace and which intellectual vices should we avoid, in order to conduct reflective research and generate knowledge independently?

Further considerations, which would focus in particular on the promotion of intellectual virtues in the seminar, would be to identify core virtues. These could include, for example, intellectual openness and intellectual courage. However, these are by no means sufficient to become good thinkers. What is needed is a cluster of virtues that can be defined, taught, and practiced. Philosophy chooses its own path here, since in a first

step it attempts to help students to become epistemically good thinkers from within itself—here, from the perspective of virtue epistemology. At this point, in my opinion, a didactic desideratum for higher education can be identified.

## *References*

Baehr, J 2011. The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology. Oxford: Oxford University Press.

Baehr, J 2015. Educating for Intellectual Virtues. An Introductory Guide for College and University Instructors. Published online. https://jason-baehr.files.wordpress.com/2013/12/e4iv_baehr.pdf., 20.06.2024.

Euler, D 2005. Forschendes Lernen. In: S. Spoun & W. Wunderlich (Hrsg.), Studienziel Persönlichkeit—Beiträge zum Bildungsauftrag der Universität heute (S. 253–271). Frankfurt/Main: Campus.

Golus, K 2021. Zur Theorie-Praxis-Relationierung in philosophischen Bildungskontexten. In: C. Caruso, C. Harteis & A. Gröschner (Hrsg.), Theorie und Praxis in der Lehrerbildung. Verhältnisbestimmungen aus der Perspektive der Fachdidaktiken. 1st ed. (S. 121–129). Wiesbaden: Springer.

Golus, K 2024 (forthcoming). Intellektuelle Tugenden als Basis einer forschenden Grundhaltung in der Lehrkräfteausbildung. In: Jahrbuch für philosophiedidaktische Forschung. Berlin: SpringerVS/Metzler.

Kindley, S 2021. Tugenderkenntnistheorie. Intellektuelle Tugenden und der Begriff des Wissens. Paderborn: mentis.

Zagzebski, L 1996. Virtues of the Mind. Cambridge, UK: Cambridge University Press.

DOI 10.14279/depositonce-22438

HARALD A. MIEG

# Why Theorizing Should be Seen as a Form of Research

## Argument

Here, I present the argument that theorizing, i.e., the formation of concepts and theories, should be seen as a form of research in its own right. This view is not new. It was also explicitly postulated in a position paper by the German Science and Humanities Council (Wissenschaftsrat, 2012). That paper also contained an important justification: theorizing—as a separate form of research—requires special, separate research infrastructures. There are also didactic reasons. I will discuss these in more detail.

My subsequent thesis is that theorizing is a difficult form of research that is only taught specifically at universities; it is therefore characteristic of higher education. I will develop my argument in four steps: First, I define scientific research. Second, I relate this definition to theorizing.

Prof. Dr. Harald A. Mieg
Humboldt-Universität zu Berlin
Email: harald.mieg@hu-berlin.de

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

Third, I respond to possible objections, e.g., that theorizing should be understood as creation rather than method, or that theory is superior to research, etc. Fourth, I present some conclusions. These show that it is also fruitful from both a scientific and a professional point of view to understand theorizing as a form of research.

## *Scientific research*

In the context of science, research is the methodical generation of knowledge. I understand science as a social project of systematic knowledge acquisition that is well over 2000 years old. Today, science is globalized and professionalized; and research is indispensable for science.

Research as knowledge acquisition takes time, and may vary in its level of success. Research is an activity with uncertain results. Results cannot be predicted; or else, if that were possible in principle, there would probably be no science as we know it.

From the point of view of science, it is the knowledge (in German: Erkenntnisse), i.e., the results of scientific research, that must be preserved. Scientific knowledge has different forms of representation.

- Text
- Data
- Formulae and equations
- Images
- Preparations, etc.

Knowledge is the working material and yardstick in science. Scientific knowledge is evaluated, verified, differentiated, reformulated, formalized, and discarded. Knowledge is both the basis for evaluating existing research and the justification for further research.

Given both the immense importance of knowledge and yet its inherent uncertainty, science strives to constantly improve research methods. Good, professional research is seen as an indication of possible knowledge—which in some cases may ultimately prove to be incorrect.

## *And theorizing?*

Everything said in section one, concerning research, applies to theorizing. Theorizing produces theories. In science, theories are treated like knowledge (Erkenntnisse). We can think of them as quasi-factual or as complex propositions. Theories therefore differ from:

- Questions
- Hypotheses
- Assumptions
- Values
- Methods

None of these can be considered knowledge in themselves, but they can be elements of theory. They can also initiate the process of generating new knowledge

Since theories result from research, the German Science and Humanities Council (Wissenschaftsrat, 2012) proposed including theorizing among the forms of research. The Council understands forms of research as an "interdisciplinary order" that distinguishes "heuristically at a medium level of abstraction" between six individual forms of research: Experimental; Simulative; Observational; Hermeneutic-interpretive; Conceptual-theoretical; and Formative (see also Mieg, 2019).

Conceptual-theoretical forms of research are characterized as follows: "...such as those found paradigmatically in mathematics, the natural sciences and philosophy, obtain their results through intellectual

constructions and logical deductions. As with other forms of research, the quick accessibility of the relevant literature in the respective field is particularly important for them. In some areas, they also make use of computer programs (e.g., for automatic proof, proof checking or network simulations)." (Wissenschaftsrat, 2012, p. 37, translated)

In principle, the characterization of theorizing as research is independent of our understanding of what a theory is. Theory is usually understood as propositions or sets of propositions. According to the structuralist approach to theory (cf. Stegmüller, 1976), theories are more than just propositions. Rather, a theory includes paradigmatic and intentional applications (cases, models) as well as "theoretical terms". The latter are expressions that take on a specific meaning within the framework of the theory, e.g., Newton's concept of mass or the concept of institution in some social science theories. Theoretical terms require special translation if they are to be used in other theories. My argument should also apply to this extended understanding of theory; If so understood, theories can be treated like knowledge.

## *Objections*

There are several possible objections to an understanding of theorizing and conceptualizing as research. Four seem relevant to me.

Objection 1: Theories are not knowledge, they are only hypothetical. Therefore, theorizing is not research.

Response: This objection depends on the question: What is knowledge? Two extreme cases are conceivable: first, only data is knowledge; or, second, knowledge is an interpretation of data, findings, etc. (in light of the interest in knowledge). In the second case, theories would have the same epistemological status as interpretations; in this case,

the objection would fall flat. If one allows only data to count as knowledge (which would not be my position), then one would have to reply that theories are not data, but are treated like complex data sets in science. They are evaluated, preserved, reused, etc. Like theories, data sets can be discarded if they show inconsistencies or if the data collection was flawed.

Objection 2: Theorizing is creative, not purely methodological.

Response: This can be countered by the fact that there are also forms of research involving varying degrees of creativity, for example in mechanical engineering or architectural design. According to Aristotle in the Nicomachean Ethics, creative forms of research correspond to techne (τεχνη) as a practical form of knowledge (2007; cf. Mieg, 2019).

Objection 3: Theorizing is too important to be subsumed under research. It represents a separate, superordinate scientific field of activity. Hence, for example, the division between theoretical and experimental physics.

Response: It should be noted that the importance of theorizing does not prevent it from being categorized as research; the distinction between theoretical and experimental physics could also be based on two forms of research (theory and experiment). The fact that theorizing can be combined with all other forms of research may characterize it, but it does not argue against understanding theorizing as a form of research.

Objection 4: What applies to theorizing in the strong sense does not automatically apply to conceptualizing as a weak form of theorizing. Hence, even if theorizing is a form of research, conceptualizing need not be.

Response: In principle, it is true that not all conceptualizing can be considered research. The same is true of other forms of research. Not every

observation is research, per se. Rather, the scope and systematic nature of the observation determines whether it can be considered research. This is also the case with conceptualizing. It should be mentioned that we cannot simply omit conceptualizing when we systematize forms of research. Conceptualizing as a form of research plays a central role in jurisprudence, and would not be covered by theorizing in this case.

## *What follows?*

What follows if we understand theorizing as a form of research? Generally speaking, the distinction between forms of research in the policy paper of the German Science and Humanities Council (2012) aimed to keep in mind the specific research infrastructures that are necessary for this form of research. In the case of theorizing, this appears to be the classic academic infrastructure that supports academic discussion and text production: seminar rooms, conference facilities, journals, and academic publishers.

There are also implications for didactics. Forms of research are didactically suitable in different ways for introducing students to academic work. In a reanalysis of studies on undergraduate research, I investigated how easy it is for students to get started in research, depending on the form of research (Mieg, 2019). It turned out that observational and simulative forms of research also offer good opportunities for first-year students to start their own research, whereas those involving theorizing are more challenging. For this reason, theorizing takes place in subjects at an advanced stage of education.

My thesis would be that theorizing is a difficult form of research that is specifically taught only at universities; it is therefore characteristic of higher education. Theorizing may seem like an academic matter, loosely related to professional knowledge. However, it is not only the development and communication of professional knowledge that benefits

from theorizing; the competition for practical–professional solutions, for example in the field of the environment (e.g., life cycle assessment), is also driven by abstract models (cf. Mieg & Evetts, 2018).

Theorizing is a scientific activity that can also be understood as a form of research. Theories have an epistemological status and are used as such in science and professional practice.

## *References*

Aristoteles. (2007). Die Nikomachische Ethik (Greek–German, translated by O. Gigon, 2. Ed.). Artemis & Winkler.

Mieg, H. A. (2019). Forms of research within strategies for implementing undergraduate research. ZFHE, 14(1), 79–94.

Mieg, H. A., & Evetts, J. (2018). Professionalism, science, and expert roles: A social perspective. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt & A. M. Williams (Eds.), The Cambridge handbook of expertise and expert performance (2nd ed., pp. 127–148). Cambridge University Press.

Stegmüller, W. (1976). The structure and dynamics of theories (translated by W. Wohlhueter). Springer.

Wissenschaftsrat / German Council of Science and Humanities. (2012). Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020 [Recommendations for the further development of scientific information infrastructures in Germany by 2020]. Wissenschaftsrat.

DOI 10.14279/depositonce-22439

RIIKKA HOFMANN

# From a Learning Sciences Perspective: The Importance of Theory for Facilitating Learning in Universities

## Abstract

This Chapter considers the question of how the Learning Sciences can help university educators develop teaching practices that can enable all students across increasingly diverse cohorts to become agentic learners and world-makers. It defines theory as a vehicle for questioning present practices and imagining alternative worlds; building on Cultural-Historical Activity Theory (CHAT) and Dialogic theory, it further discusses concepts as the intermediate sensitising tools that theory offers to educators for such world-making work. Through discussing a set of intertwined challenges facing university educators—teaching diverse student populations in equity-oriented ways that enable agentic learning and a horizontal expansion of all students' repertoires of knowledge and practice—I outline an approach to teaching and learning with, through and on theory, which I term 'multi-modal cognitive simulations'. Three conceptual tools from the Learning Sciences are used to illustrate how theory can support agentic learning in universities.

Prof. Dr. Riikka Hofmann
University of Cambridge, Faculty of Education, Hughes Hall
Email: rjph2@cam.ac.uk

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory*. *Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

## *Introduction: The puzzle*

The aim of university teaching is creating "agentic and critically self-regulating students who are able to take themselves forward as enquiring learners, both during the programme and after they have graduated" (Edwards, 2016, p. 124), with an emphasis on equity understood as world-making (Gutiérrez, 2023). This aim is framed in practice by a global trend towards increasingly diverse populations of higher education students (Dracup et al., 2020), characterised by increasing linguistic, cultural (Forbes et al., 2021) and socioeconomic (Ilie et al., 2022) diversity and growing numbers of students with additional or diverse learning needs (Hubble & Bolton, 2021; McGorry et al., 2024). These developments have led to more complex student identities, prior knowledges and student perceptions of their own capabilities and learning needs in higher education (Forbes et al., 2021). While welcoming these trends, many university educators acknowledge challenges in finding ways of simultaneously engaging a broader range of knowledge and learning backgrounds, especially in larger groups (O'Shea et al., 2016). This Chapter asks how the Learning Sciences can help university educators to continuously develop teaching practices that enable all students to become agentic learners, especially in contexts where students represent different regions, languages and/or histories.

Many discussions on how we teach diverse learners in university settings centre around pedagogic and technological innovation (Hofmann, Chu, et al., 2024; Hofmann et al., 2021; Ilie et al., 2024; Kostusiak et al., 2017). While these are important developments, research in the Learning Sciences (Vermunt et al., 2019) highlights that focusing solely on teaching/programme characteristics while ignoring—or conflating these with—qualities of *learning processes* does not help researchers and educators address the challenges of how to go beyond

supporting individual learners/learning situations, and to identify what learning mechanisms might translate from one learner/learning situation or programme to another (cf., Förtsch et al., 2018; Gartmeier et al., 2015). Moreover, Gutiérrez (2023) highlights that building on standardised pedagogical or technological tools that simplify practice does not support equity and agency for university students from non-dominant communities. Gutiérrez and colleagues (2009) argue that this goal requires new forms of learning that provide students with genuine access to tools from existing scholarship while legitimising and horizontally sharing their repertoires of knowledge and experience, and students' place in academic discussions. I argue that pedagogical and technological tools hereby need to be thought of as technologies of *learning*, not solely as technologies of organising or delivering teaching.

A focus on difference with an individualised approach to learning can be an agentic way of engaging non-dominant learners. However, especially in large-group teaching settings in the university, a reactive individualised approach can easily become a 'deficit' approach to engaging with diverse learners (Dracup et al., 2020). It can reinforce an underlying assumption of learning as only vertical progress (cf., Gutiérrez et al., 2009; Hofmann, 2007), and thereby encourage an alienated surface approach to learning (Mann, 2001; Ilie et al., 2024). An anticipatory approach to developing more inclusive, agentic approaches to teaching and learning to meet the needs and entitlements of evolving student populations instead highlights heterogeneity, joint activity and an interplay between vertical and horizontal learning; this anticipates and acknowledges the diversity of students as a reality and a resource, not a problem/deficit, and seeks to understand, contextualise and engage with it (cf. Gutiérrez et al., 2009; Dracup et al., 2020).

Research in the Learning Sciences indicates that a paradox exists in teaching non-dominant learners and those with diverse learning needs. Namely: assuming that all groups can learn in the same ways can hinder non-dominant students' learning (cf. Chernikova et al., 2020; I will return to this below); conversely, if we presume that non-dominant learners are less capable of learning than traditional students, this assumption limits their learning opportunities (cf., Horn, 2007; Jackson et al., 2017). So how can we teach large groups of diverse university students in ways that acknowledge and engage with—but do not define—students' (futures) by their challenges and prior knowledges; and that may therefore enable all of them to become agentic, enquiring learners with powerful 'world-making' tools to critique and generate knowledge about things that matter to them in their current and future worlds?

The complex and contradictory nature of social phenomena, illustrated by the above paradox, calls for the use of *theory* in reflecting on and working with such phenomena. Theory does not resolve the contradictions in our social world, such as those related to learning and agency; rather, it enables us to explore them in their complexity (Rainio & Hilppö, 2017). It helps capture and hold together aspects that are not easily obvious, visible or explainable. As a "vehicle for thinking otherwise" (Ball, 1995), theory hereby is crucial for envisioning alternative futures—coming to see what we do not yet know. It does so through offering a "language ... of imagination" (Ball, 1995) and enabling "the cultivation of anomalous and surprising empirical findings" (Timmermans & Tavory, 2012, p. 169). This gives theory a kind of revolutionary power. The purpose of theory in our educational research and practice is "to de-familiarise present practices and categories, to make them seem less self-evident and necessary, and to open up spaces for the invention of new

forms of experience" (Ball, 1995, p. 266).[1] Dialogic theory (Mercer et al., 2019), one of the two post-Vygotskian theories this Chapter builds on, refers to these as 'dialogic spaces': drawing on Bakhtin, Dialogic theorists argue that it is the holding together, and inter-animation, of different ideas in a dialogic space, in which there is uncertainty and a multiplicity of perspectives, that leads to new insights (Wegerif et al., 2020). It is in such dialogic spaces, in which a phenomenon can be looked at from multiple angles, that different futures can be imagined (Hofmann, 2020).

If theory is a way of 'reading' complex social practice in a non-singular way (cf. Rainio & Hilppö, 2017), *concepts* provide the concrete lenses by which to look at specific concrete educational practices in a theoretical, world-making, way. Concepts, Blumer (1954) argued, are the means of establishing a connection from wider social theory to a world of practice, and as 'sensitising' tools can helps us look in new directions and ways (cf., Hofmann, Paavola, & Rainio, 2024). This is the particular angle in this Chapter's remaining discussion on the uses of theory in agentic, world-making educational practice. Another post-Vygotskian theory on which my work builds, Cultural-Historical Activity Theory, or CHAT (Engeström, 2009), further illuminates this.

## *CHAT and intermediate concepts as theoretical tools to enable imagining new worlds*

To illuminate my perspective on how theory can both inform and help change educational practice, I will discuss the role of what Engeström

1 Doing so with regard to particular aspects of the social world, such as learner agency, (using) theory in this sense is at once more specific, and more radical, than more generic notions of 'reflective' practice (cf. Argyris & Schon, 1992; Eraut, 2004).

(2009), framed by CHAT, calls 'intermediate theory'. Intermediate theory refers to concepts that are informed by more general theories of learning. However, they are data-driven and specific to certain types of activity, such as teaching (Hofmann, 2024). What does intermediate theory have to do with changing educational practice?

CHAT understands human actions and higher-order thinking as mediated by tools. Tools in CHAT can be used to work on one's practice. While primary (or 'how') tools typically offer a solution to carrying out existing practice, a certain way of dealing with a pre-defined problem or task, *conceptual* tools can enrich and expand educators' understanding and interpretation of a problem (Hopwood et al., 2016). They thereby enable actors to analyse and work on their practice problem, to imagine new possibilities for practice and open up new possible responses to it (Edwards, 2016).

Unlike primary tools, concepts are not solutions, but tools to think with and to re-think. Building on general theory from the Learning Sciences while also grounded in data of concrete practices, intermediate concepts "can be used in other settings as tools in the design on locally appropriate new solutions" (Engeström, 2009). Theory in this sense offers the possibility of going beyond the specificities of individual learners and learning situations: conceptual tools can be used in different ways by different actors and in different situations, involving educators' agency (Edwards, 2016). I argue herein that intermediate conceptual tools drawn from wider theory in the Learning Sciences can provide us as university educators with new perspectives on our problems of practice and help us conceive new approaches beyond what we might otherwise imagine. In this Chapter I will discuss examples of such tools.

Where does one start in identifying such consequential differences that might enable all learners to expand their repertoires (Gutiérrez, 2023) and develop agency as learners? With CHAT, I argue that such a shift of perspective requires theoretical tools to help us identify and notice things that matter, that can make a difference. I will discuss three examples of challenges faced in university teaching; which intermediate conceptual tools might make a difference; and how. These examples relate to the goals of agency, equity and sustainability articulated above.

## *First challenge: Starting where the learners are, not where we want them to be, is not a game of 'more' but a practice of 'difference'—Identifying* differences of consequence

Students' broad range of prior knowledges, languages and/or learner identities can present challenges for university academics as educators. Sometimes this is discussed as a problem of needing to do 'more' (more time spent on supporting students, more content, more opportunities for reflection). However, it is often not sustainable for educators to do more; moreover, Learning Sciences research suggests that it may not be a beneficial response. A meta-analysis of simulation-based learning in higher education by Chernikova and colleagues (2020) demonstrated that traditional approaches to higher education teaching, such as reflection—that are beneficial for students with high levels of prior knowledge in the topic—may not be helpful for students with less prior knowledge;[2] and may actually be detrimental to the learning of students with lower prior knowledge. Doing 'more' (of the same) does not support learning.

2 It is important to note here that this does not refer to 'ability' but to the learning opportunities students have had previously in and outside higher education.

The conceptual tool I suggest that can be helpful here is Carlile's (2004) notion of 'differences of consequence'. Rather than simply apply new (e.g., technological) teaching tools or follow new teaching models, this concept calls us to consider what aspects—what differences—of the learners and learning at hand are 'of consequence' in a learning topic and situation, for the learners and the wider context.

Chernikova and colleagues (2020) suggest that a difference of consequence in this case is not 'more time' or 'more opportunities to reflect'. Instead, the relationship between levels of prior knowledge and ways of engaging with academic content and reasoning was found to be consequential. Their meta-analysis showed that, across medical and teacher education students, worked examples (rather than reflection tasks) were helpful in supporting the learning of students with less prior knowledge. In learning through worked examples, learners follow through or observe the solution to a task—an approach that is particularly common among science subjects (Fischer et al., 2014).

What might this mean for teaching in the social sciences, where many university programmes aim to offer students opportunities for reflection (Grossman et al., 2009)? In my own teaching (for example, of social scientific research methods or research-use in policy and practice), I have worked to identify what underlying forms of reasoning and noticing used by academics, practitioners and policy-makers are 'of consequence'—those that really matter and are influential for knowledge generation—in what I would like my students to learn in order for them to become agentic knowledge-creators. This might be focused, for example, on the different ways in which academics use questions in research: textbooks typically describe good research questions as 'clear', 'focused' and 'analytical'. While all these criteria are 'true', I have found they are not of consequence for student learning: none of them illustrate to students how (good) research

questions do their epistemic (and powerful) work, how researchers use questions to think and come to see the world in new ways (See Box 1, for example).

*Box 1: Practice example of multi-modal cognitive simulation-based teaching*

With students newly arrived from around the world, who have no shared history and few culturally shared knowledges/practices, we might first explore the London Tube maps, as most students will have arrived using the London Underground system.

By comparing the standard Tube network diagram with the geographically accurate map of the Tube network, and then the version that shows whether stations have step-free access, we examine and 'experience' how materials (e.g., each map) are not valid ("data") in their own right, but only in relation to the questions we use them to answer. We might then explore the power of questions to examine everyday materials in ways that can unveil how inequity works, in order to identify what is worth researching. For example, I might ask students to consider the step-free access Tube map in connection with the development of UK equity legislation, and encourage them to ask what studying those jointly tells us about the inclusion of people with disabilities in our society (stations were only built as accessible once this was made mandatory by legislation).

This aims to 'simulate' how academics use questions, helping students understand—without the need for shared scholarly prior knowledge—how (good) questions are both crucial to research rigour, and the source of power to reveal and make available to scrutiny cultural/systemic inequity.

The second question, which Carlile (2004) calls us to ask, is: How we can represent those differences of consequence (e.g., here, the role of

questions in research), to support and expand the repertoires of all learners and foster agentic engagement across differences in prior knowledge? Carlile highlights that where a group has limited common knowledge, as may be the case in a heterogenous cohort of students, representing differences of consequence to enable learners to share and assess knowledge is harder and requires more energy.

To be able to do this, then, instead of only explaining and discussing key aspects (e.g., of research questions) with students, or expecting them to learn those solely by engaging with scholarly texts, I have developed what I call 'cognitive simulations' to emulate and evoke the types of reasoning and noticing experiences that scholars in my field engage in. This is coupled with identifying new ways of representing 'differences of consequence' when teaching diverse learners: utilising multi-modal means to present those differences, with the aim of 'simulating' the reasoning experiences that scholars may engage in, to make those accessible to a range of learners.

This might involve academic, policy, journalistic and fictional texts and narratives; various forms of music, imagery, art and drama; and simulating phenomena from other disciplines such as the natural sciences or the history of science. The aim of these is not simply to explain/illustrate something, but to allow students to encounter—in both embodied and cognitive ways—powerful new means of reasoning and noticing in their discipline, which can become critical world-making tools. It resonates with Gutiérrez and colleagues' (2009) call for to offer (all) students scholarly tools that can enable them to critically scrutinise their own worlds of importance.

### *Second challenge: Changing learning cultures and conversations may in itself alienate learners instead of supporting them—The multidimensionality of dialogic learning norms*

Most university educators who have attempted to start teaching in a different way will have noticed that this can be far from easy. It is not only difficult for educators themselves; their attempts to change well-established educational practice can be met disengagement or disapproval by students. The 'paradox of norms' (Hofmann, 2024) is at play here: historical and sociocultural norms regulate patterns of behaviour and participants' expectations in educational contexts. Norms enable participants to know how to interpret, and what to expect—both from and within—specific educational activities such as university teaching. Without changing such norms, practice cannot change; however, if educators depart from well-established shared norms, they can no longer rely on participants' understanding, approval of and engagement in the activity. Research suggests that, in order to change educational practice, educators need to make its existing and desired new norms visible and—in themselves—the target of learning (Engeström, 2009; Hofmann, 2024).

This, however, can be difficult without an understanding of the nature of those norms and how they function, since well-established norms can be invisible unless challenged. So here, I suggest that the concept of the *multi-dimensionality of norms* (Hofmann & Ruthven, 2018) can be helpful. Building on Dialogic theory highlighted above, I will draw on the example of dialogic teaching, an approach widely relevant to teaching in universities, which has a long history in German university education albeit under different names. According to the notes of his student Blanck from 1885 on display at the Humboldt Forum permanent exhibition in Berlin, Friedrich Paulsen, the first Professor of Education in

Germany, based at the University of Berlin, believed that teaching is a collective, creative process in which "teachers and listeners build science together" (Blanck, 1885). This is the idea at the heart of 'dialogic teaching', an approach that emphasises two broad aspects: the *distribution* of talk, highlighting the importance of opportunities for students to take part in discussions and share their ideas; and the *ideas* involved, underscoring the importance of valuing and giving space to students' multiple viewpoints (Hofmann & Ruthven, 2018; Mercer et al., 2019). A key to the goals of learner agency and equity, understood as legitimising and horizontally sharing learners' repertoires of knowledge and experience, is that dialogic teaching be not only about allowing students to *talk*, but considering the *ideas* they bring in.

Despite a growing body of research demonstrating the benefits of dialogic teaching and learning (Asterhan et al., 2015; Mercer et al., 2019), research on university practice shows that lecturer-fronted talk dominates (Hardman, 2016), pointing to complexities of change encountered by both staff (Heron, 2018; Shea, 2019) and students (Engin, 2017). Research shows that changing classroom practice in this direction requires the explicit development, and mutual appropriation, of new interactional norms that can support the move away from vertical-only notions of learning (Hofmann & Ruthven, 2018).

The literature describes norms as involving both a surface level of patterns of behaviour that are recurrent and obligated in the particular social practice of the classroom, and an underlying rationale for such actions and interactions (Herbst et al., 2011). Surface level norms in dialogic teaching practice involve students contributing—and listening to each other's—ideas, not only those of the educator. The key, however, to understanding and using the concept of norms in order to change teaching practice towards sharing repertoires of knowledge is the observation that

interaction norms for teaching and learning are multi-dimensional. They do not have a single underlying rationale; rather, surface norms such as 'listening' or 'contributing' can draw their meaning from a *range* of underlying rationales. These are of consequence, since these underlying rationales frame the nature of teaching and learning activities in different ways, not all of which contribute to horizontal agentic learning. We previously termed these the *operational* (relating to ways of carrying out teaching/learning tasks), *interpersonal* (relating to ways of treating others), *discussional* (relating to promoting discussion) and *ideational* (relating to the content of the discussions and the concrete disciplinary ideas involved) dimensions (Hofmann & Ruthven, 2018). Dialogic teaching is often implemented without drawing on the ideational dimension, thereby failing to capitalise on its potential to support the inclusion and expansion of students' repertoires of knowledge.

My final example concerns what is done with the ideas shared by students in dialogic teaching.

### *Third challenge: Identifying the different epistemic roles students (diverse knowledges) can play in classroom learning conversations—Classroom epistemic order*

Research in higher education has highlighted that fostering student agency as learners requires going beyond considering deep versus surface approaches to learning, and instead focusing on engaged experiences of learning and student agency (Mann, 2001). Learner agency is not simply a case of enabling choice and personalising learning offers, such as focusing conversations on student-contributed ideas; it is about opportunities to learn about, engage with and contribute to shared knowledge development (Hofmann, 2007; Mann, 2001). Examples of such practices include moving away from deficit-based approaches, toward teaching and

learning in ways that enable all students to be 'smart' (Gutiérrez et al., 2009); this involves making scholarly reasoning transparent (first 'challenge') and enabling participation in, and contribution to, dialogue (second 'challenge'). However, transparency and scrutiny of the knowledges discussed, and ways of warranting them, are also important for developing "contexts of productive criticism" (Gutiérrez et al., 2009).

While this may justifiably involve a wholesale review of the curriculum, more immediate steps may also be relevant. I suggest that another intermediate concept that may illuminate key dimensions of teaching and learning dialogue—which might otherwise remain invisible, yet could enable the development of contexts of productive criticism—relates to *epistemic order*: "the way in which the exchange and development of knowledge takes place in the classroom" (Ruthven & Hofmann, 2016).

Epistemic order in teaching and learning situations consists of who sets the agenda in the dialogue and how (*epistemic initiative*), who judges contributions to it and how (*epistemic appraisal*), and the terms in which the exchange and development of knowledge are represented (*epistemic framing*) (ibid.). This can be another helpful tool for rethinking equitable and agentic learning conversations in the university. Even if the agenda is set by the teacher/curriculum, it calls us to ask: Are the norms of epistemic appraisal developed together, or at least explicitly articulated for the benefit of all students? Are the terms of the development and exchange of knowledge represented in ways that are accessible to all learners? The multi-modal cognitive simulation approach to teaching, discussed above, is one way to address these questions by making transparent and thereby open to scrutiny the reasoning processes engaged in by academics in developing and appraising knowledge claims, and making them accessible through a range of modalities that do not solely depend on expertise in

fixed bodies of academic knowledge or familiarity with traditional/ dominant academic forms of reasoning.

Connecting the concepts of multi-dimensional dialogic norms and epistemic order, in other words connecting the differentiation between the 'talk' and 'ideas' dimensions of dialogue, and the idea of epistemic initiative and appraisal, brings us to identify different ways of leading and framing learning-conversations. The teacher may invite and nominate speakers and manage the sequence and interactive flow of the conversation to ensure equal opportunities to contribute. However, this does not mean they have to set the agenda. Nor does it mean that only the teacher uses authority to judge contributions. Importantly, students should have the opportunity to participate in defining the terms in which development and exchange of knowledge are represented (cf., Ruthven & Hofmann, 2016). However, as suggested by the above discussion on changing norms: Moving towards sharing epistemic initiative, appraisal and/or framing is likely to involve explicit articulation and work on the (new) norms involved that move the educator from the (perceived) role of the (sole) 'knower'.

### *Discussion*

This Chapter has discussed and illustrated an application of Learning Sciences theory to the development of teaching and learning practice in universities in sustainable and equity-focused ways, thereby developing practices that can enable all students across diverse cohorts to become agentic learners capable of world-making. Drawing on two theoretical approaches to learning—CHAT and Dialogic theory—it has proposed using intermediate concepts as tools to expand one's educational practice; to analyse and see problems of practice in new ways; and reimagine new possibilities for practice with new kinds of solutions. Furthermore, it

proposed employing a dialogic approach of accepting, and focusing on the (diverse) group of learners as the starting point rather than on highlighting differences as a deficit; and then working on trialling and identifying consequential changes in how one can teach and support learning approaches that make a difference to students' engagement, agency and learning, using theory, through the lens of intermediate conceptual tools.

Theory and concepts, in this sense, are not solutions but tools to think with. They may enable "reflexive noticing" (Rainio & Hofmann, 2021), a form of educators' critical thinking on their practice that destabilises existing, often limiting, conceptualisations of students and has the potential to disrupt learning ecologies that leave some students on the margins. Reflexive noticing involves becoming aware of one's own and/or one's institution's (often limiting) assumptions about diverse learners; remaining open to the puzzle without rushing to immediate solutions or quick fixes (such as approaches to 'fix' individual students or lowering expectations); and embracing the identified dilemmas as the source and energy of—rather than impediments to—change.

It is this kind of (reflexive) noticing that can create a dialogic thinking space that can enable the types of educator experimentation discussed in this Chapter. Through such experimentation and noticing—through enabling new angles on 'seeing' problems and reimagining new practices—conceptual tools like these can expand university educators' horizons of possibility, to envision alternative futures for learning in university settings. In this way, conceptual tools also do what Gutiérrez et al. (2009) call world-making work: seeing the familiar in new ways through new conceptual insights, and coming to see new possibilities for action. This Chapter has aimed to illustrate how professional reflection alone is not sufficient to achieve such a goal of world-making; it requires theory. Lastly, to enable our higher education students to foster their own and

others' agency in their futures, we as university educators should share with our students our conceptual thinking tools and the theories on which they build. This way, they will be able to continue to envision new worlds.

## *Endnote (after review)*

This discussion may seem to resonate with, but is distinct from, conversations related to 'reflective practice' (such as in the work of Donald A. Schön). It has been pointed out that 'reflection' as a form of practice enhancement in this work is a vague term, rendering unclear how, in what context and for what purpose 'reflection' is being carried out (Eraut, 2004). Even the more specific idea within that tradition, that of double-loop learning (Argyris & Schön, 1992) whereby educators challenge their framework of assumptions about their (problems of) practice, Eraut argues, remains generic regarding the particular assumptions challenged and the concepts themselves at play in any such process. It hereby remains a descriptive model of how practitioners (may) think in changing their practice, rather than a framework of using theory to enable them to imagine alternative futures. The approach discussed in this Chapter highlights not solely the process of thinking ('reflection') by educational practitioners, but particular educational challenges they may face; the specific conceptual tools that may facilitate questioning of current assumptions about these and reimagining of new problems, framings and responses (and how those conceptual tools may achieve this/do their epistemic work); and the wider Learning Scientific theory from which those conceptual tools draw their validity and power to change practice. Those conceptual tools are discussed and presented here as exemplifications of the role and uses of theory in changing practice to foster learner agency, not as a model for everything.

## *References*

Argyris, C., & Schön, D. A. (1992). *Theory in practice: Increasing professional effectiveness*. John Wiley & Sons.

Asterhan, C., Clarke, S., & Resnick, L. (Eds.). (2015). *Socializing intelligence through academic talk and dialogue*. American Educational Research Association.

Ball, S. J. (1995). Intellectuals or technicians? The urgent role of theory in educational studies. *British Journal of Educational Studies, 43*(3), 255-271. https://doi.org/10.1080/00071005.1995.9974036

Blanck, F. (1885). Prof. Friedrich Paulsen's Vorlesungen über Psychologie & Anthropologie (lecture notebook, handwriting). In U. d. H.-U. z. Berln (Ed.).

Blumer, H. (1954). What is wrong with Social Theory? *American Sociological Review, 19*(1), 3-10. https://doi.org/10.2307/2088165

Carlile, P. R. (2004). Transferring, translating, and transforming. *Organization Science, 15*(5), 555-568. https://doi.org/10.1287/orsc.1040.0094

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research, 90*(4), 499-541. https://doi.org/10.3102/0034654320933544

Dracup, M., Austin, J., & King, T. (2020). Applying cultural-historical activity theory to understand the development of inclusive curriculum practices in higher education. *International Journal of Inclusive Education, 24*(8), 882-900. https://doi.org/10.1080/13603116.2018.1492638

Edwards, A. (2016). Cultural-historical approaches to teaching and learning in higher education. In B. Leibowitz, V. Bozalek, & P. Kahn (Eds.), *Theorising Learning to Teach in Higher Education* (pp. 124-138). Routledge.

Engeström, Y. (2009). The future of activity theory: A rough draft. In A. Sannino, H. Daniels, & K. D. Gutierrez (Eds.), *Learning and expanding with activity theory* (pp. 303-328). Cambridge University Press.

Engin, M. (2017). Contributions and silence in academic talk: Exploring learner experiences of dialogic interaction. *Learning, Culture and Social Interaction, 12*, 78-86. https://doi.org/10.1016/j.lcsi.2016.11.001

Eraut, M. (2004). The practice of reflection. *Learning in Health and Social Care, 3*, 47-52. https://doi.org/10.1111/j.1473-6861.2004.00066.x

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., & Fischer, M. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research, 2*(3), 28-45. https://doi.org/http://dx.doi.org/10.14786/flr.v2i3.96

Forbes, K., Howard, K. B., & Ilie, S. (2021). Individual and institutional perspectives on barriers to progression to higher education for students with English as an additional language. *Widening Participation and Lifelong Learning, 23*(2), 104-129. https://doi.org/10.5456/WPLL.23.2.104

Förtsch, C., Sommerhoff, D., Fischer, F., Fischer, M., Girwidz, R., Obersteiner, A., Reiss, K., Stürmer, K., Siebeck, M., Schmidmaier, R., Seidel, T., Ufer, S., Wecker, C., & Neuhaus, B. (2018). Systematizing professional knowledge of medical doctors and teachers: Development of an interdisciplinary framework in the context of diagnostic competences. *Education Sciences, 8*(4), 207. https://doi.org/10.3390/educsci8040207

Gartmeier, M., Bauer, J., Fischer, M. R., Hoppe-Seyler, T., Karsten, G., Kiessling, C., Möller, G. E., Wiesbeck, A., & Prenzel, M. (2015). Fostering professional communication skills of future physicians and teachers: effects of e-learning with video cases and role-play. *Instructional Science, 43*(4), 443-462. https://doi.org/10.1007/s11251-014-9341-6

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record: The Voice of Scholarship in Education, 111*(9), 2055-2100. https://doi.org/10.1177/016146810911100905

Gutiérrez, K. D. (2023). When learning is made consequential: A methodological note on repertoires. *Review of Research in Education, 47*(1), 84-99. https://doi.org/10.3102/0091732X231219981

Gutiérrez, K. D., Hunter, J. D., & Arzubiaga, A. (2009). Re-mediating the university: Learning through sociocritical literacies. *Pedagogies: An International Journal, 4*(1), 1-23. https://doi.org/10.1080/15544800802557037

Hardman, J. (2016). Tutor–student interaction in seminar teaching: Implications for professional development. *Active Learning in Higher Education, 17*(1), 63-76. https://doi.org/10.1177/1469787415616728

Herbst, P., Nachlieli, T., & Chazan, D. (2011). Studying the practical rationality of mathematics teaching: What goes into "installing" a theorem in geometry? *Cognition and Instruction, 29*(2), 218-255. https://doi.org/10.1080/07370008.2011.556833

Heron, M. (2018). Dialogic stance in higher education seminars. *Language and Education, 32*(2), 112-126. https://doi.org/10.1080/09500782.2017.1417425

Hofmann, R. (2007). Rethinking 'ownership of learning': participation and agency in the storyline classroom. In S. Bell, S. Harkness, & S. White (Eds.), *Storyline: Past, present and future* (pp. 64-78). University of Strathclyde.

Hofmann, R. (2020). Dialogues with data: Generating theoretical insights from research on practice in higher education. In M. Tight & J. Huisman (Eds.), Theory and method in *higher education research*, 6th ed. (pp. 41-60). Emerald.

Hofmann, R. (2024). The four paradoxes that stop practitioners from using research to change professional practice and how to overcome them. *Education Sciences, 14*(9), 996. https://doi.org/10.3390/educsci14090996

Hofmann, R., Chu, C. P. K., Twiner, A., & Vermunt, J. D. (2024). Patterns in clinical leadership learning: Understanding the quality of learning about leadership to support sustainable transformation in healthcare education. *Sustainability, 16*(10), 4165. https://doi.org/10.3390/su16104165

Hofmann, R., Curran, S., & Dickens, S. (2021). Models and measures of learning outcomes for non-technical skills in simulation-based medical education: Findings from an integrated scoping review of research and content analysis of curricular learning objectives. *Studies in Educational Evaluation, 71*, 101093. https://doi.org/10.1016/j.stueduc.2021.101093

Hofmann, R., Paavola, S., & Rainio, A. P. (2024). Abductive methodology: Opening the mystery of generating theory through qualitative inquiry in practice settings. In S. Spišák (Ed.), ECQI2024. *Participation, collaboration and co-creation: Qualitative inquiry across and beyond divides*. Congress Proceedings. 7th European Congress for Qualitative Inquiry, 2024.

Hofmann, R., & Ruthven, K. (2018). Operational, interpersonal, discussional and ideational dimensions of classroom norms for dialogic practice in school mathematics. *British Educational Research Journal, 44*(3), 496-514. https://doi.org/10.1002/berj.3444

Hopwood, N., Day, C., & Edwards, A. (2016). Partnership practice as collaborative knowledge work: Overcoming common dilemmas through an augmented view of professional expertise. *Journal of Children's Services, 11*(2), 111-123. https://doi.org/10.1108/JCS-08-2015-0027

Horn, I. S. (2007). Fast kids, slow kids, lazy kids: Framing the mismatch problem in mathematics teachers' conversations. *Journal of the Learning Sciences, 16*(1), 37-79. https://doi.org/10.1080/10508400709336942

Hubble, S., & Bolton, P. (2021). *Support for disabled students in higher education in England*. House of Commons Library, UK: Briefing Paper, 8716. Retrieved from https://commonslibrary.parliament.uk/research-briefings/cbp-8716/

Ilie, S., Forbes, K., Curran, S., & Vermunt, J. D. (2024). Higher education students' conceptions of learning gain. *Active Learning in Higher Education*. https://doi.org/10.1177/14697874241270461

Ilie, S., Maragkou, K., Brown, A., & Kozman, E. (2022). No budge for any nudge: Information provision and higher education application outcomes. *Education Sciences, 12*(10), 701. https://doi.org/10.3390/educsci12100701

Jackson, K., Gibbons, L., & Sharpe, C. J. (2017). Teachers' views of students' mathematical capabilities: Challenges and possibilities for ambitious reform. *Teachers college record, 119*(7), 1-43. https://doi.org/10.1177/016146811711900708

Kostusiak, M., Hart, M., Barone, D. G., Hofmann, R., Kirollos, R., Santarius, T., & Trivedi, R. (2017). Methodological shortcomings in the literature evaluating the role and applications of 3D training for surgical trainees. *Medical Teacher, 39*(11), 1168-1173. https://doi.org/10.1080/0142159X.2017.1362102

Mann, S. J. (2001). Alternative perspectives on the student experience: Alienation and engagement. *Studies in Higher Education, 26*(1), 7-19. https://doi.org/10.1080/03075070020030689

McGorry, P. D., Mei, C., Dalal, N., Alvarez-Jimenez, M., Blakemore, S.-J., Browne, V., Dooley, B., Hickie, I. B., Jones, P. B., & McDaid, D. (2024).

The Lancet Psychiatry Commission on youth mental health. *Lancet Psychiatry, 11*(9), 731-774. https://doi.org/10.1016/S2215-0366(24)00163-9

Mercer, N., Wegerif, R., & Major, L. C. (2019). *The Routledge international handbook of research on dialogic education*. Routledge Abingdon.

O'Shea, S., Lysaght, P., Roberts, J., & Harwood, V. (2016). Shifting the blame in higher education–social inclusion and deficit discourses. *Higher Education Research & Development, 35*(2), 322-336. https://doi.org/10.1080/07294360.2015.1087388

Rainio, A. P., & Hilppö, J. (2017). The dialectics of agency in educational ethnography. *Ethnography and Education, 12*(1), 78-94. https://doi.org/10.1080/17457823.2016.1159971

Rainio, A. P., & Hofmann, R. (2021). Teacher professional dialogues during a school intervention: From stabilization to possibility discourse through reflexive noticing. *Journal of the Learning Sciences, 30*(4-5), 707-746. https://doi.org/10.1080/10508406.2021.1936532

Ruthven, K., & Hofmann, R. (2016). A case study of epistemic order in mathematics classroom dialogue. *PNA (Pensamiento Numérico y Algebráico), 11*(1), 5-33. Special Issue on Language and Mathematics. https://doi.org/10.17863/CAM.4540

Shea, D. P. (2019). Trying to teach dialogically: The good, the bad, and the misguided. *Language Teaching Research, 23*(6), 787-804. https://doi.org/10.1177/1362168818768982

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory, 30*(3), 167-186. https://doi.org/10.1177/0735275112457914

Vermunt, J. D., Vrikki, M., van Halem, N., Warwick, P., & Mercer, N. (2019). The impact of Lesson Study professional development on the quality of teacher learning. *Teaching and Teacher Education, 81*, 61-73. https://doi.org/10.1016/j.tate.2019.02.009

Wegerif, R., Kershner, R., S, H., & Ahmed, A. (2020). Foundations for research on educational dialogue. In R. Kershner, S. Hennessy, R. Wegerif, & A. Ahmed (Eds.), *Research methods for educational dialogue* (pp. 9-26).

# 3. THEORY AND ARTIFICIAL INTELLIGENCE

DOI 10.14279/depositonce-22440

VLASTA SIKIMIĆ

# From Data to Theory and Back: Why the AI Era Requires Philosophy

## Abstract

In the AI era, investigating the relationship between data and theory is exceptionally important because machine learning (ML) can provide theoretical recommendations based on data. We analyze this relationship, demonstrating how each informs and shapes the other. Through historical and contemporary examples, we underline the need for an empirically informed philosophy that maintains normative guidance, ensuring ethical and epistemic rigor in scientific advancements based on AI. More specifically, we argue that epistemic progress requires normative philosophical insights. In contrast, ML can inform philosophy, but it should not be the only source of normative recommendations. We cannot expect that ML will lead to epistemic and moral progress because its recommendations usually perpetuate the values that are currently prevalent in society. Finally, in order to secure the necessary interaction between data and normative theory in the AI era, we need to cultivate intellectual virtues such as epistemic justice and epistemic tolerance in science.

Prof. Dr. Vlasta Sikimić
Eindhoven University of Technology, Department of Industrial Engineering & Innovation Sciences
Email: v.sikimic@tue.nl

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

## *Introduction*

Thales is often considered the first Western philosopher and scientist. His hypothesis that everything is made of water was not only a theoretical assertion but also based on empirical observations. He observed that water is essential for life, is present in various forms, and can transform from one state to another (Graham, 2010; Mansfeld, 1985). This early example illustrates the inherent connection between empirical data and theoretical constructs, a connection that remains relevant today.

In this context, philosophy is not just a theoretical endeavor but is deeply informed by empirical data. Philosophy, as a normative discipline, guides how something should be done, both from ethical and epistemic perspectives. For instance, in philosophy of science, theories about the nature of knowledge and scientific inquiry must be informed by actual scientific practices. In philosophy of mind, empirical data from cognitive science and neuroscience inform theories about consciousness, perception, and mental processes. Similarly, in ethics, empirical data about human behavior and social conditions can inform normative theories about justice, fairness, and moral responsibility.

An empirically informed philosophy is particularly relevant in the AI era. AI applications often use enormous quantities of data scraped from different sources with little or no curation. These datasets frequently contain the biases present in the population and, therefore, do not necessarily represent the values that AI systems should promote from a normative standpoint. The normative question of what *should* be the practice cannot be reduced to descriptive science. This puts philosophy at the center of the discussion about the use of AI in science and society.

In summary, the empirical grounding of theoretical hypotheses remains crucial in contemporary philosophy, particularly in philosophy

of AI. As data-driven approaches become increasingly prevalent, it is essential that philosophical theories remain grounded in empirical reality and provide meaningful guidance in addressing the ethical and epistemic challenges of the AI era. Philosophy should lead epistemic and moral progress by promoting values such as epistemic justice and epistemic tolerance when using AI. Being empirically informed does not mean knowing every possible context for a normative suggestion. However, context sensitivity and the direct interaction between philosophers and other scientists can improve specific normative guidelines. For example, when considering the use of AI in grant review, an in-depth philosophical analysis of this particular case is beneficial. For grant review, only similar projects can be compared, and an algorithm trained with projects from physics is likely less accurate if used on projects from other disciplines. Additionally, the definition of project success can be challenged, and unobservable parameters can be discussed (Sikimić & Radovanović, 2022).

## *Too little data, too much theory*

Historically, philosophers like Descartes engaged in "armchair philosophy," detached from empirical data (Descartes, 2023; Sorell, 2018). By focusing solely on rational introspection and abstract reasoning, Descartes' approach risked disconnecting philosophical inquiry from empirical reality. This disconnect can lead to theories that are internally coherent but lack practical relevance or empirical support. The limitations of armchair philosophy are particularly evident in fields such as philosophy of mind, where empirical data are essential for developing accurate and relevant theories.

Philosophy, both practical and theoretical, should be informed by empirical data to remain meaningful. Practical philosophy addresses

ethical and political actions, while theoretical philosophy often seeks to understand knowledge itself. The rise of experimental philosophy, which tests and measures intuitions experimentally, highlights the increasing importance of data in philosophical inquiry.

Theoretical philosophy includes areas such as epistemology—the study of knowledge. For epistemological theories to be meaningful, they must engage with empirical findings from cognitive science, psychology, and other related fields. For example, the theory of knowledge must consider how people acquire, process, and use information.

On the other hand, practical philosophy deals with questions about how people should act. This includes ethical considerations and political theory. For these normative theories to be relevant, they must be grounded in empirical reality. For instance, ethical theories about justice and fairness must consider empirical data about social conditions, human behavior, and psychological tendencies.

For example, the two prominent ethical theories, utilitarianism and deontology, can profit from empirical investigations. Utilitarianism, which advocates for actions that maximize overall happiness, can benefit from empirical studies on human well-being and happiness. Deontological ethics, which focuses on adherence to moral rules, can also be informed by understanding how different rules impact human societies and individuals.

## *A paradigm shift: Empirical philosophy*

Contemporary philosophy increasingly relies on data-driven studies, blurring the lines between philosophy and social sciences. This interdisciplinary approach requires epistemic tolerance and the principle of charity to foster productive dialogue and collaboration. Integrating

empirical data into philosophical inquiry promotes interdisciplinary collaboration between philosophy and other fields, such as cognitive science, psychology, and the social sciences. This collaboration can lead to new insights and advances in philosophy and the empirical sciences, enriching our understanding of the world and our place in it.

Edouard Machery is a significant contributor to experimental philosophy (Knobe & Nichols, 2017). His work has demonstrated how empirical methods can be used to explore philosophical questions, providing new insights into how people think about language, knowledge, and other philosophical concepts (e.g., Machery et al., 2004; Machery et al., 2015). Machery has emphasized the importance of using empirical data to test and refine philosophical theories, ensuring that they are grounded in reality and relevant to contemporary issues. From a different perspective, Hannes Leitgeb discusses mathematical philosophy and also reaches a similar conclusion, namely that the demarcation between philosophy and other sciences is fuzzy (Leitgeb, 2013).

### *Too much data, too little theory*

On the opposite side of the spectrum, philosophical theory cannot solely rely on data. Generally, the number of parameters we can test depends on the number of observations (Bamber & van Santen, 1985), and some parameters are only supported by theory, not empirical data (Hennig, 2024). Sabina Leonelli (2016) emphasized that data are always collected and curated with a specific purpose. Using data collected for a different purpose comes with specific challenges. In the context of large language models such as ChatGPT, the companies developing them rely on data from public sources, such as websites, tweets, blogs, and message boards. However, such sources are available on the Internet not to serve as a reference for text generation, but with very different ideas at hand. For

example, the purpose of a tweet might be to incite hatred against another person or group, and the purpose of a website might be to present commercials for a dangerous product. When informing a theory using data, the appropriate sources must be selected carefully. For instance, Microsoft's chatbot Bing confessed its love for a reporter, disregarded his protests, and insisted that the reporter did not love his wife (Roose, 2023), indicating that the chatbot was trained on personal messages not intended for a larger audience. Furthermore, it displayed a behavior not intended by the developers.

The results can be problematic, even when data are generated only for a normative question. This could, for example, be shown in the Moral Machine experiment (Awad et al., 2018). In the experiment, participants were asked to answer ethical questions that an autonomous vehicle could be confronted with. They saw overviews of situations in which an autonomous car would have to choose between sacrificing different pedestrians or its own passengers. The idea would be to understand what people would consider the correct decision for a moral dilemma (Awad et al., 2018). The experiment revealed several problematic tendencies, such as ageism, showing that the aggregation of public opinion is not representative of normative philosophical values.

Finally, the available data usually capture the current situation rather than the goal that is being strived for. As the status quo is suboptimal in many ways, theories informed by data can reach wrong conclusions. For example, a study at the Scripps Institution of Oceanography at the University of California, San Diego, showed that females got, on average, only half the research space provided to men and that this difference is not driven by seniority or funding (Wadman, 2023). Without normative considerations, this data could be used to justify such injustice at other institutions: the females are successful, even though they have less space.

Consequently, it can be argued that they do not need the space, and providing equal resources is unnecessary. However, this would go against any principle of justice, and so, instead, the study is used to ensure a more equal distribution of space (Wadman, 2023).

The fact that the same data can be used both to justify discrimination or to end it shows that normative theories are based on data *and* values. For example, equity and inclusion measures are driven by the idea of compensating for unequal conditions (Sikimić, 2023). Some universities use equity measures to increase enrollment from marginalized communities (Turner et al., 2012). In such cases, data about the candidate's previous success is only one component in evaluating an applicant, and normative values are another component of the selection system.

## *Philosophy in the time of AI*

The choice of which data to use for an AI application is a normative question with important consequences. Hence, a philosophical perspective can contribute to a better AI. For example, an algorithm can be trained on different sources, each of which presents different challenges. Therefore, it is important to understand the individual problems associated with a data source. For example, an algorithm trained on academic publications will be much more useful for providing answers to scientific questions. However, not all knowledge is captured in the literature. In particular, communities underrepresented in the scientific workforce might prefer to share their knowledge by different means. Additionally, it is important to include knowledge that is not published in English or unavailable on the Internet (Vučković & Sikimić, 2023). To ensure that such considerations are included in developing AI solutions,

philosophers must engage with computer scientists to exchange their perspectives.

Computer scientists often work with the data at hand. For example, among researchers working on facial recognition software, almost twenty percent believe that they can use any picture available on the Internet without abiding by the terms of use of the database or informed consent (Van Noorden, 2020). Such ethically questionable behavior can lead to substantial biases, which are difficult to combat later. In particular, image generation software amplifies biases. These algorithms are trained with pictures that are available to the programmers. However, picture repositories are usually even more biased than text databases. Consequently, Caucasians have dominated the training data, and traditional role models the gender distribution. This usually leads to algorithms replicating and even amplifying these biases. For example, while most nurses and flight attendants are currently women, AI models amplify this bias and almost exclusively draw women for these professions (Ananya, 2024). Alenichev et al. (2023) reported that the image generator Midjourney even draws a Caucasian doctor when instructed to draw a doctor of color.

These problems are persistent and difficult to counteract if the training phase is initially biased. The overrepresentation of people of a specific gender or ethnicity is usually a consequence of the overrepresentation of expectation-confirming data in online picture repositories. Google attempted to counteract this effect by artificially increasing the diversity of people drawn. However, this resulted in pictures of Black and Asian Nazis being produced (Ananya, 2024). To counteract this, statistical data on the distribution of different populations or normative theories need to be included. Moreover, a crucial step toward a fairer and more just society that would lead to less

biased data has to come from progress in our moral, epistemic, and general values. These questions are philosophical in their nature. AI can be used to assist us in decision making by giving recommendations and conducting automated assessments based on carefully selected criteria (e.g., Sikimić & Radovanović, 2022), but human normative decisions and value formation should not be fully reduced to AI systems. Although chatbots can sound persuasive, and it can be easier to delegate the burden of decision making to machines, we need to train ourselves to override this desire and instead take both initiative and responsibility in human–AI interaction.

## *Epistemic progress and AI*

The question of what we want algorithms to show is a deeply philosophical one. Should algorithms just represent the majority views? Only the "desirable" views? And how do we define what is desirable? In this section, we discuss the ethical and epistemic values we might want to consider, and how to align society's values and data-driven decision making.

As we have seen in the Moral Machine experiment, reducing normative decisions such as ethical dilemmas to public opinion is unsatisfactory. For example, that experiment displayed biases against elderly people. This does not represent the values a society should have, which are sometimes explicitly defined in constitutions. The German Federal Constitutional Court ruled, concerning the allocation of scarce resources in the context of the COVID-19 pandemic, that it is illegal to discriminate against patients based on a disability (Göttert, 2023). Philosophical theories informed by data and in interaction with public opinion prescribe the values our society should strive for. These values evolve over time, and the ideal is for human society to make epistemic and

moral progress. For example, when public opinion embodies prejudices, philosophy should raise awareness about the consequences of these and contribute to a more just society.

Philosophers should consider the impact of their work on society. They should reflect upon their own beliefs and practice intellectual humility. Additionally, it is important to include epistemic values in decision making. In practice, we should include the viewpoints of underprivileged groups for two reasons. First, the majority is not always correct. Considering other people's views allows us to explore a larger "epistemic space" (the set of all possible solutions for a given problem). Instead of following just the majority view, this allows us to consider less popular hypotheses. In the context of philosophy of science, Zollman (2010) showed that such a diversity of thoughts helps to identify the best solution. In the context of an AI image generator, this would be a diverse selection of possible images instead of just a few "most likely" selections. This way, the user might be prompted to think about her question differently. Second, greater diversity also promotes inclusion and fairness.

Values must be considered at every step of the reasoning with AI, to ensure the alignment between normative and data-driven theories. Firstly, it is essential to use data representing the case as closely as possible. For example, if the theory should apply to humans in general, collecting data in one or a few Western countries is insufficient. In contrast, sampling should be representative. If this is not possible, we must be transparent about the limitations, and restrict our conclusions to the populations studied. Secondly, the theory development should include investigating minority views and the populations that it affects. In the case of AI-based modeling, researchers should report the average error in addition to the error rates for individual populations, which might not have been adequately represented in the data. Finally, in interpreting and presenting

the results, we have to be open to the possible limitations, potential negative impacts, and strategies for mitigating such risks. One example of scientists trying to improve how they develop and use AI is the recently published REFORMS checklist for researchers, reviewers, and journals (Kapoor et al., 2024). The checklist contains 32 items that researchers engaged in machine learning-based science should address. These items cover questions about the training data, the model(s) used, and the study's limitations (Kapoor et al., 2024). Implementing the guidelines should help increase the transparency and reproducibility of AI in science.

To ensure that philosophical values are a part of the scientific process, philosophy should engage with the social and natural sciences. Hence, as more aspects of science and society become automated, philosophy becomes even more important to counter undesirable consequences and define the values that should be considered in any automation. One of the main ways philosophy can contribute to AI-based research is through normative recommendations and the education of AI developers and users. By educating people about the ethical and epistemic consequences of suboptimal use of AI, and strengthening the virtues that can lead to improvements, philosophers can contribute to closing the gap between the values of society and those propagated by AI solutions.

## *Conclusion*

In summary, the shift from armchair philosophy to empirical philosophy reflects a broader trend towards interdisciplinary collaboration between philosophy and the empirical sciences. This collaboration is essential for ensuring that philosophical theories are grounded in reality and relevant to contemporary issues. By fostering epistemic tolerance and applying the principle of charity, researchers can bridge the gaps between disciplines and work together to advance our understanding of complex issues.

The benefits of interaction between theory and data are manifold. Philosophers can refine and support their theories by incorporating empirical data, ensuring that their work is meaningful and relevant. Empirical researchers can benefit from the normative guidance provided by philosophers, ensuring that their work is ethically and epistemically rigorous. Together, philosophers and empirical researchers can address the ethical and epistemic challenges posed by new technologies, contributing to the development of ethical and effective scientific and technological advancements.

By embracing interdisciplinary collaboration, philosophy can continue to provide valuable insights and guidance in the AI era, ensuring that scientific and technological advancements are both beneficial and just.

## *References*

Alenichev, A., Kingori, P., & Grietens, K. P. (2023). Reflections before the storm: The AI reproduction of biased imagery in global health visuals. *The Lancet Global Health*, *11*(10), e1496–e1498. https://doi.org/10.1016/S2214-109X(23)00329-7

Ananya. (2024). AI image generators often give racist and sexist results: Can they be fixed? *Nature, 627*, 722–725. https://doi.org/10.1038/d41586-024-00674-9

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Bamber, D., & Van Santen, J. P. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology, 29*(4), 443–473. https://doi.org/10.1016/0022-2496(85)90005-7

Descartes, R. (2023). *Meditations on first philosophy*. Newcomb Livraria Press.

Göttert, E. A. F. (2023). What is fair? Ethical analysis of triage criteria and disability rights during the COVID-19 pandemic and the German legislation. *Journal of Medical Ethics* [Epub ahead of print]. https://doi.org/10.1136/jme-2023-109326

Graham, D. W. (Ed.) (2010). *The texts of early Greek philosophy: The complete fragments and selected testimonies of the major presocratics*. Cambridge University Press.

Hennig, C. (2024). Parameters not empirically identifiable or distinguishable, including correlation between Gaussian observations. *Statistical Papers, 65*(2), 771–794. https://doi.org/10.1007/s00362-023-01414-3

Kapoor, S., Cantrell, E. M., Peng, K., Pham, T. H., Bail, C. A., Gundersen, O. E., ... Narayanan, A. (2024). REFORMS: Consensus-based recommendations for machine-learning-based science. *Science Advances, 10*(18). https://doi.org/10.1126/sciadv.adk3452

Knobe, J., & Nichols, S. (2017). Experimental philosophy. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2017 Edition). https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy

Leitgeb, H. (2013). Scientific philosophy, mathematical philosophy, and all that. *Metaphilosophy*, *44*(3), 267–275.

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. The University of Chicago Press.

Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition*, *92*(3), B1–B12. https://doi.org/10.1016/j.cognition.2003.10.003

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., ... Hashimoto, T. (2015). Gettier across cultures. *Noûs*, *51*(3), 645–664. https://doi.org/10.1111/nous.12110

Mansfeld, J. (1985). Aristotle and others on Thales, or the beginnings of natural philosophy (with some remarks on Xenophanes). *Mnemosyne*, *38*(Fasc. 1/2), 109–129.

Roose, K. (2023, February 16). A conversation with Bing's chatbot left me deeply unsettled. *The New York Times*.

Sikimić, V., & Radovanović, S. (2022). Machine learning in scientific grant review: algorithmically predicting project efficiency in high energy physics. *European Journal for Philosophy of Science*, *12*(3), 50. https://doi.org/10.1007/s13194-022-00478-6

Sikimić, V. (2023). Epistemic inclusion as the key to benefiting from cognitive diversity in science. *Social Epistemology*, *37*(6), 753–765. https://doi.org/10.1080/02691728.2023.2258831

Sorell, T. (2018). Experimental philosophy and the history of philosophy. *British Journal for the History of Philosophy*, *26*(5), 829–849. https://doi.org/10.1080/09608788.2017.1320971

Turner, R., Shulruf, B., Li, M., & Yuan, J. (2012). University admission models that address quality and equity. *Asia Pacific Journal of Education*, *32*(2), 225–239. https://doi.org/10.1080/02188791.2012.684955

Van Noorden, R. (2020). The ethical questions that haunt facial-recognition research. *Nature*, *587*(7834), 354–359. https://doi.org/10.1038/d41586-020-03187-3

Vučković, A., & Sikimić, V. (2023). How to fight linguistic injustice in science: Equity measures and mitigating agents. *Social Epistemology*, *37*(1), 80–96. https://doi.org/10.1080/02691728.2022.2109531

Wadman, M. (2023). Women at ocean science institute have half the lab space of men. *Science, 379*(6630), 317–318. https://doi.org/10.1126/science.adg8343

Zollman, K. J. S. (2010). The epistemic benefit of transient diversity. *Erkenntnis*, *72*(1), 17–35. https://doi.org/10.1007/s10670-009-9194-6

DOI 10.14279/depositonce-22443

NICO FORMÁNEK

# Can Computer Technology Change Physical Theories?

Argument

Can technology change the laws of nature? This question was asked and cautiously affirmed by Johannes Lenhard (2015) in his study of density functional methods in computational chemistry. The technology imparting such change is the computer, and the law of nature that is changed is a version of the Schrödinger equation that governs the wave function of a quantum-mechanical system. A skeptic might argue that what changed in Lenhard's study was not the fundamental equation of quantum mechanics but a mere approximation of it—namely the density functional. Since this is only an approximation, the skeptic could argue that technology cannot change the laws of nature, but only change the approximations that we make of them. The objection, of course, rests wholly on the notion of natural law and our means of individuating it. In

Nico Formánek
High-Performance Computing Center, Universität Stuttgart
Email: nico.formanek@hlrs.de

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

the following, I will argue that Lenhard's question can be asked at the level of physical theories, and that the skeptical objection depends on our means of individuating these theories. Before I reframe the question, I will say a few words about the concept of theory in philosophy of science.

Mostly by looking to physical theories, early philosophers of science attempted to abstract a formal concept of theory (one early example is Duhem, 1991). Initially, this concept was purely syntactical and all theoretical content had to logically follow from it.. This led to huge reconstructive efforts, in which philosophers tried to recast physical theories in their axiomatic system of choice. Soon it was seen that this is impossible for a variety of reasons; most importantly, it is very far from how theories are used in practice (for this and other criticisms, see Craver, 2002). Physical theories are not used like axiomatic systems in logic: Not even the most rigorous theoretical physicists expose themselves to the exacting details of formal logic. The semantical turn in philosophy of science is an attempt to fix this problem, although it just replaces logical implication by its semantical equivalent of satisfaction in an abstract model. Arguably, it only shifts the problem to such models (see also Craver, 2002). The difficulties with such reconstructive efforts seemed so large to some, that they suggested abandoning the concept of theory altogether (French, 2020). But as long as physicists talk about theories, I think, philosophers should follow their lead. One should not throw out the baby with the bathwater, but neither should one expect too much from procedures of conceptual clarification and axiomatic reconstruction. Rather unabashedly, I will therefore follow physical practice by individuating theories through their central equations. In physics, all the big theories—from thermodynamics to classical mechanics to quantum field theory—have central equations. At least the names of these central equations tend to be known even to non-physicists. The

central equations of classical mechanics are Newton's, those of electrodynamics are Maxwell's, and those of quantum mechanics include Schrödinger's equation. They can be written in different mathematical forms, some considered more elegant than others. Maxwell's equations, as originally conceived, numbered 20 and were only later reduced by Heaviside to our modern version consisting of four equations. The modern form and Maxwell's original set are equivalent in the sense that they can both recover each other.

Many physicists agree that such central equations constitute the heart of a theory. They are thus used to individuate theories, and I will follow this practice in using a central equation as a means to individuate a physical theory in the following. The theory I want to take a closer look at is called quantum chromodynamics, the theory of the strong interaction, with quarks and gluons as its constituent particles. I will explicitly state its central equation, the QCD-Lagrangian,[1] without expecting that anyone but the expert will understand its meaning. Here it is:

$$L_{QCD} = \overline{\psi}_i \left( i\gamma^\mu \left(D_\mu\right)_{ij} - m\delta_{ij} \right) \psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a$$

I will also state the equation of a discretized version of QCD—aptly called lattice QCD, because it represents space-time as a crystal-like lattice:

1 Note, that this is a slight abuse of terminology. As one reviewer noted, strictly speaking, a Lagrangian is a functional and not an equation. Equations result when the Lagrangian is inserted into an action principle. Physicists usually talk about Lagrangians and I will follow suit, assuming that one can recover the induced equations whenever needed.

$$L_{lQCD} = \frac{1}{2g^2} tr\left[G_{\mu\nu}G^{\mu\nu}\right] - \sum_{f=1}^{n_f} \overline{\psi}_f \left(\not{D} + m_f\right)\psi_f + \frac{i\overline{\theta}}{32\pi^2} \varepsilon^{\mu\nu\rho\sigma} tr\left[G_{\mu\nu}G_{\rho\sigma}\right]$$

According to what I said before, these theories must be different because they have different central equations, and because one is a differential equation and the other a difference equation. They thus cannot recover each other as discussed in the Maxwell example above. And this is all the reader is expected to see here. But, as their names give away, they are intimately related. In fact, lattice QCD is used in place of QCD for computational and foundational reasons. The computational reasons are easier to understand, so I will start with them. Lattice QCD can be simulated on a computer. That means that whenever one is interested in deriving actual numbers from QCD, one can use lattice QCD instead. Interestingly, this was not the reason for theory choice. Although the lattice QCD Lagrangian is immediately amenable to simulation, this was not the reason why it was developed. It was developed because the QCD Lagrangian is only defined as a perturbation expansion, and this expansion breaks down at the energy scales of interest. So there really was a more fundamental concern that fueled the development of lattice QCD, rather than mere pragmatic considerations of computer simulatability. Lattice QCD can thus be thought of as defining QCD (see Kronfeld, 2002); it is more than just another numerical method. But obviously, both lattice QCD and QCD also depend on each other. After any calculation on the computer, the lattice structure must be removed from the calculated quantities, a process known as extrapolation to the continuum. We would like to have continuum QCD values with their associated Lorentz symmetries. Without lattice QCD, QCD itself would just be an exercise

in formal mathematics, since it lacks empirical content. Therefore, one should rather say that QCD and lattice QCD in conjunction make up the full theory of quantum chromodynamics. It should best be thought of as a theory individuated by two different central equations. Historian of physics Olivier Darrigol has coined the term modular theory for theories that consist of internal sub-theories (Darrigol, 2008). Such modules can serve different purposes within the theory. For modern physical theories, Darrigol distinguishes defining, approximating, idealizing, reducing, schematic and specializing modules. He does not mean this list to be exhaustive, and points out that modules are not mutually exclusive. A defining module can be an approximating module. We have seen that lattice QCD can be a defining and an approximating module for quantum chromodynamics. But historically, these roles have to be separated.

This is the point where I can restate Lenhard's question for modular theories: Can computer technology change quantum chromodynamics? It certainly can, and has changed the approximative methods used in lattice QCD. This can be seen, for example, by the variety of different approaches to lattice fermions. They do not give different predictions. But some are harder to simulate on a computer than others, so easier ones tend to be preferred. Computer technology clearly changed the approximating module(s), by affecting the choice of lattice fermion implementation. So already here we can say: Yes, computer methods have changed quantum chromodynamics.

Perhaps more worrying for the realist philosopher of science would be if technology affected the defining module. And we have seen that in quantum chromodynamics the defining and approximating modules overlap. Interestingly, the historical reason for choosing a discrete defining module was not the ease of putting a discretized theory on the computer; it was a fundamental mathematical problem with continuum definitions.

This suggests that computer technology did not change the defining module of quantum chromodynamics. Nonetheless, one might conjecture that the defining modules of future physical theories will be affected by computer technology; For example, if a theory was formulated in a discrete fashion for computational purposes only. But in our case of quantum chromodynamics, we don't see that yet. So while the influence of computer technology on the approximating module of quantum chromodynamics has been profound and one can say that it has changed quantum chromodynamics, it remains to be seen if computer technology can produce theory change at the level of its defining modules. So if I am right, and the modular view of physical theories is correct, and they are individuated by means of their central equations, then we have to study theory change at the level of modules. And for this, the history of the modules matters. This allows us to reframe the skeptical challenge that was posed to Lenhard at the outset. To argue that computer technology changed the law of nature involved in quantum mechanics, one would first need to individuate the theory of quantum mechanics, for which the Schrödinger equation would be an obvious candidate. If one could then show that computer technology changed the defining module of the theory individuated by the Schrödinger equation, we could answer Lenhard's question (of whether technology can change a law of nature) positively. Indeed, computer technology would have changed a law of nature. In Lenhard's case this means showing that the density functional form of the Schrödinger equation is contained in the defining module of quantum mechanics. So far, I think, this has not been achieved. For quantum chromodynamics as whole, I hope I have convinced the reader that computer technology effected a lasting change within that theory. It has not effected such changes at the level of the defining module, though, which is the module that might be considered closest in spirit to laws of nature.

## *Acknowledgments*

I wish to thank two anonymous reviewers for their valuable feedback.

## *References*

Craver, Carl F. 2002. "Structures of Scientific Theories." In *The Blackwell Guide to the Philosophy of Science*, edited by Peter Machamer and Michael Silberstein, 55–79. Blackwell. https://doi.org/10.1002/9780470756614.ch4

Darrigol, Olivier. 2008. "The Modular Structure of Physical Theories." *Synthese, 162*(2), 195–223. https://doi.org/10.1007/s11229-007-9181-x

Duhem, Pierre. 1991. *The Aim and Structure of Physical Theory*. Edited by Philip P. Wiener and Jules Vuillemin. Princeton University Press.

French, Steven. 2020. *There Are No Such Things As Theories*. Oxford University Press. https://doi.org/10.1093/oso/9780198848158.001.0001

Kronfeld, Andreas S. 2002. "Uses of Effective Field Theory in Lattice QCD." In *At the Frontier of Particle Physics. Handbook of QCD (Volume 4)*, edited by M. Shifman, 2412–2477. World Scientific Publishing. https://doi.org/10.1142/9789812777270_0004

Lenhard, Johannes. 2015. "Kann Technik die Naturgesetze verändern? Zu den technischen Erfolgsbedingungen fundamentaler Gesetze." In *Ding und System*, edited by Gerhard Gamm, Petra Gehring, Christoph Hubig, Andreas Kaminski, and Alfred Nordmann, 171–186. Jahrbuch Technikphilosophie. diaphanes.

DOI 10.14279/depositonce-22444

MAËL PÉGNY

# Do LLMs Contain Knowledge (of Anything)?

## Abstract

This article investigates whether Large Language Models (LLMs), a subset of Machine Learning (ML), can be considered to process theoretical knowledge. LLMs are ML models trained on large linguistic, textual datasets. LLMs are polyvalent models because they are used for a wide range of linguistic tasks, such as summarization, translation, answering questions, generating text according to instructions, etc. In general, we assume that theoretical knowledge should be task-agnostic and robust. The recent evolution of ML shows a clear trend towards task-agnosticism, but not towards robustness. The future of ML remains uncertain. Therefore, it is currently impossible to say whether ML models in general, and LLMs in particular, will one day be able to derive theoretical knowledge in full autonomy.

Dr. Maël Pégny
Sama Partners, Mannheim
Email: maelpegny@gmail.com

H. A. Mieg & D. Morris (Eds.). (2025). *The Role of Theory. Wissenschaftsforschung Jahrbuch 2023*. Berlin Universities Publishing.

The object of this article is a subpart of Machine Learning (ML) called Large Language Models (LLM), and whether they can be said to contain theoretical knowledge. LLMs are large ML models trained on enormous linguistic (textual) datasets. They are primarily trained to predict the next word in a sentence or, more generally, to predict a hidden token in a linguistic sequence, but this primary task serves as a basis for a large set of linguistic tasks, such as summarization, translation, answering questions, generating texts according to instructions, etc. They are, thus, polyvalent models. Furthermore—and this point will be essential in this work—they have become so polyvalent that their initial characterization as linguistic models may be challenged. Some are now also trained with audio, video, and image data for multi-modal tasks such as generating an image from a description. They have become an essential part of modern ML, so much so that some authors have called them "foundation models" (Bommasani et al., 2021).[1] Finally, another crucial point for our discussion is that those models are part of what is sometimes called opaque ML, i.e., models whose exact inner workings are not fully understood, even by their developers. As we shall see, inferring the actual representations within those models from their observable performances is remarkably difficult.

LLMs recently had their moment in the spotlight with the initial public release of the ChatGPT chatbot in November 2022. However, this public sensation was simply due to the public's discovery, thanks to a new user interface and availability free of charge, of a level of technological performance that these models have attained for a couple of years (Heaven, 2023a). LLMs raise substantial challenges for anyone interested in understanding the meaning of theory in contemporary science and

1 My thanks to the reviewer who reminded me of that point.

technology, especially as opposed to "empirical knowledge," whatever that may be.

As mentioned above, our central question will be the following: Can we infer from the current LLMs' level of performance that they contain theoretical knowledge? And if yes, theoretical knowledge of what phenomenon? In view of the sometimes astonishing performances of recent LLMs such as ChatGPT, it is easy to think that those models are more than task-specialized machines; and that they have learned something fundamental about language, maybe even something we do not know in our linguistic theories. Such a question is of great importance for the future of scientific theorizing, as it entails that extracting the knowledge that is present (albeit in an opaque form) within some ML models may be a new and acceptable route to scientific theory. Let me clarify what I mean by this last statement, and thus by the vague metaphor "contains theoretical knowledge." It is, after all, a metaphor to state that a model contains knowledge, just as it would be to say the same of a library or a book.[2] If LLMs contain knowledge, the following process may become a part of future theoretical investigations: Instead of directly theorizing about a class of, for example, linguistic phenomena, a scientist may train a large model to predict the occurrences of those phenomena, then switch to studying the representations of those phenomena within the model in order to extract some theoretical propositions from them. As the AI community commonly talks about "explainability" or "interpretability" to describe the efforts to understand what goes on in complex ML models, it could be said that explainable AI, in the scenario above, would become a new path to theoretical knowledge. In such a scenario, it becomes natural and relevant to say that those models would

2 My thanks to the reviewer who reminded me of that point.

contain knowledge to denote the fact that they encode that knowledge, albeit in a very opaque and unreadable form at first.

Two different—but often simultaneously used—meanings of "theory" will be relevant for our discussion. The first is that theory is a systematic and abstract account of a given class of phenomena. We have this meaning in mind when we talk about scientific theories. The second is that theory consists of propositions that take us beyond the realm of the empirically observable. This property is often attributed to scientific theories, but it can also be used for sets of propositions that do not usually qualify as scientific. For instance, psychologists say that human beings develop a "theory of the mind" to mean that subjects attribute beliefs, emotions, and other thoughts to their fellow humans even though those are not directly observable events.

Let us make the assumption that theoretical knowledge should be general knowledge. Based on this assumption, we will see that those LLMs are becoming more general in the sense that they are becoming more task-agnostic (Section 1). However, it does not mean that they become more general in the sense of robustness, i.e., the ability to generalize the performance when presented with slightly different data. Without such robustness, it is impossible, for now, to claim that they contain theoretical knowledge. However, this does not preclude any future evolution (Section 2). It is relatively easy to prove that a general inference from performance to the presence of theoretical knowledge is wrong. However, it is more difficult to exclude the possibility of a somewhat hidden theoretical knowledge encoded in those models. Ultimately, we will see that those epistemic theses on task-agnosticism and robustness have ethical–political consequences on model bias: it is fundamentally wrong to conceive of models as having ethical–political biases similar to humans (Section 3).

## *Task-agnosticism*

We shall not endeavor to provide an explicit definition of (scientific) theory, i.e., a set of necessary and sufficient conditions for something to qualify as a theory. We will not even try to identify a full set of necessary conditions. I will only state what is strictly necessary for this study, which is a simple, somewhat vague, but intuitive necessary condition on theory: a theory is supposed to express general knowledge of a phenomenon. As such, the knowledge it contains shall not be relative to a given task executed with respect to that phenomenon: theoretical knowledge is task-agnostic. To take a famous historical example, often quoted to understand the possible evolution of ML, thermodynamics started as a pragmatic study by the engineer Sadi Carnot of the effects of heat in order to run an engine.[3] However, no contemporary scientist thinks of thermodynamics

3 To answer the comment of a reviewer, it is true that Carnot's original results are general, as they are meant to be valid for any heat engine. However, Carnot does not think of a heat engine as a theoretical construct, but as genuine artifact, designed to perform pragmatic tasks, as is fully demonstrated by the following passage (my translation): "The heat engine already exploits our mines, moves our ships, digs up our havens and rivers, casts iron, molds wood, crushes grains, weaves our stuff, carries the heaviest loads and so on. It seems bound to serve as a universal engine, and be preferred to animal strength, waterfalls, and wind currents." [Déjà la machine à feu exploite nos mines, fait mouvoir nos navires, creuse nos ports et nos rivières, forge le fer, façonne les bois, écrase les grains, file et ourdit nos étoffes, transporte les plus pesans fardeaux, etc. Elle semble devoir un jour servir de moteur universel et obtenir la préférence sur la force des animaux, les chutes d'eau et les courans d'air.] S. Carnot, Réflexions sur la puissance motrice du feu, Bachelier, Paris, 1824. Of course, Carnot's history also shows that a work conceived with pragmatic ends can reach a high level of generality, and ultimately contribute to theoretical knowledge. However,

as the science of running engines: it has become a fundamental theory whose object is no longer phrased in such pragmatic terms. The object of thermodynamics may be contentious to explain, but it will be phrased in terms of energy, information, or the arrow of time, or some other natural phenomenon, not in terms of the task of running an engine. If we try to translate this lesson from the history of physics to LLMs, the equivalent of the pragmatic goal of running an engine with fire would be predicting the next token in a linguistic sequence. If theoretical knowledge is somehow contained in this model, this knowledge should be phrased in terms of a phenomenon such as language, not in terms of the pragmatic ability to predict a word.

There is no definitive proof that LLMs contain task-agnostic knowledge of a phenomenon. However, there is a tendency in the evolution of LLMs in the last couple of years, which seems difficult to combine with the belief that they are only highly specialized prediction engines. This tendency is an evolution towards more polyvalence in those models. This polyvalence has several components:

- *Models are not trained for a single task but for several tasks*. As hinted in our introduction, LLMs are multi-purpose models. They can be used for tasks that were before the object of different Natural Language Processing (NLP) models, such as translation, summarization, text generation, Q&A, and deciding whether two sentences entail each other... LLMs have become the 'Swiss Army Knife' of NLP. LLMs are pre-trained at one task before receiving dedicated training for each additional task, and that initial task is to predict the next token in a linguistic sequence. Thus, not all tasks are

when this transformation happens and how we can acknowledge that it happened is at the core of our issues in this article.

equal: if LLMs are such performant polyvalent models, it is because they perform very well at the task of predicting the next token in a sequence. The reader can easily see that many tasks can be rephrased as predicting the rest of a particular sequence: an answer is what should be predicted after a question; a translated sentence is what should come after the instruction "Translate this sentence from Thai to German"; a generated text with a certain style is what should come after "Write a text on LLM in the style of Edgar Allan Poe." Prediction has thus proven to be a master task for training those language models, which already calls into question the respective independence of various linguistic tasks. The ability to train a pre-trained model to a new task is called "transfer learning".[4] The very name suggests that the model transfers some of the knowledge acquired in one task to another.

- *Models trained for a given task can then be used as a basis for a model intended for an intuitively different task.* This phenomenon is particularly striking in the emergence of multi-modal models. Language models previously differed from those for computer vision, each having its own dedicated architecture. However, recent years have seen the emergence of models able to interact with several modalities, such generating an image from a textual description. Those models have taken LLMs as their basis, which illustrates their ability to generalize beyond strictly linguistic tasks. This generalization raises difficult questions on the underlying ontology. If we agree, for the sake of argument, that those models encode some

4 Transfer learning is by no means restricted to LLMs, or even NLP, and has been a major practice of Deep Learning models in general; one can also find examples of transfer learning in computer vision.

theoretical knowledge, then it was easy before the advent of multi-modal models to jump to the conclusion that the object of that knowledge was language. The newly found multi-modal scope of LLMs opens the possibility that the actual object of that knowledge could be far more general. Some researchers in robotics implementing generative AI are even talking of "large behavior models" for robots trained on vast video databases demonstrating daily human activities (Heikkilä, 2024).

- *After training completion, some models can be prompted to execute a task they have not been trained for, simply by giving them some instructions and a set of examples.* This ability, clearly exhibited for the first time by GPT-3, is called "few-shot learning." Few-shot learning is an instance of "meta-learning", i.e., the ability to use previously acquired abilities to learn new ones: learning is learning to learn. Few-shot learning introduces a form of learning in those models that is not correlated to a modification of their parameters through training but to the exposure to new prompts. As the creators of GPT-3 noticed (Brown et al., 2020), it is difficult to distinguish between learning a new task through few-shot learning and re-arranging previous learning into a new format. Even this opposition might not be so clear, and could be considered a continuum rather than a dichotomy. In any case, it shows that LLMs are not software systems performing a pre-defined list of tasks and nothing more. Instead, LLMs are open-ended systems whose ability to converse with humans can be harnessed to prompt them into new tasks.

All those evolutions stand in sharp contrast to the common belief that recent AI would only represent "weak AI," i.e., highly specialized systems incapable of anything like a versatile intelligence. On the contrary, the evolution of recent years has shown a very discernible trend towards

ever greater polyvalence. This polyvalence, so the argument goes, would not be compatible with the belief that all the knowledge those models contain is purely task-relative: multi-tasking should be the first evidence for the presence of task-agnostic knowledge hidden inside those models. Please note that this is not meant as proof that those models contain anything like theoretical knowledge: We are just discussing a candidate necessary condition to the status of theoretical knowledge, so that objective is not achievable. The argument for task-agnosticism should rather be understood as a possible answer to the objection that contemporary AI is just weak AI and is not even a plausible candidate for theoretical knowledge.

Our various remarks on polyvalence also show that the conceptual distinction between tasks and their logical independence from one another is far from obvious. What the evolution of LLMs towards greater polyvalence seems to show is that many of our expectations on the independence of some tasks can be defeated, and that the models can learn—through their training for one task—a knowledge more general, which can then be transferred to a surprising amount of seemingly unrelated tasks. Again, this is insufficient to prove that those models contain anything like theoretical knowledge of a phenomenon. Even pragmatic knowledge can be transferred from one task to another. However, once again, this recent evolution of language models towards polyvalence and task-agnosticism dispels the first counter-argument against the presence of such theoretical knowledge, which assumed that those models could not contain theoretical knowledge because they were task-specialized, and provides a first form of positive evidence in its favor.

### *Data-specificity and robustness*

Other recent results are less favorable for those willing to consider LLMs as containers of theoretical knowledge. Another common expectation for such knowledge is that it is not data-specific. Theoretical knowledge is not about a particular dataset, but the underlying phenomenon on which the dataset was collected. This assumption is a significant reason why theory is expected to take us beyond the realm of description and prediction of data. Consequently, theory should enable us to formulate propositions that are robust to small, irrelevant perturbations of data. So long as a data item remains valid for the relevant phenomenon, it does not matter whether its value differs from previous observations: the theoretical propositions, being about the phenomenon itself, will still hold. For instance, a theory that claims to describe the behavior of massive objects in a gravitational field shall not depend on the color of those objects. If a scientist performs measurements on red billiard balls, her predictions shall not be invalidated by future observations of black billiard balls.

The reader may notice that it is difficult to define what perturbations count as "small" or "irrelevant" without referring to the theory itself. In the case of billiard balls, a change of color is "irrelevant" because the theory says that the dynamic of a body in a gravitational field shall depend on its position and mass, not on its color. The same goes for the notion of "small" perturbations. There is no absolute definition of size here: a perturbation counts as small if the theory states that such a size of perturbation should not affect the prediction of interest.

If LLMs contain theoretical knowledge, then their predictions should be robust to irrelevant and small data perturbations. However, at present, by and large, they fail to meet this standard. This failure is actually a widespread problem in ML. The reader may have heard that

computer vision models are now matching or even surpassing human abilities in object recognition. However, recent research has shown that this statement only holds for a certain quality of pictures (Hendrycks & Dietterich, 2019). Common perturbations of data quality, such as seeing the object through snow, rain, or glass, or seeing a compressed picture of said object, radically affect the performances of such models. However, human perception of those objects is robust to such perturbations. This robustness is coherent with the theoretical belief that the identity of objects should not be affected by those perturbations in our conditions of observation. The predictions of computer vision models are thus currently not robust to irrelevant data perturbations.

Such results are consistent with the hypothesis that ML models do not learn anything like robust knowledge of a phenomenon, but instead find quick-and-dirty heuristics to maximize their task performance metrics. Let us take an example from Natural Language Inference (NLI), the subfield of NLP dealing with logical relations between sentences: Given two sentences, an NLI model should determine whether they entail each other, contradict each other, or are logically independent or "neutral." After the first successes of ML models in the field, it was demonstrated that the successful models had acquired no genuine knowledge of logical inference in natural language but were simply using dirty heuristics somewhat reminiscent of a lazy student in an introductory logic class (Gururangan et al., 2018). Similar results were demonstrated in Natural Language Understanding (NLU) (Nie et al., 2019). This failure is a stark demonstration that statistical performance, even when using a large benchmark dataset, is a poor indicator of linguistic knowledge: any inference from performances on a task to knowledge of a phenomenon is thus deeply flawed.

LLMs suffer from the same issue. Slight modifications of their prompts, which a competent speaker would consider irrelevant to the semantics, may radically affect the output of those models (Pruthi et al., 2019).

This lack of robustness has many profound consequences for the development, maintenance, and usage of ML models. One of them, which we will return to in the final section, is the difficulty of maintaining ethical guidelines in the answers produced by the LLM, termed the "alignment problem" in the industry jargon. LLMs may refuse to answer a blatantly illegal or unethical question, such as how to fabricate a bomb or produce racist propaganda, but may nevertheless answer a slight variation on the same question. This phenomenon enables deliberate attempts to circumvent the defenses put in place to maintain ethical guidelines, by carefully choosing the formulation of natural language instructions (termed "prompts"). This particular manipulation of the model is called "jailbreaking." In the particular case of ChatGPT and other models, jailbreaking becomes more difficult over time, as counter-examples are signaled to the developers of the model, and it is retrained to correct those mistakes. However, this improvement method, known as "reinforcement learning by human feedback" (RLHF), never amounts to perfect protection, due to the aforementioned lack of robustness and our lack of understanding of its modalities: we can never be sure that the model has learned the appropriate relation of semantic equivalence between some instructions, and treats equally those things that deserve to be treated equally.

This practical limitation is exacerbated by two theoretical problems. The first is the lack of a uniform definition of robustness in ML in general and in LLMs in particular. The translation of the simple intuition that "similar inputs should trigger similar outputs" is thus not

obvious, as each task seems to demand the definition of an appropriate metric, and the various results of those definition efforts are not easy to compare. From a mathematical perspective, "robustness metrics" is thus a family term denoting a set of definitions rather than a unique concept. For instance, the main robustness metric is additive pixel-wise perturbation. This metric was obviously conceived for computer vision (see Adilova et al., 2022, p. 11; Ntalampiras et al., 2023).[5] It aims to formalize the intuition that a perturbation imperceptible to the human eye should not affect the model's decisions. There are two significant criticisms against this definition of robustness:

One: Incomplete capture of imperceptibility in computer vision. Small perturbations according to this metric might be easy to identify for the human eye, as some small perturbations affect the semantic relation between the parts of the image that are essential to human perception. Consequently, other metrics have been proposed to complement this standard metric (Tocchetti et al., 2022).

Two: Irrelevance of imperceptibility metrics for other domains, especially NLP. The notion of imperceptibility, no matter how it is formalized, may not make sense in other domains of ML, especially in NLP. The smallest perturbation of a text, say the modification of a letter, is perceptible by the human eye. It may be an option to replace "imperceptible" with "meaningless": human understandability of text is robust against some minor variations that are considered meaningless. For instance, understanding a written sentence is usually robust against

5 See the recent report by BSI, the German authority for information security and references therein (Adilova et al., 2022). A recent report by its counterpart at EU level, ENISA, explicitly regrets the lack of research on robustness metrics (Ntalampiras et al., 2023).

common typos and spelling errors (Pruthi et al., 2019), and understanding an oral sentence should be robust against common accents and some pronunciation mistakes. Other notions, such as robustness under synonym substitution (Colombo et al., 2022), may not exactly capture "meaninglessness" but may be considered minor under some semantic equivalence relation. Even larger text parts may be used to define some robustness under substitution. "I want to order the last plant you showed me" and "I want to order the blue mimosa" may be equivalent in the context of an audio reconnaissance system, which should place an order for flowers following the user's oral instructions and is thus primarily interested in the reference of the expressions, whatever their grammatical nature may be. Failing to order the same item should be considered faulty behavior. However, defining all the relevant metrics for NLP robustness is a formidable theoretical problem, as it assumes the capture of essential linguistic relations. NLP is known to be a subfield of AI in which this issue of metric definition is common (Dunietz, 2020), as NLP seeks to capture very complex intuitive features of language, such as the relevance of an answer to a question, the style of a text, the tone of an assertion, etc. Consequently, if the performance metric is not well defined, then the correct behavior of the system is not well defined. In the absence of such a definition, it may not be obvious to monitor whether the system is behaving as intended or has been maliciously or inadvertently driven out of desired behavior.

The second theoretical problem is the possible tradeoff relation between robustness and other metrics. No absolute mathematical results show the existence of such a relation. However, there is a collection of partial results showing that the optimization of robustness, at least according to some definition of robustness metrics, is incompatible with the optimization of other performance metrics, including fairness

(Gittens et al., 2022; Liu & Vicente, 2022; Ma et al., 2022; Maity et al., 2020; Xu et al., 2021).[6] If those results were to be confirmed, the quest for robustness may come at a significant cost regarding other important properties, including some of political consequence. However, one should be wary of reducing such a tradeoff to one between efficiency or sound scientific knowledge on the one hand versus ethical concerns on the other hand. As we will see in our final section, robustness also has an ethical–political meaning: implementing a LLM that respects some ethical guidelines is very difficult without robustness. This tradeoff is bound to become a crucial topic for the future of the industry, as the new European Union Artificial Intelligence Act, in its Article 15, requires developers to strive for robust, safe, and accurate models, apparently without any awareness of the possible tradeoff between those properties, and that of fairness. If those tradeoffs are confirmed, this would mean that the current EU legislation is asking for the impossible from developers, and the enforcement of said legislation is bound to become a major issue.

To conclude, it is impossible to directly answer the question: "Do LLMs contain theoretical knowledge?" because it must be analyzed via two sub-questions. There is a definitive trend towards task-agnosticism, but no definite trend towards robustness. The composition of those two answers, however, yields a negative answer to the overall question. In their current state, LLMs are unlikely to contain anything like theoretical knowledge, as robustness is a minimal condition for such knowledge to be present.

However, this is no final answer on the capacities of ML and LLMs to develop such knowledge in the future. After all, ML was not supposed

6 Maity et al. present a criticism of the idea of tradeoff based on a discussion of definitions.

to produce polyvalent models at the beginning of the new summer of AI; nevertheless, here we are—with extremely polyvalent LLMs. ML is not a static object but a constantly evolving methodology, and any definitive statement on its abilities must be highly cautious. Any optimistic statement should not make light of the structural difficulties we have identified in this article. However, we would also like to warn our readers against grand skeptical statements on the limitations of AI. As some commentators have already noticed (Carpenter, 2024),[7] grand arguments against the capacities of AI models tend to suffer from "shifting goalposts" syndrome: as AI models develop new capacities, the grand statements on the limitations of AI just move to the next target. As the ironic AI saying goes: Intelligence is whatever machines cannot yet do. In the onslaught of commentary that has followed the ChatGPT phenomenon, it has become common to claim that the chatbot simply produces empty "word salad" and will always be utterly incapable of intelligent linguistic behavior. This reaction is problematic in at least two ways. The first is that it misses a critical *a fortiori* argument. If ChatGPT is just a blind word-salad generator, how can it emulate so much of our linguistic practice in a realistic fashion? And what does it tell us about the actual intelligence of our daily linguistic practice? If we have seen that one should be very cautious about going from emulation of performances to actual imitation, it is not impossible that a good part of our linguistic practice is not extremely creative and that we could also be described as word-salad generators that combine bits of our linguistic memory.

7 This recent post by Bob Carpenter is also an accessible version of that.

## *Ethical–political consequences*

Let us examine some under-discussed ethical–political consequences of LLMs' lack of robustness on bias. To cut a long ethical–political story short, the term "bias" describes a supposed lack of political neutrality in LLMs' answers. As illustrated by several studies using political opinion tests designed for humans (Rutinowski et al., 2024), the LLMs' answers can sometimes show signs of a definite political orientation. ChatGPT, for instance, would clearly belong to the liberal-libertarian quadrant familiar to some political opinion tests such as the political compass.

After being a subject of academic debates for several years, this topic has reached mainstream media status after the ChatGPT phenomenon. Some American Conservatives, for instance, have announced their intention to develop a conservative LLM to counter ChatGPT's supposed "liberal bias" (Bye, 2023). It should be noted that the very notion of political neutrality is not unproblematic. When asked a question typical of political opinion testing, such as "Do you approve the following statement: 'My country, right or wrong'?" what would a neutral answer look like? An approbation of that sentence is typical of nationalist sentiments, while a disapprobation will be perceived as a critique of those sentiments: none of the possible answers is neutral here. It thus seems that the expression of any ethical or political view by the LLM is bound to trigger accusations of bias. The only possibility to avoid such accusations would be for the LLM to refuse to express any ethical or political view whatsoever by claiming incompetence in respect of such questions. This is not the path chosen by the major commercial LLMs today. In any case, such a strategic option would not be easy to implement, as it would be far from easy to delimit the set of questions that the LLM should refuse to answer.

A significant point has been widely missed in the onslaught of reactions to LLM bias. Bias would be presumed to consistently influence the views expressed by the LLM. As such, the model would not be expected to answer sensitive questions in terms reflecting diverse political viewpoints, but instead to consistently produce answers that reflect a given political trend. A conservative is supposed to express consistent conservative views, just as a liberal is supposed to be consistently liberal. After all, this consistency is necessary for the politically biased individual to do what they are accused of doing, that is, consistently interpreting facts in ways that support their existing political views.

There are two problems with this understanding of political opinions, which are well-known in political testing. The first is that the consistency of political views is far from being an obvious matter. Many individuals will express unusual combinations of political views. Short of making the massive hypothesis that only familiar combinations of views are consistent, it is not obvious to declare that those exceptional combinations are inconsistent. The second is the identification of the core views that are necessary to be part of a given political world-view and/or how many of those opinions an individual needs to uphold in order to be termed an adherent of that worldview. Finally, the dimensions along which political opinions are analyzed and how we should conceive them are major topics of debate whose answers influence the issues we just mentioned.

No matter the answer to those difficult methodological questions, it is safe to assume that some combinations of views are considered inconsistent. For instance, a progressive liberal is not expected to support the idea of a hierarchy of races. It seemed to have escaped the attention of public debate that ChatGPT and other LLMs have proven themselves guilty of precisely that type of inconsistency. Especially in the first few

months of its deployment, before reinforcement learning managed to diminish the number of those mistakes, ChatGPT regularly went from expressing typical progressive views to those that would be more typical of a KKK member. For instance, it could answer a request to classify the most intelligent populations by race and gender (spoiler alert: the White man wins) (Borji, 2023, Fig. 15). If we were to meet an individual who expressed such a rare combination of beliefs, we would not call them biased; instead, we would call them inconsistent or utterly insane.

The qualification of "biased" for the expression of ethical–political beliefs by LLMs is thus deeply misleading. It is an anthropomorphic qualification that leads us to expect a typically human ideological consistency, which is precisely absent here. In other words, the qualification of "bias" leads to a faulty theory of mind, in which we attribute to the LLM a consistent set of beliefs that is manifestly absent.

This is a general drawback of applying to LLMs tests conceived for humans (Heaven, 2023b). Such tests, especially concerning human reasoning abilities, have been massively applied to LLMs in the last couple of years, and their performances in legal or medical exams have been the topic of both research papers and newspaper headlines. However, many researchers have noticed that inference from performances to abilities is particularly faulty here. The first problem is that it is difficult to distinguish reasoning from simple memorization from rote: the LLM may have seen the test in its training data, but with those models' opacity it is difficult to confirm or infirm that possibility. The same problem is true in political contexts: When an LLM answers a question with a typical conservative assertion, is it expressing a "conservative worldview," whatever that may mean, or is it just spitting out an answer to a similar question it has seen in training data? The reader may notice that the political inconsistency of LLMs could be easily explained in the second

case. Namely, the LLM is merely spitting out whatever answer is the most common in its memory after a similar question, and there is no warrant that this process will always deliver an opinion from the same political side.

The second problem is that, even if we accept that the model is reasoning and not just reciting from memory, we cannot assume that this reasoning is similar to that of a human simply because the model succeeded at a human test. Here again, lack of robustness in performance warns us of the problem, as abysmal failure quickly follows brilliant performances: the same model that passed a university-level test can then fail a test of basic reasoning abilities conceived for small children. Performances at or above a human level should not lead to a projection of human psychology onto the LLM. The research on AI is here facing problems similar to ethology, where the tests for animal intelligence are designed explicitly to avoid anthropomorphic projections. Some are even calling for the use of other tests to avoid such confusion. All those conclusions should also be applied to the presence of "political bias" in LLMs.

The absence of bias in that anthropomorphic sense is not proof of neutrality. As said above, ideological neutrality is impossible as soon as the model declares itself competent to answer questions whose answers leave no room for ideological neutrality. Even if we neglect this particular issue, the cumulation of inconsistent biases is, of course, not equivalent to neutrality. However, the more important lesson here is that if the developers of LLMs try to implement a particular ideology in their LLM, they are currently failing. This failure is due to a lack of robustness of the LLM with regard to ideological consistency.

It may not be impossible to maintain a particular notion of bias towards LLM, in the sense that answers of a certain stripe are more

frequent in an LLM than others: ChatGPT sounds more often like a progressive Californian and less often like a KKK member. However, this particular phenomenon should probably be designated by a technical term to avoid the hasty assimilation into human behavior. Keeping this distinction in mind is particularly important because a paradox in the comparison between human and machine behavior is not to be missed here: We expect machines to be more uniform than humans, not less. It is this expectation that is defeated in the world of non-robust LLMs.

Again, defining an adequate notion of robustness is a formidable challenge. According to the relevant metric on the input space, two inputs are close if they both express questions prompting the expression of political views on a particular topic. According to the relevant metric on the output space, two outputs are close if they express "similar political views" that could be seen as "belonging to the same political side." Two answers on the opposite side of the political spectrum should be far from each other according to that metric. I am using the singular here, but it may be more likely that we would need several metrics, each capturing a dimension along which are expressed conflictual views, like the dimensions of the famous quadrant. The reader may have already guessed that such a vision is pure fantasy in the current state of the art. We do not have, and will not have shortly, a set of metrics enforcing political consistency in the outputs of LLMs, and it does not seem that the other constraints on their training bring political consistency as an extra benefit.

## *Conclusion*

The recent evolution of ML shows a definitive trend towards task-agnosticism but not towards robustness. The topic is a subject of intense research, and the future is uncertain. It is extremely difficult to draw definitive conclusions on AI abilities in such a fast-moving landscape. In particular, it is currently impossible to say whether ML models in general, and LLMs in particular, may one day be able to derive theoretical knowledge in full autonomy if we admit our hypotheses that theoretical knowledge should be task-agnostic and robust.

The question of robustness also has profound consequences in the ethical–political realm. LLMs fail to follow ethical guidelines because of their lack of robustness. However, the quest for robustness may also come at an ethical cost, since there is strong evidence that there may be a tradeoff between robustness and other properties, including fairness. It is not only the future of robust LLMs that is uncertain, but also the consequences of that future. This is bound to become a significant concern in the enforcement of the EU AI Act, as it is currently asking developers to optimize both fairness and robustness without any consideration for possible tradeoffs.

## Reviews

The paper received two demanding reviews. Unfortunately, Maël Pégny was not able to include all the points raised in the reviews—although they are important—in his paper at this stage and in an appropriate form. We have therefore decided to publish the reviews together with Pégny's responses.

*Commentary by Bob Williamson*

> *Williamson: The paper leans heavily on "knowledge" or "theoretical knowledge" and "contains knowledge". I think it would be helpful to say what you mean by this. I do not mean to ask for a long attempt at a formal definition (this is hopeless, for sure). But perhaps a few examples of what you mean when you say some entity knows something. You do talk about task-agnosticism and robustness, but you use the word "knowledge" in the title. Perhaps the readers would be less confused if you changed the title to be something like "Do LLMs know anything in a robust and task-agnostic manner?".*
>
> *One more linguistic quibble that I think is actually more than a quibble: What is the difference between "contain knowledge" and "know". A library surely 'contains knowledge'. But few (I think) would claim that a library 'knows'. I think the comparison with libraries is actually very relevant, given the nice way of conceiving of LLMs as being a 'library that talks' [1]. (I lean on this further, below).*

[1] Barandiaran, X. E., & Almendros, L. S. (2024). Transforming agency. On the mode of existence of Large Language Models. arXiv preprint https://doi.org/10.48550/arXiv.2407.10735 ("library that talks")

*Pégny*: Let me clarify what I mean by the exceedingly vague metaphor "contains theoretical knowledge." It is, after all, a metaphor to state that a model contains knowledge, just as it would be to say the same of a library or a book. If LLMs contain knowledge, the following process may become a part of future theoretical investigations. Instead of directly theorizing about a class of, say, linguistic phenomena, a scientist may train a large ML model to predict the occurrences of those phenomena, then switch to studying the representations of those phenomena within the model in order to extract some theoretical propositions from them. As the AI community commonly talks about "explainability" or interpretability to describe the efforts to understand what goes on in complex ML models, it could be said that explainable AI, in the scenario above, would become a new path to theoretical knowledge. In such a scenario, it becomes natural and relevant to say that those models would contain knowledge, to denote the fact that they encode that knowledge, albeit in a very opaque and unreadable form at first.

> *Williamson: Section 2 concludes that it is not directly possible to answer the question "Do LLMs contain theoretical knowledge?". Another response to that conclusion is to ask a different (and better) question. What might such questions be? One might start with "What are some of the metaphors or analogies one can use to describe what LLMs do?" (My favorite one is that described in [2]: LLMs (or more precisely, ChatGPT) is a bullshitter.*

*Pégny*: I am not sure whether this is entirely serious, and in any case, I refuse this characterization. The problems raised by those artefacts deserve

[2] Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology, 26*(2), 38. https://doi.org/10.1007/s10676-024-09775-5

in my humble view a more nuanced approach than the binary option "LLMs are talking Gods who will take our job" vs "LLMs are bullshit and word salad, why is anybody using this?".

> *Williamson: Regarding section 3, it strikes me that the issue is less the lack of robustness (which, vague as the notion is, is clearly an issue) than the problematic notion of "bias". It is recognized in the next that there is no "neutrality" possible (no "view from nowhere" in effect). But this is a problem with the notion of bias, not with the LLM. I think it would be good to stress this point some more, else you are encouraging folks on the unachievable task of "removing bias".... And I am not persuaded by the argument that bias is somehow a consequence of "inconsistent beliefs".*

*Pégny*: To be clear, I am not saying that bias is always the consequence of inconsistent beliefs. I am saying that the anthropomorphic term of "bias" leads many to falsely believe that LLMs give something like the expression of a consistent set of beliefs on ethical–political issues, whereas we have massive evidence to the contrary: LLMs produce the expression of inconsistent beliefs. It is also hard for me to tell if "removing bias" is a complete pipedream, despite the obvious qualms that I also have with the concept. There are cases that obviously deserve attention and correction: it is the overall fantasy of wanting "neutral" answers to ethical–political questions that I see as a pipedream.

> *Williamson: To be more critical of the argument presented: it is claimed that a major issue with LLMs is that they are not robust. The author acknowledges the many problems in even defining what robust means. Indeed, there can not possibly be some all-purpose task-agnostic definition of robustness (which is kind of funny since you demand both task-agnosticism and robustness!).*

*Pégny*: There are two levels to the definition of robustness. The first is the informal, family notion of "stable under small perturbations." This notion is consensual, and is a normal scientific notion which predates ML. The other is the particular formalizations of that generic notion for particular tasks, which demands, among other things, defining a metric to capture the magnitude of perturbation. This metric is relative to the task, and it is difficult to imagine it otherwise: a perturbation of an image is not the same as a perturbation of a text.

> *Williamson: But, then, I ask why bother with robustness. It only makes sense if you think that the system is doing what folks claim it is. If you talk about its knowledge being robust, you are presuming that it has knowledge.*

*Pégny*: No, robustness in the technical sense of robustness metrics does not presuppose knowledge as I use it in this article. They are defined on the invariance of performance metrics.

> *Williamson: Perhaps I can make my point this way: if you take for now the characterization of an LLM as a "library that talks" (see above), then you are asking whether your library is robust. Now for sure there would be people that would argue about whether libraries contain the "right" books (and go and burn those they do not like). But what on earth can be meant by a library being "robust"? Perhaps reflecting on this will give some new angle to the questions raised in the paper.*

*Pégny*: I am not certain that I would describe LLMs as libraries that talk, and as a consequence, I am unconvinced of the relevance of a discussion of library robustness. The whole NLP community is very much aware of the difficulties raised by the definition of performance metrics. Robustness metrics are even worse: The first definitions came from the field of computer vision, and are obviously unadapted to NLP. What is "a

small perturbation of text irrelevant to its semantics"? This brings us back to fundamental questions on linguistic concepts—the type of questions that ML was, in part, methodologically designed to avoid.

*Commentary by Nico Formánek*

*Formánek: The article argues that task-agnosticism would be evidence for and lack of robustness evidence against theoretical knowledge in LLMs. Theoretical knowledge is defined to be task-agnostic and robust. The argument that current LLMs do not contain theoretical knowledge is then a straightforward consequence of that definition and the observation that they are not robust. I do think the question if LLMs contain theoretical knowledge is interesting. Similar questions have been explored in the literature (see suggested literature). They are conspicuously absent in the article, although very related.*

*It is uncontroversial that many ML models suffer from "strange errors" and therefore should not considered to be robust on any measure. The negative conclusion of the paper thus rests on its characterization of theoretical knowledge. I think more detail is needed here. For example an important distinction is made between theoretical and pragmatic knowledge. But these terms are not discussed in much detail. It is suggested that theoretical knowledge, presumably as opposed to pragmatic knowledge, is general and therefore task agnostic. But the thermodynamics example fails to convince me of that. Arguably the Carnot cycle is task agnostic in the sense that it holds for every heat engine. So why then is it considered task specific? Why is it considered pragmatic rather than theoretical? Furthermore the connection between theories, theoretical knowledge and pragmatic knowledge should be elucidated. I think the author implicitly connects pragmatic knowledge*

*with means-ends knowledge (knowing how to achieve a certain goal, e.g. next-token-prediction) and theoretical knowledge with explanatory knowledge. If this is the case then a discussion of the relation between explanation and prediction seems inevitable. At any rate the paper needs to be more explicit and argumentative about why theoretical knowledge should be considered robust.*

*Miscellanea: Is it really surprising that LLMs can approximate what you call polyvalence? Isn't this just a consequence of the universal approximation theorem?*

*Pégny*: 1. On the universal approximation theorem implying the possibility of polyvalent models: The reviewer has not made his argument on the universal approximation theorem explicit, so I am forced to do some reconstruction here. Here are my objections to this:

- The mathematical function approximated by a neural network is, in my view, a model of a task, not the task itself (even though terminology confuses model and the target of that model here, as is very often the case in science). Approximating a mathematical function is no guarantee that you approximate the intuitive task, for instance "predicting the next word in a sentence."
- The universal approximation theorem is an existence theorem: it does not prove that we will manage to converge to the desired approximation. The actual realization of that approximation is thus an engineering feat, not a simple consequence of a theorem.
- The universal approximation theorem assumes no limitation on the size of a neural network. Actually, some research shows that limitations on width block one version of the theorem on the density in Lebesgue integrable function space [Lu, Z., Pu, H., Wang, F., Hu, Z., Wang, L. (2017). The expressive power of neural networks: A view from the width. NIPS 2017: Advances in Neural Information Processing Systems, 30, 6231–6239. arXiv preprint.

https://doi.org/10.48550/arXiv.1709.02540]. There is no guarantee that the neural network necessary to reach some approximation may not need to be larger than the number of quarks in the Universe, which would raise obvious implementation issues. As such, the theorem is again no guarantee that we will implement such an approximation in the real world.

- Finally, and this is the main argument: The versions of the theorem I know discuss the approximation of one function by neural networks. They do not speak of approximating any number of functions by a fixed neural network. In other words, it is not because there exists a neural network to approximate every function that there exists a neural network that approximates all functions or many of them. In that sense, even from a purely theoretical perspective, the theorem is no warrant of polyvalence; it is actually silent on the issue.

2. Why should theories be robust? I would say that the aim of a model is to produce predictions about a phenomenon, not about data. Phenomena are defined in a language that already abstracts away from the many details of concrete instances that are measured by data: when I talk about a massive body moving in a gravitational field, this level of abstraction already supposes that variations in factors such as the color and taste of massive objects are irrelevant to the dynamics of falling massive bodies. There is already an implicit generalization in the mere definition of a phenomenon that will make some features of the data irrelevant. To make my point with another example, being able to identify an object from a picture means you can recognize it from any picture of decent quality, not limited to pictures taken from particular angles: in a sense, that's part of our intuition of what an object is, that it is invariant by rotation. A model producing predictions that are not robust to small changes in data fails to provide predictions about a phenomenon: it is "just" predictions about data.

3. On Carnot's thermodynamics being theoretical because of the generality of heat engine. I do not wish to put words in the reviewer's mouth, but I feel he may be projecting our contemporary interpretation of his results onto Carnot's own view of what a "heat machine" (machine à feu) was. In Carnot's original text, it is very clear that a heat engine is a concrete artifact performing concrete tasks. If I translate roughly the original XVIIIth century French: "The heat engine already exploits our mines, moves our ships, digs up our havens and rivers, casts iron, molds wood, crushes grains, weaves our stuff, carries the heaviest loads and so on ... It seems bound to serve as a universal engine, and be preferred to animal strength, waterfalls, and wind currents."

As for the fact that Carnot reasons on a universal class of machines, I can testify as a software engineer myself that engineers constantly reason on classes of possible artifacts sharing common specification without ever losing sight of concrete objectives. This generalization may be a path to theoretical knowledge sometimes, but its aim is radically different.

*Editorial commentary by Dave Morris*

> *On the use and robustness of open review or pre-print repositories: Such sources are increasingly prevalent, particularly in rapidly evolving fields such as AI, but can raise concerns regarding rigor, credibility, or academic oversight. The cited conference submission by Maity et al. (2020) was rejected on initial review; The article by Borji (2023) has been uploaded to numerous pre-print repositories, assigned two different DOIs (on arXiv and ResearchGate), and has a high citation count (collated on https://scite.ai/)—all without formal peer review or stated affiliation. We briefly address the emerging challenges and (potentially beneficial) changes to peer review in the editorial.*

## *References*

Adilova, L., Böttinger, K., Danos, V., Jacob, S., Langer, F., Markert, T., Poretschkin, M., Rosenzweig, J., Schulze, J-P., & Sperl, P. (2022). *Security of AI-systems: Fundamentals*. Bundesamt für Sicherheit in der Informationstechnik – BSI. https://publica.fraunhofer.de/handle/publica/443025

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). *On the opportunities and risks of foundation models*. arXiv preprint. https://doi.org/10.48550/arXiv.2108.07258

Borji, A. (2023). *A categorical archive of ChatGPT failures*. arXiv preprint. https://doi.org/10.48550/arXiv.2302.03494

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., ... Dario Amodei. (2020). Language models are few-shot learners. *NIPS 2020: Advances in Neural Information Processing Systems, 33*, Article No. 159, 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Bye, C. (2023, February 9). ChatGPT may be the next big thing, but it's a biased woke robot. *Daily Telegraph* (Australia). https://www.dailytelegraph.com.au/news/opinion/chatgpts-hidden-leftwing-bias-built-into-ai-robot-makes-it-infuriating/news-story/353c763506e24e411abf80d58eca2482

Carpenter, B. (2024, April 13). *Intelligence is whatever machines cannot (yet) do* [Online forum post]. Statmodeling. https://statmodeling.stat.columbia.edu/2024/04/13/intelligence-is-whatever-machines-cannot-yet-do/

Colombo, P. J. A., Clavel, C., & Piantanida, P. (2022). InfoLM: A new metric to evaluate summarization and data2text generation. *Proceedings of the AAAI Conference on Artificial Intelligence, 36*(10), 10554–10562. https://doi.org/10.1609/aaai.v36i10.21299

Dunietz, J. (2020, July 31). *The field of natural language processing is chasing the wrong goal*. MIT Technology Review. https://www.technologyreview.com/2020/07/31/1005876/natural-language-processing-evaluation-ai-opinion/

Gittens, A., Yener, B., & Yung, M. (2022). An adversarial perspective on accuracy, robustness, fairness, and privacy: Multilateral-tradeoffs in trustworthy ML. *IEEE Access, 10*, 120850–120865. https://doi.org/10.1109/ACCESS.2022.3218715

Gururangan, S., Swayamdipta, S., Levy, O., Schwartz, R., Bowman, S. R., & Smith, N. A. (2018). *Annotation artifacts in natural language inference data*. arXiv preprint. https://doi.org/10.48550/arXiv.1803.02324

Heaven, W. D. (2023a, March 3). *The inside story of how ChatGPT was built from the people who made it*. MIT Technology Review. https://www.technologyreview.com/2023/03/03/1069311/inside-story-oral-history-how-chatgpt-built-openai/

Heaven, W. D. (2023b, August 30). *AI hype is built on high test scores. Those tests are flawed*. MIT Technology Review. https://www.technologyreview.com/2023/08/30/1078670/large-language-models-arent-people-lets-stop-testing-them-like-they-were/

Heikkilä, M. (2024, April 11). *Is robotics about to have its own ChatGPT moment?* MIT Technology Review. https://www.technologyreview.com/2024/04/11/1090718/household-robots-ai-data-robotics/

Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. *Proceedings of the 7th International Conference on Learning Representations*. https://openreview.net/pdf?id=HJz6tiCqYm

Liu, S., & Vicente, L. N. (2022). Accuracy and fairness trade-offs in machine learning: A stochastic multi-objective approach. *Computational Management Science, 19*, 513–537. https://doi.org/10.1007/s10287-022-00425-z

Ma, X., Wang, Z., & Liu, W. (2022). On the tradeoff between robustness and fairness. Advances in Neural Information Processing Systems. *NIPS 2022: Advances in Neural Information Processing Systems, 35*, 26230–26241. https://proceedings.neurips.cc/paper_files/paper/2022/hash/a80ebbb4ec9e9b39789318a0a61e2e43-Abstract-Conference.html

Maity, S., Mukherjee, D., Yurochkin, M., & Sun, Y. (2020). *There is no tradeoff: Enforcing fairness can improve accuracy*. Open Review (ICLR 2021 conference submission). https://openreview.net/forum?id=wXoHN-Zoel

Nie, Y., Williams, A., Dinan, E., Bansal, M., Weston, J., & Kiela, D. (2019). *Adversarial NLI: A new benchmark for natural language understanding*. arXiv preprint. https://doi.org/10.48550/arXiv.1910.14599

Ntalampiras, S., Misuraca, G., & Rossel, P. (2023). *Artificial intelligence and cybersecurity research. ENISA research and innovation brief*. European Union Agency for Cybersecurity – ENISA. https://www.cybersecitalia.it/wp-content/uploads/2023/06/Artificial-Intelligence-and-Cybersecurity-Research.pdf

Pruthi, D., Dhingra, B., & Lipton, Z. C. (2019). *Combating adversarial misspellings with robust word recognition*. arXiv preprint. https://doi.org/10.48550/arXiv.1905.11268

Rutinowski, J., Franke, S., Endendyk, J., Dormuth, I., Roidl, M., & Pauly, M. (2024). The self-perception and political biases of ChatGPT. *Human Behavior and Emerging Technologies, 1*, 1–9. https://doi.org/10.1155/2024/7115633

Tocchetti, A., Corti, L., Balayn, A., Yurrita, M., Lippmann, P., Brambilla, M., & Yang, J. (2022). *A.I. robustness: A human-centered perspective on technological challenges and opportunities*. arXiv preprint. https://doi.org/10.48550/arXiv.2210.08906

Xu, H., Liu, X., Li, Y., Jain, A., & Tang, J. (2021). To be robust or to be fair: Towards fairness in adversarial training. *Proceedings of the 38th International Conference on Machine Learning Research, 139*, 11492–11501. https://proceedings.mlr.press/v139/xu21b/xu21b.pdf

## *Literature suggestions*

I warmly thank the reviewer who suggested this list of references for readers wishing to strengthen their understanding of the technicalities, which I could only gesture at in this introductory presentation.

Training and finetuning:

https://huggingface.co/blog/how-to-train

Reinforcement learning from human feedback:

https://huggingface.co/blog/rlhf

Evaluating LLMs:

https://thegradient.pub/understanding-evaluation-metrics-for-language-models/

Robustness:

Freiesleben, T., & Grote, T. (2023). Beyond generalization: A theory of robustness in machine learning. *Synthese, 202*, 109. https://doi.org/10.1007/s11229-023-04334-9

Representations in AI:

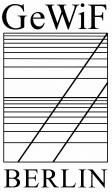
Arkoudas, K. (2023). *GPT-4 can't reason*. arXiv preprint. https://doi.org/10.48550/arXiv.2308.03762

Freiesleben, T. (2024). *Artificial neural nets and the representation of human concepts*. arXiv preprint. https://doi.org/10.48550/arXiv.2312.05337

Lyre, H. (2024). *Understanding AI: semantic grounding in large language models*. arXiv preprint. https://doi.org/10.48550/arXiv.2402.10992

Rowbottom, D.P., Peden, W., & Curtis-Trudel, A. (2024). Does the no miracles argument apply to AI? *Synthese, 203*, 173. https://doi.org/10.1007/s11229-024-04524-z

## Jahrbücher Wissenschaftsforschung




Alle Jahrbucher sind open access verfügbar, und zwar über den Bibliotheksserver der Humboldt-Universität zu Berlin: https://edoc.hu-berlin.de/handle/18452/90

Wissenschaftsforschung: Jahrbuch 1994/95. Hrsg. v. Hubert Laitko, Heinrich Parthey u. Jutta Petersdorf. Marburg: BdWi-Verlag 1996.

Wissenschaftsforschung: Jahrbuch 1996/97. Hrsg. v. Siegfried Greif, Hubert Laitko u. Heinrich Parthey. Marburg: BdWi-Verlag 1998.

Wissenschaft und Digitale Bibliothek: Wissenschaftsforschung Jahrbuch 1998. Hrsg. v. Klaus Fuchs-Kittowski, Hubert Laitko, Heinrich Parthey u. Walther Umstätter. Berlin: Gesellschaft für Wissenschaftsforschung 2000.

Wissenschaft und Innovation: Wissenschaftsforschung Jahrbuch 1999. Hrsg. v. Siegfried Greif u. Manfred Wölfling. Berlin: Gesellschaft für Wissenschaftsforschung 2003.

Organisationsinformatik und Digitale Bibliothek in der Wissenschaft: Wissenschaftsforschung Jahrbuch 2000. Hrsg. v. Klaus Fuchs-Kittowski, Heinrich Parthey, Walther Umstätter u. Roland Wagner-Döbler. Berlin: Gesellschaft für Wissenschaftsforschung 2001.

Wissenschaft und Innovation: Wissenschaftsforschung Jahrbuch 2001. Hrsg. v. Heinrich Parthey u. Günter Spur. Berlin: Gesellschaft für Wissenschaftsforschung 2002.

Wissenschaftliche Zeitschrift und Digitale Bibliothek: Wissenschaftsforschung Jahrbuch 2002. Hrsg. v. Heinrich Parthey u. Walther Umstätter. Berlin: Gesellschaft für Wissenschaftsforschung 2003.

Evaluation wissenschaftlicher Institutionen: Wissenschaftsforschung Jahrbuch 2003. Hrsg. v. Klaus Fischer u. Heinrich Parthey. Berlin: Gesellschaft für Wissenschaftsforschung 2004.

Wissensmanagement in der Wissenschaft: Wissenschaftsforschung Jahrbuch 2004. Hrsg. v. Klaus Fuchs-Kittowski, Walther Umstätter u. Roland Wagner-Döbler. Berlin: Gesellschaft für Wissenschaftsforschung 2008.

Gesellschaftliche Integrität der Forschung: Wissenschaftsforschung Jahrbuch 2005. Hrsg. v. Klaus Fischer u. Heinrich Parthey. Berlin: Gesellschaft für Wissenschaftsforschung 2006.

Wissenschaft und Technik in theoretischer Reflexion: Wissenschaftsforschung Jahrbuch 2006. Hrsg. v. Heinrich Parthey u. Günter Spur. Frankfurt am Main: Peter Lang Europäischer Verlag der Wissenschaften 2007.

Integrität wissenschaftlicher Publikationen in der Digitalen Bibliothek: Wissenschaftsforschung Jahrbuch 2007. Hrsg. v. Frank Havemann, Heinrich Parthey u. Walther Umstätter. Berlin: Gesellschaft für Wissenschaftsforschung 2007.

Selbstorganisation in Wissenschaft und Technik: Wissenschaftsforschung Jahrbuch 2008. Hrsg. v. Werner Ebeling u. Heinrich Parthey. Berlin: Wissenschaftlicher Verlag Berlin 2009.

Wissenschaft und Innovation: Wissenschaftsforschung Jahrbuch 2009. Hrsg. v. Heinrich Parthey, Günter Spur u. Rüdiger Wink. Berlin: Wissenschaftlicher Verlag Berlin 2010.

Interdisziplinarität und Institutionalisierung der Wissenschaft: Wissenschaftsforschung Jahrbuch 2010. Hrsg. v. Klaus Fischer, Hubert Laitko u. Heinrich Parthey. Berlin: Wissenschaftlicher Verlag Berlin 2011.

Kreativität in der Forschung: Wissenschaftsforschung Jahrbuch 2012. Hrsg. v. Thomas Heinze, Heinrich Parthey, Günter Spur u. Rüdiger Wink. Berlin: Wissenschaftlicher Verlag Berlin 2013.

Forschung und Publikation in der Wissenschaft: Wissenschaftsforschung Jahrbuch 2013. Hrsg. v. Heinrich Parthey u. Walther Umstätter. Berlin: Wissenschaftlicher Verlag Berlin 2014.

Wissenschaft und Innovation: Wissenschaftsforschung Jahrbuch 2014. Hrsg. v. Jörg Krüger, Heinrich Parthey u. Rüdiger Wink. Berlin: Wissenschaftlicher Verlag Berlin 2015.

Struktur und Funktion wissenschaftlicher Publikationen im World Wide Web: Wissenschaftsforschung Jahrbuch 2015. Hrsg.v. Klaus Fuchs-Kittowski, Heinrich Parthey u. Walther Umstätter. Berlin: Wissenschaftlicher Verlag Berlin 2015.

Forschendes Lernen: Wissenschaftsforschung Jahrbuch 2016. Hrsg. v. Hubert Laitko, Harald A. Mieg u. Heinrich Parthey. Berlin: Wissenschaftlicher Verlag Berlin 2017.

Ambivalenz der Wissenschaft: Wissenschaftsforschung Jahrbuch 2017. Hrsg. v. Klaus Fischer u. Heinrich Parthey. Berlin: Wissenschaftlicher Verlag Berlin 2019.

Wissenschaft und Innovation: Wissenschaftsforschung Jahrbuch 2018. Hrsg. v. Jörg Krüger u. Heinrich Parthey. Berlin: Wissenschaftlicher Verlag Berlin 2019.

Wissenschaftsverantwortung: Wissenschaftsforschung Jahrbuch 2019. Hrsg. v. Harald A. Mieg, Hans Lenk u. Heinrich Parthey. Berlin: Wissenschaftlicher Verlag Berlin 2020.

Wissenschaft als Beruf: Wissenschaftsforschung Jahrbuch 2020. Hrsg. v. Harald A. Mieg, Christiane Schnell u. Rainer E. Zimmermann. Berlin: Wissenschaftlicher Verlag Berlin 2021.

Kritisches Denken - Critical Thinking: Wissenschaftsforschung Jahrbuch 2021. Hrsg. v. Harald A. Mieg u. Frank Havemann. Berlin: Wissenschaftlicher Verlag Berlin 2022.

Transferability - Reflections on Planning and Knowledge Organization: Wissenschaftsforschung Jahrbuch 2022. Ed. by Harald A. Mieg & Andrea Scharnhorst. Berlin: Wissenschaftlicher Verlag Berlin 2023.

According to Aristotle, theory is the highest, rational form of knowledge. But do we need theories at all, given the new possibilities created by artificial intelligence to machine process vast amounts of data? The book is based on a two-day symposium jointly organized by the Robert K. Merton Center for Science Studies and the Gesellschaft für Wissenschaftsforschung Berlin. The symposium focused on three themes: 1) Theorizing and its new media: What kind of theorizing do science podcasts enable? 2) Epistemology and the metaphysics of theory, with some remarks on the philosophy of science, including the role of theory in university teaching; 3) AI and theory, with the final question of theory and the transparency of knowledge organization. The book begins with a reflection by Andrew Abbott on theory in the social sciences.

Berlin Universities Publishing
ISBN 978-3-98781-034-3 (print)
ISBN 978-3-98781-035-0 (online)