STUDIA METODOLOGICZNE

39

WYDAWNICTWO NAUKOWE UAM

STUDIA
METODOLOGICZNE

ADRES REDAKCJI

Wydział Filozoficzny UAM
ul. Szamarzewskiego 89C
60-568 Poznań
e-mail: studmed@amu.edu.pl; rzepinskit@wp.pl

Strona internetowa czasopisma:

http://www.studiametodologiczne.amu.edu.pl

Wydanie drukowane jest podstawową wersją czasopisma

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU

STUDIA METODOLOGICZNE

DISSERTATIONES METHODOLOGICAE

39

Issue on Culture(s) of Modelling in Science(s)

WYDAWNICTWO
NAUKOWE

POZNAŃ 2019

Publikacja dofinansowana ze środków grantu nr 0266/NPRH/H2b/83/2016
oraz przez Instytut Filozofii UAM i Wydział Nauk Społecznych UAM

Komputerowe opracowanie okładki: K. & S. Szurpit
Redaktor: Bożena Kapusta
Redaktor techniczny: Elżbieta Rygielska
Łamanie komputerowe: Reginaldo Cammarano

ISSN 0039-324X

WYDAWNICTWO NAUKOWE UNIWERSYTETU IM. ADAMA MICKIEWICZA W POZNANIU
61-701 POZNAŃ, UL. A. FREDRY 10
www.press.amu.edu.pl
Sekretariat: tel. 61 829 46 46, faks 61 829 46 47, e-mail: wydnauk@amu.edu.pl
Dział sprzedaży: tel. 61 829 46 40, e-mail: press@amu.edu.pl

Wydanie I. Ark. wyd. 18,50. Ark. druk. 18,50

DRUK I OPRAWA: VOLUMINA.PL DANIEL KRZANOWSKI, SZCZECIN, UL. KS. WITOLDA 7–9

Contents

Foreword: Culture(s) of Modelling in Science(s)

This volume of *Studia Metodologiczne* (*Dissertationes Methodologicae*) addresses the question of culture(s) of modelling in science(s), bringing together two issues significant for contemporary methodology of sciences, namely scientific modelling and scientific culture. For a long time these two issues used to be treated separately in philosophical discussions and with discernable priority given to scientific modelling. Thus, once natural, computer, cognitive and social sciences became broadly populated by models, the ontological nature, cognitive status and practical types of scientific models and modelling were taken as primary objects of numerous philosophical investigations. Recognizing that many scientific disciplines are populated by models of different nature, status and type does not preclude, however, that scientific modelling is still immersed in symbolic and material culture. This is where the idea of scientific culture comes into play.

Since the very term "scientific culture" has been systematically equipped with many different meanings, there are numerous conceptual tools at hand for philosophical reconstructions and analyses of various manifestations of symbolic and material culture in daily scientific research practice. Let us recall here only a few of them: 'material and theoretical cultures' (Peter Galison), 'thought styles' (Ludwik Fleck), 'epistemic cultures' (Karin Knorr-Cetina), 'styles of reasoning' (Ian Hacking), 'epistemological cultures' (Evelyn Fox Keller), 'experimental cultures' (Hans-Jorg Rheinberger), 'local scientific cultures' (Barry Barnes, David Bloor, John Henry), 'evaluation cultures' (Donald MacKenzie), 'scientific imagination' (Fiora Salis, Roman Frigg) or 'norms of science' (Robert Merton).

Regardless of the multiplicity of available conceptual tools that help theoretically grasp the symbolic and material culture in daily scientific research practice, the question how to discuss culture(s) of modelling in science(s) remains open. With this volume the editors aim to contribute to this discussion. Our idea is to equip the reader with a conceptual framework that may help him or her in a two-fold way: better orientate in heterogeneity of conceptualizations of cultural dimensions of scientific research in general and scientific modelling in particular, as well as increase awareness of interconnections between these various conceptualizations. For this purpose, we distinguish three subproblems within the leading problem of culture(s) of modelling in science(s):

- culture(s) of science – which relates to the question of a multitude of cultures with the spectrum of possibilities from monism (monoculturalism), via dualism (biculturalism) to pluralism (multiculturalism or polyculturalism);
- culture in science(s) – which relates to the question of a range of culture with the spectrum of possibilities from global culture, via regional culture to local culture;
- culture(s) of modelling – which relates to the question of a function of modelling with the spectrum of possibilities from the culture of representing, via the culture of intervening to the culture of exploring.

This volume consists of contributions by scholars with different disciplinary background who either investigate the culture(s) of modelling in science(s) or reflect on cultural dimension of their own modelling practice. The first article offers an analysis of the very term of 'model' by exploring different meanings attached to this terms in different domains (logic, mathematics, science, everyday life), as well as different uses that model may serve; the author further presents his account of the general theory of models (Bernhard Thalheim). The second paper deals with the view on scientific modelling by the physicist Sir Rudolf Peierls whose taxonomy of scientific models exhibits points of convergence with contemporary philosophical accounts of how scientific models function; the author argues that Peierls' view warrants the recent philosophical shift from a focus on model-based representation to non-representational (e.g., exploratory) uses and functions of models (Axel Gelfert). In the third article the authors present their account on the past and future of modelling in biology and invite philosophers of biology to provide normative research guidance for biologists; such a call comes amid unprecedented availability of ecological, evolutionary, and molecular data, of computational resources, and of mathematical and statistical tools (Steven Hecht Orzack, Brian McLoone).

The fourth paper focuses mainly on the proper uses and difficulties of formal theory (e.g., rational choice theory, game theory) in political science; according to the author, the roots of the formal approach can be traced to Thomas Hobbes and William Riker's second launch of 'Hobbesian advice' who put the field of formal theory on the map of political science; the author supports his historical analyses by both offering an example that explains the necessity of formal political science and discussing a trap for a barefoot empiricism (Piotr Świstak). In the fifth article the authors concentrate on the social sciences and present the variety of computational methodologies from both data-driven (such as 'black box') and rule-based (such as 'per analogy') approaches; what is more, they show how to build simple models and discuss both the greatest successes and the major limitations of modelling societies and populations (Andrzej Jarynowski, Michał B. Paradowski, Andrzej Buda). The sixth paper focuses on the nature of knowledge about the world that models and modelling give us; it puts forward the thesis that models are producers of beliefs about their targets and concludes that these beliefs should not be interpreted in terms of probabilities but rather as claims about prototypical characteristics of entities being under investigation (Łukasz Hardt). In the seventh article the author discusses how the developments in game theory and social choice theory transformed our understanding and modeling of social rationality in the social sciences due to the erosion of the concept of social optimum (Marek M. Kamiński).

The eighth contribution to the volume shows idealizations and limiting cases in models as playing an exploratory role in science; the authors distinguish four senses of explorations and illustrate their claims with three case studies from physics; finally they compare their account of idealization with Michael Weinsberg's three-fold taxonomy (Elay Shech, Axel Gelfert). The ninth paper calls for the need to introduce analysis of value judgements into literature on economic modelling; the author uses the prescription formulated by Max Weber that social scientists should openly state values and policy ends they accept while doing research and he adds this requirement to Uskali Mäki's 'model of a model' (Robert Mróz). In the final article the authors explicate the very term of 'integration of sciences' in order to disentangle it from the concepts of unification and interdisciplinarity; they support their account on integration with a case study and argue that the methodology of humanities may play an important function in integration trials (Jarosław Boruszewski, Krzysztof Nowak-Posadzy).

The editors of this volume of *Studia Metodologiczne (Dissertationes Methodologicae)* believe that this issue will foster more systematic and deepened insights into the culture(s) of modelling in science(s). Still, there are other research practices in science to be explored through the cultural lens, namely theorizing, measurement, experimentation or simulation.

Jarosław Boruszewski, Krzysztof Nowak-Posadzy

STUDIA METODOLOGICZNE
NR 39 • 2019, 11–48
DOI: 10.14746/sm.2019.39.1

Bernhard Thalheim

Models and their Foundational Framework

Abstract. The term model is mainly used in two meanings which are considered to be different: a model of a problem domain as a conceptualisation; a model of a set of formulas as an interpretation in which every formula within this set is true. A general theory of models has not yet been developed. H. Stachowiak proposes a phenomenal approach and 'defines' models by their properties of mapping, truncation and pragmatics. Meanwhile, a notion of the model has been developed. At the same time, it seems that there are rather different understandings of model in sciences and especially Mathematical Logics. Sciences treat models as reflections of origins. Mathematical logics considers models as an instantiation in which a set of statements is valid. So, mathematical model theory is often considered to be a completely different approach to modelling. We realise however that mathematical model theory is only a specific kind of modelling. We show that the treatment of models in logics and in sciences can be embedded into a more general framework. So, the theory of models is based on a separation of concern or orientation.

Keywords: models, mathematical model, instrument, framework, unifying theory.

1. Introduction

Modelling is a topic that has implicitly been in the center of research in science and engineering since its beginnings. It has been considered as a side issue for long time. During the last 40 years it has gained more attention and becomes nowadays a subdiscipline in many disciplines. The compendium [TN15] introduces models in agri- culture, archeology, arts, biology, chemistry, computer science, economics, electrotechnics, environmental sciences, farming, geosciences, historical sciences, languages, mathematics, medicine, ocean sciences, pedagogical science, philosophy, physics, political sciences, sociology, and sports. The models used in these disciplines are instruments used in certain scenarios. So, essentially it is an old subdiscipline of most natural sciences with a history of more than 2.500 years [Mu¨ l16][1]. It is often restricted to Mathematics and mathematical models what is however to much limiting the focus and the scope.

The *modelling method* is a specific science method that uses models as instruments with certain intention or goal, e.g. for solving a task. The model represents or deputes origins. The model is used instead of the origin due to its properties, esp. adequacy and dependability. The modelling method thus consists (i) of the development of 'good' models, (ii) of the utilisation of the model according to the goal, (iii) of the compilation of the experience gained through model utilisation according to the goal, and finally (iv) of generalisation of the experience back to the origins. So, a model must be well-build for this goal, must be enhanced by methods that support its successful deployment, and must support to draw conclusions to the world of its origins.

1.1. A Model is an Adequate and Dependable Instrument

A **model** *is a well-formed, adequate, and dependable instrument that represents origins* [Tha14, Tha17a].

Its criteria of well-formedness, adequacy, and dependability must be commonly accepted by its *community of practice* within some *context* and correspond to the *functions* that a model fulfills in *utilisation scenarios*.

As an instrument or more specifically an artifact a model comes with its *background*, e.g. paradigms, assumptions, postulates, language, thought community, etc. The background its often given only in an implicit form. The background is often implicit and hidden.

A well-formed instrument is *adequate* for a collection of origins if it is *analogous* to the origins to be represented according to some analogy criterion, it is more *focused* (e.g. simpler, truncated, more abstract or reduced) than the origins being modelled, and it sufficiently satisfies its *purpose*. Well-

[1] The earliest source of systematic model consideration we know is Heraklit with his λογος (logos). Model development and model deployment is almost as old as the mankind, however.

formedness enables an instrument to be *justified* by an empirical corroboration according to its objectives, by rational coherence and conformity explicitly stated through conformity formulas or statements, by falsifiability or validation, and by stability and plasticity within a collection of origins. The instrument is *sufficient* by its *quality* characterisation for internal quality, external quality and quality in use or through quality characteristics such as correctness, generality, usefulness, comprehensibility, parsimony, robustness, novelty etc. Sufficiency is typically combined with some assurance evaluation (tolerance, modality, confidence, and restrictions). Model functions determine which justification is required and which sufficiency characteristics are important. A well-formed instrument is called *dependable* if it is sufficient and is justified for justification properties and sufficiency characteristics.

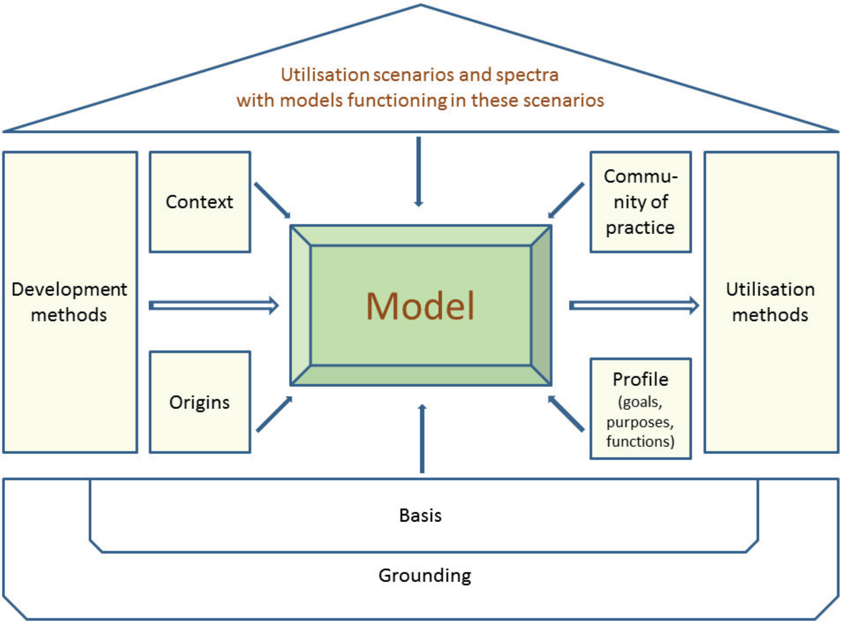

Utilisation scenarios and spectra
with models functioning in these scenarios
Development methods
Context
Origins
Model
Commu-nity of practice
Profile (goals, purposes, functions)
Utilisation methods
Basis
Grounding

Figure 1: The model as an instrument that is adequate and dependable for its driving directives (origins, profile (functions, purposes, goals), community of practice, context) within its background (grounding, basis) and that properly functions in utilisation scenarios as a deputy of its origins

Figure 1 represents a model of the model. The development and utilisation methods form the enabling aspects of the modelling method. Driving directives are (1) origins to be represented by the model, (2) purposes or goals or functions of models[2], (3) the community of its users and developers, i.e. the community of practice, and (4) the context into which the model is embedded. Models function as instruments in application or utilisation scenario. Typical functions of models are (a) cognition, (b) explanation and demonstration, (c) indication, (d) variation and optimisation, (e) projection and construction, (f) control, (g) substitution, and (h) experimentation. A model is not built on its own. It has an undisputable grounding that has to be accepted. The basis of the model - similar to the cellar - can however be disputed. Grounding and basis form the background of a model. We observe that the background is often given only in an implicit form. The same kind of concealment can also be observed for the utilisation scenario which are implicitly given by sample and generalisable case studies for the utilisation frame.

The model is not simply an image of its origins. The mapping property [Kas03, Mah09, Mah15, Sta73] might be too restrictive for models. Instead, we use analogy. Models can also be material artifacts, e.g. exhibition models, architecture models, models used in religion, and three-dimensional demonstration models used in engineering and mathematics [CH04]. A model can be a model of another model. For instance, the topographical Königsberg bridge sketch is a model that is the origin for the graph-theoretical model for the Euler path existence [MV10]. Models might follow different structuring

[2] Goals, purposes, and functions are often considered to be synonyms. We follow the separation of concern as discussed in [TN15]. The *goal* of a model can be defined as a ternary relation between initial states, final states and the community of practice that accepts the final state and considers the initial state. It describes the aim, the ambition, the destination, the end, the intent, the intention, the objective, the prompt, or the target of a model. It is not of interest whether the goal is realistic. The *purpose* of a model extends the goal of the model by means (or instruments) that potentially enable the community of practice to achieve the goal. The *function* of a model embeds the model into practices and scenarios of its application. The function can be considered as an extension of the purpose by `application games' of the model. It specifies the role and the play of a model in the scenarios, i.e. how, when, for which/what or why, at what/which etc. the model functions. The function often implicitly adds conventions of deployment, customs, exertions, habits, specific usage and uses, and handling pattern to the purpose. The model thus functions in the scenarios in the given mould.

and behaviour than the origins. For instance, many models in classical physics concentrate on aspects and do not represent reality, e.g. the Bohr atom model. The Ptolemaic world model has a complete different behavior for most of the bodies of the universe but is was useful (see the nautical tools used as that time or the clock of Antikythera). The hydrological model of the electrical circuit has an analogical behavior that is not based on a mapping. Usefulness and utility according to goals govern the selection of a model instead of quality characteristics such as validity. Finally, a model comes with its background. It cannot be properly understood and used if the background is concealed. Let us distinguish the concepts of goal, of purpose, and function in the sequel. The goal of a model is in general the association between a current state and the target state that is accepted by stakeholders or – more general – by members of a community. The purpose enhances the goal by means that allow to reach the target state, e.g. methods for model development and utilisation. The function extends the purpose by practices or – more systematically – by scenarios in which the model is used. A typical scenario is the modelling method and its specific forms.

1.2. Models in Science and Daily Life versus Models in Mathematical Logics

Models in sciences and model theory in mathematical logic are often considered to be completely different issues [Bal16]. This point of view is correct as long there is no consolidated understanding of a notion of a model. Models in model theory are instantiations of a set formulas. This set of formulas is satisfied by a model according to a logical definition frame. The model is a structure that is defined with the same signature as the set of formulas.

So, we might come to the conclusion that there are at least three different understandings of the model. We will oppose this conclusion in the sequel. It is only true for the Fuzzy or phenomenalistic view.

Models in science typically follow the modelling methods. They may be composed of a number of models and be based on other models. A model must not be true. It should however be coherent to some extent within its discipline

The origin in science is not limited to material origins. The origin itself can be virtual or be again a model, e.g. a mental model. So, the modelling methods may also be iteratively applied.

Models often used in daily life. One kind are metaphors or parables. The typical kind is, however, a pattern for explanation, negotiation, and communication. Models carry a meaning. It is often debated whether a fashion model or a diagram or a visualisation can be considered as a specific kind of a model.

The modelling method presented so far is associated with its origins. We might however also use models for construction of other origins or models. In this case, the model is not generalised but used as a blueprint for another artifact. So, we observe that the modelling method must be extended.

1.3. Models and their Utilisation Scenarios

Models are used in various scenarios, e.g. communication, system construction, perception, analysis, forecasting, documentation, system modernisation and optimisation, control, management, and simulation. Let us in the sequel concentrate on the first three scenarios.

The extended modelling method is embedded into a more general form of activities, i.e. scenarios. The model itself is used as an instrument in a scenario or a bundle of scenarios which we call usage spectrum. It has a function or a number of functions in these scenarios. This functioning must be effectively supported by utilisation methods and is used by members of a community of practice in most cases. For instance, models of situations/states/data are often used for structuring, description, prescription, hypothetic investigation, and analysis. So, we observe that the function (or simpler the purpose or the goal) of the model is determined by the concrete way how a model is used.

A model might be oriented towards this community of practice. It can however also represent the scenarios themselves. It might represent the context of these scenarios, e.g. the scientific or engineering background, the relation to time and space, the application area insight, and the knowledge accepted by the community. It might also be oriented to representation of either a situation and state under consideration or a evolutionary change process.

The different orientations are the basis to distinguish the six concerns for models: community of practice, back- ground/knowlegde/context, application scenario and stories of model utilisation with their specific frames, situation/ state/data, dynamics/evolution/change/operations, and models as representations and instruments. Figure 2 shows the relation between the concerns and the functions a model might have[3]

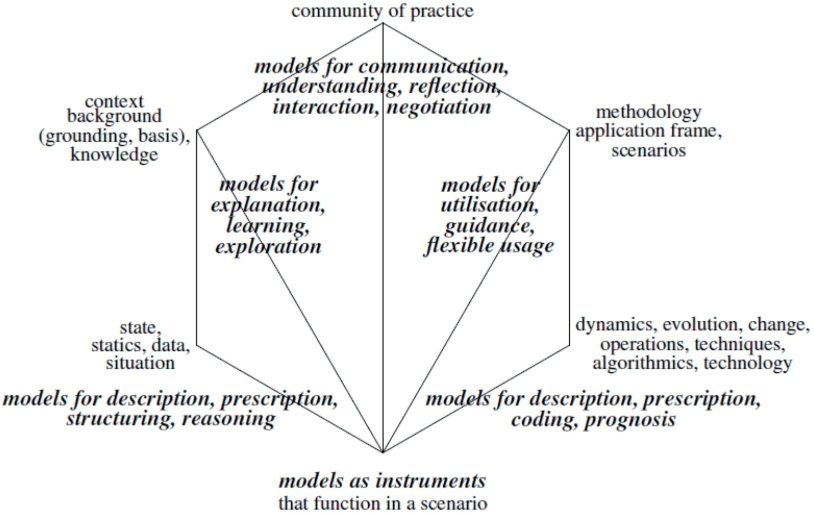

community of practice
models for communication, understanding, reflection, interaction, negotiation
context background (grounding, basis), knowledge
methodology application frame, scenarios
models for explanation, learning, exploration
models for utilisation, guidance, flexible usage
state, statics, data, situation
dynamics, evolution, change, operations, techniques, algorithmics, technology
models for description, prescription, structuring, reasoning
models for description, prescription, coding, prognosis
models as instruments that function in a scenario

Figure 2: Models and the five concerns in model-based reasoning, investigation, and engineering

1.4. The Storyline of the Paper

A general theory of models, of modelling activities and of systematic modelling has not yet been developed although modelling has already attracted a large body of knowledge and research[4]. The notion of the model is not yet commonly accepted. Instead we know a large variety of rather different no-

[3] Modified and revised from [Tha17c].

[4] It is not our purpose to develop a bibliography of model research. Instead we refer to bibliographies in [TN15] and the more than 5.000 entries in R. Müller's website, e.g. [Mül16].

tions. Model development activities have been a concern in engineering. The process of model development has not yet attracted a lot of research. Model deployment also needs a deeper investigation. The model is mainly used as an instrument in certain application scenarios and must thus function in these scenarios. So, a model is a medium.

We have already introduced the general notion of a model as a starting point. The next step could be the development of a general theory of modelling. It is often claimed that modelling is rather different in science and engineering. So, we might conclude that there is no general theory of modelling. This paper is going to show that there is a general theory of modelling. We start with a case study in Section 2. These lessons gained in this cases study are a starting point for a general theory of models, of modelling activities, and of systematic modelling. In Section 3 first elements of this theory are developed.

2. Models in Everyday Life and Sciences: A Case Study

Analysing model notions we realise that there are at least four different approaches:

1. The general phenomenalistic definition uses properties such as mapping, truncation and pragmatic properties for the association between origins and models. Most research on models starts with this approach.

2. The axiomatic definition follows frames used in Mathematical Logics and defines models as exemplifications of formal systems and formal theories. Models thus depute and represent a certain part of reality.

3. The mapping-based definition is based on a direct homomorphic mapping between origin and model. We might have another mapping between model and implemented system that is a realisation of the model.

4. The construction-oriented definition defines a model as being a result of a modelling process by some com- munity of practice.

There is a fifth approach to models which simply uses artifacts as models without any definition, e.g. in human communication and also in sciences[5].

[5] One of the prominent definition is given by John von Neumann [vN55]: "*The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is*

The definition given above follows, however, the mathematical way of defining things through definitional extensions.

Models are used as

- *perception models* reflecting someone's understanding,
- *mental models* that combine various perception models and that make use of cognitive structures and operations in common use,
- *domain-situation models* representing a commonly accepted understanding of a state of affairs within some application domain,
- *experimentation models* that guide experimentation,
- *formal models* based on some kind of formalism,
- *mathematical models* that are expressed in some mathematical language and based on some mathematical methods,
- *conceptual models* which combine models with some concept and conception space,
- *computational models* that are based on some (semi-)algorithm ,
- *informative models* that used to inform potential users about origins,
- *inspiration models* that provide an intuitive understanding of something,
- *physical models* that use some physical instrument,
- *visualisation* models that provide a visualisation,
- *representation models* that represent things like other models,
- *diagrammatic models* that are based on some diagram language with some kind of semantics,
- *exploration models* for property discovery,
- *prototype models* that represent a class of similar items,
- *mould models* that are used for production of artefacts,
- *heuristic models* that are based on some Fuzzy, probability, plausibility etc. relationship, etc.

meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work - that is correctly to describe phenomena from a reasonably wide area. Furthermore, it must satisfy certain esthetic criteria - that is, in relation to how much it describes, it must be rather simple. I think it is worthwhile insisting on these vague terms - for instance, on the use of the word rather. One cannot tell exactly how ``simple" simple is. Some of these theories that we have adopted, some of the models with which we are very happy and of which we are proud would probably not impress someone exposed to them for the first time as being particularly simple."

Although this categorisation provides an entry point for a discussion of model properties, the phenomenon of being a model can be properly investigated. Each category is rather broad and combines many different aspects at the same time. We already introduced a general notion of model. In this Section we will investigate whether the general definition covers all these kind of models for science and also daily life and whether it can be supported by a holistic treatment of models.

2.1. Models in Mathematical Logics

Let us consider only one kind of logics: classical Mathematical Logic based on first-order or higher-order predicate logics. Mathematical logics considers models as an instantiation in which a set of statements is valid. An arbitrary structure of the same signature as the logical language of the statements which satisfies this set of statements is called model or realisation of this set [Tar56]. Similar observations can be drawn for other mathematical logics as well. Mathematical logic has a long tradition of model research. Model theory became its branch and has a deep theoretical foundation. The main language is the first-order predicate logic. This language is applied in a rigid form [ST08] that became a canonical form of Mathematical Logics: It uses a canonical way of associating syntactic and semantic types. Additionally, the semantic type and the syntactic type have the same signature. The expressions of syntactic types are inductively constructed starting with some basic expressions of certain construct by application of expressions of some other construct. For instance, we may start with truth values and variables. Terms and formulas are then based on these basic expressions. The context is not considered. The world of potential structures is typically not restricted. The rigidity however allowed to gain a number of good properties. For this reason, first-order predicate logics became a first-class fundament for Computer Science.

A typical example of a standard set of logical formulas characterizing real numbers (commutativity and associativity of addition, existence of an additive identity and an additive inverses (0), commutativity and associativity of multiplication, existence of an multiplicative identity (1), existence of a multiplicative inverses with the exception of *the additive identity, distributivity of multiplication over addition, linear order). The set of real numbers with classical operations (+,*),predicates (=, <) and the numbers 0,1 as additive and multiplicative identity is the standard model of these logical formulas. Another model is the set of Robinson numbers which are used in a nonstandard approach to analysis based on infinitesimals. Infinitesimals are greater than 0 but smaller than any positive real number.*

In general, a model in Mathematical Logics is defined through its relationship to a set of formulas. These formulas are valid in the model. Additionally, axioms and rules of the first order predicate logics are valid in the model since they are valid in any structure of given signature. Models are thus instantiations (or exemplifications) for a set of statements. The theory of deduction is the main basis for reasoning. Therefore, the five concerns in Figure 2 have the specific peculiarity shown in Figure 3.

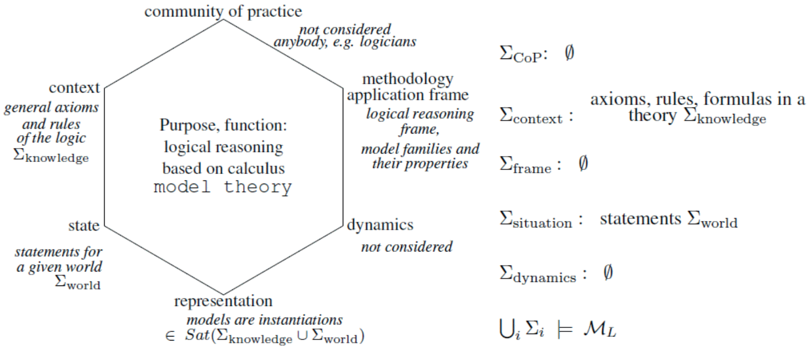

community of practice
not considered
anybody, e.g. logicians
context
general axioms
and rules
of the logic
$\Sigma_{\text{knowledge}}$
Purpose, function:
logical reasoning
based on calculus
model theory
methodology
application frame
logical reasoning
frame,
model families and
their properties
state
statements for
a given world
Σ_{world}
dynamics
not considered
representation
models are instantiations
$\in Sat(\Sigma_{\text{knowledge}} \cup \Sigma_{\text{world}})$
Σ_{CoP}: $\emptyset$
Σ_{context} : axioms, rules, formulas in a theory $\Sigma_{\text{knowledge}}$
Σ_{frame} : $\emptyset$
$\Sigma_{\text{situation}}$: statements Σ_{world}
Σ_{dynamics} : $\emptyset$
$\bigcup_i \Sigma_i \models \mathcal{M}_L$

Figure 3: Models in logics for investigation of situations and expressible properties: axioms and rules form the context world; admissible states are characterised by a set of formulas; models are instances of potential systems that obey the system

The special side of the approach of Mathematical Logics to modelling is the consideration of the set of all potential models together with a given instantiation. This approach is however also taken into consideration for other model kinds as we shall in the sequel.

A model might become then an exemplar or prototype for a given theory. It can represent this theory and thus allows to reason on the given theory. It can be thus a final or an initial model (see the theory of abstract data types [Rei84, Wec92]) where the first one is the best and most detailed representation of the given theory and allows to reason on all potential negative statements as well.

We notice that classically the community of practice is not considered. Also, dynamics ist not an issue. There is not really defined any reasoning frame beside the calculus itself. We are free to choose Hilbert style or Gentzen style or any other derivation style for reasoning.

A specific decision within mathematical logics is the invariance of the signature, i.e. models as structures and logical languages for theory statements share the same signature. Therefore, there is a tight mapping between terms and formulas and the properties that can be stated on the model.

This specific mapping property has also been used for the phenomenal characterisation of models as structures that a based on a mapping from the origin to the model, e.g. [Bal82, Sta73, Ste66, Ste93]. We also observe that the truncation or abstraction property is a specific property of logical models.

2.2. Mathematical Models

Mathematical models are considered to be the most prominent kind of model. A mathematical representation of another 'donor' or *origin model* is based on the mathematical language. The mathematical model is used for solving of problems that have been formulated for the origin model. The association between the mathematical model and the origin model must be problem invariant. Solution faithfulness is often not given explicitly required, i.e. the solution obtained for the mathematical model must be faithful for the origin model. Mathematical modelling presumes the existence of this origin model. So, (1) it starts with an application analysis and a formulation of the problem to be considered in the application area. Next, (2) this formulation is transformed to the origin model which allows to describe the problem. (3) This origin model is then mapped to a mathematical model. (4) The fourth phase is the development of a solution of the problem within the mathematical model. (5) The solution is verified and will be validated for faithfulness within the origin model. Finally, (6) the solution is examined for its reflection in the application area. If the solution is not of the required quality then the phases are repeated. This 6-phase circular frame [GKBF13, Pol45] is a commonly accepted scenario for mathematical modelling.

> *A typical mathematical model is the Euler graph model for the Königsberg bridge problem (for detailed consideration see [MV10]). This graph represents a rough topographical model which is a model of the inner city of Königsberg. The purpose of the Euler graph is to provide a solution of the bridge path problem within a graph theory setting, i.e. providing a path – called Euler path - that uses each bridge once and only once. The solution to this problem is the condition that an Euler path exists if and only if the degree of all nodes is even except maximal two nodes with odd degree.*

We observe that the mathematical frame is similar to the logical reasoning frame. Main quality requirements satisfaction of the problem solving purpose, adequacy of the mathematical model, robustness against minor changes, and potential and capacity for problem solution. The community of practice should not influence the model properties. It may influence on the selection of various representation models. The situation and its dynamics determines the appropriatedness of the mathematical language. The mathematical model is determined by some mathematical method that has shown useful in the past.

Our model notion extends the model discussion by H. Hertz [Her84, vDGOG09]. He postulates that some artefact is a model due to its analogy to origins, its dependence within an application context, its purposefulness, its correctness, its simplicity, and its potentially only implicit given background. Models have thus a validity area.

Mathematical models are specific formal models. They are based on a formalisation that can be mapped to some mathematical language. The mapping from the formal model to the mathematical model should preserve the problem, i.e. it is invariant for the problem. The mapping should additionally also allow to associate the mathematical solution to the problem with a correct or better faithful solution in the formal model and for the origins, i.e. the model is solution-faithful [BT15]. The mathematical language has not only a capacity and potential. It also restricts and biases the solution space. The calculus used for the derivation of the model is any mathematical and not restricted to logical reasoning.

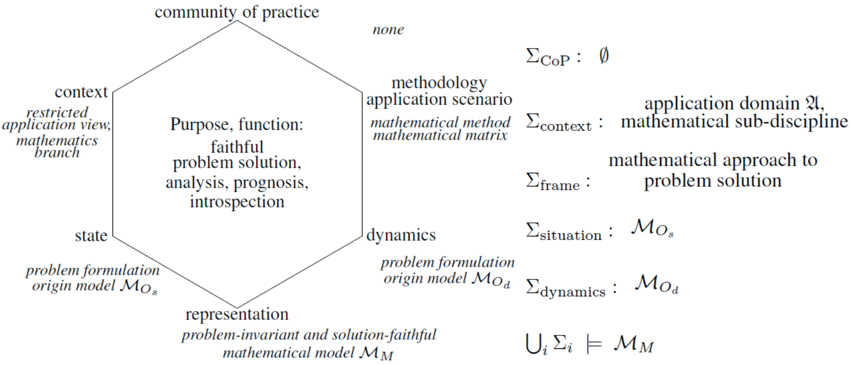
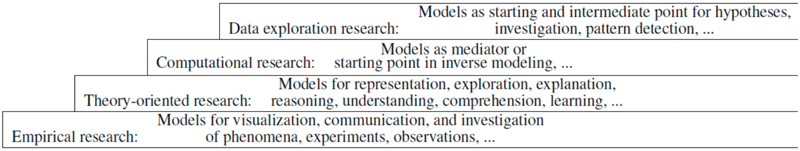

community of practice
none
$\Sigma_{\text{CoP}} : \emptyset$
context
restricted application view, mathematics branch
Purpose, function:
faithful problem solution, analysis, prognosis, introspection
methodology
application scenario
mathematical method
mathematical matrix
Σ_{context} : application domain $\mathfrak{A}$, mathematical sub-discipline
Σ_{frame} : mathematical approach to problem solution
state
problem formulation origin model $\mathcal{M}_{O_s}$
dynamics
problem formulation origin model $\mathcal{M}_{O_d}$
$\Sigma_{\text{situation}} : \mathcal{M}_{O_s}$
$\Sigma_{\text{dynamics}} : \mathcal{M}_{O_d}$
representation
problem-invariant and solution-faithful mathematical model $\mathcal{M}_M$
$\bigcup_i \Sigma_i \models \mathcal{M}_M$

Figure 4: The mathematical model as a representation of a origin model within the mathematical frame

2.3. Science Models

All sciences widely use models. Typical main purposes are explanation, exploration, hypothesis and theory development, and learning. Models are mediators, explainers, shortcuts, etc. We can consider models as the third dimension of sciences [BFA+ 16, TD16, TTF16][6]. Following [Gra07], sciences may combine empirical research that mainly describes natural phenomena, theory-oriented research that develops concept worlds, computational research that simulates complex phenomena and data exploration research that unifies theory, experiment, and simulation. Models are an essential instrument in all four kinds of research. Their function, however, is different as illustrated in Figure 5 [BFA+ 16].



Data exploration research: Models as starting and intermediate point for hypotheses, investigation, pattern detection, ...

Computational research: Models as mediator or starting point in inverse modeling, ...

Theory-oriented research: Models for representation, exploration, explanation, reasoning, understanding, comprehension, learning, ...

Empirical research: Models for visualization, communication, and investigation of phenomena, experiments, observations, ...

Figure 5: Some model functions according to the kind of scientific research

[6] The title of the book [CH04] has inspired this observation

Empiric research also uses a canonical modelling mould. Beside an ad-hoc mould we might use a sophisticated one: (1) define a research question (based, for instance, on the rhetoric frame (who, what, when, where, why, in what way, by what means (Hermagoras of Temnos, also Augustine's De Rhetorica) or W*H framework [DT15]), (2) consider threats to the research, (3) choose a research model (e.g. positivistic), (4) develop an approach how facts become theories, (5) create a generic meta-model (with some level of abstraction, with independent and dependent parameters and indicators), (6) define analysis approaches (qualitative or quantitative), (7) define the research pro- gram and agenda as a specific research process, (8) select the research method, (9) analyse the capacity and potential of quantitative data, (10) design the experiment, (11) design the case study, and (12) design the outcomes survey.

The empirical research approach often combines qualitative and quantitative approaches. The quantitative approach is often oriented on observable data whereas the qualitative approach orients towards theory, on concepts and conceptions, and on a characterisation of the situations of interest. The quantitative theories are often 'phenotypical' approaches contrary to the 'genotypical' approaches used in qualitative approaches. A typical approach is used in the collaborative research centre 1266[7]. It uses additionally an investigative reasoning approach. Figure 6 shows the differences between genotypical and phenotypical models. We use a planar representation of the three dimensions: (1) the composition dimension with sources, concepts, and theories; (2) the kind dimension with qualitative and quantitative reasoning, and (3) the model dimension that allows to concentrate on certain aspects of the first dimensions depending which function, purpose and goal the model should satisfy. A typical specific treatment of concepts is applied in modelling. Since models orient on certain aspects and represent also combined representations, concepts used in models are often not directly derived from concepts in the theory. Additionally, we should distinguish between quantitative, investigative, and quantitative models. The model kind in Figure 6 uses investigative reasoning and lends some elements from quantitative and qualitative theories beside the theory offering that are used for investigative reasoning. The quantitative theory should also be reflected in the qualitative theory.

[7] Scales of Transformation – Human-Environmental Interaction in Prehistoric and Archaic Societies: https://www.sfb1266.uni-kiel.de/en

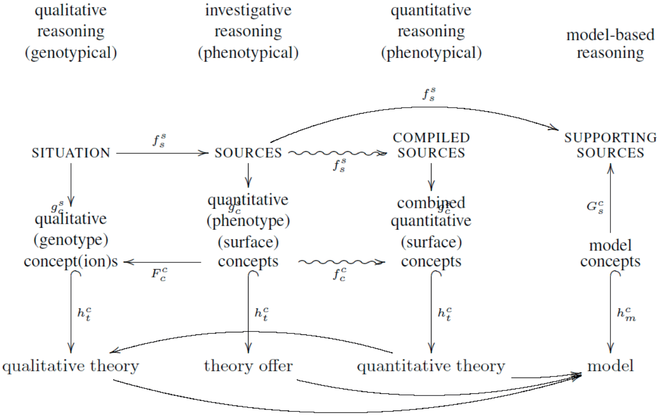


Figure 6: Models for investigative and quantitative reasoning in empirical research

A qualitative theory uses a concept or conception space that represents situations of interest (may be based on some mapping *g* from source data to concepts and some *G* mapping from concepts to sources*)*. The situation can be observed and characterised by sources (may be based on some *f* or *F* mapping between sources). Empirical research in sciences often differentiates between an investigative reasoning and quantitative reasoning. Both use phenotypical observations on proxies. Quantitative approaches aggregate and combine the source data and thus allow to reason on correlation, dependencies, time and spatial relationships. The first two reasoning approaches should be based on a commuting diagram, i.e. the *g*-mapping of a situation equals to the *f*-*g*-*F*-mapping of this situation.

Evidence-based proxy modelling and reasoning treats models in a different way.

(α) Models represent only acceptable possibilities. Each model captures a distinct set of possibilities to which the current description refers) which are consistent with the premises and the knowledge gained so far what makes them intrinsically uncertain because they mirror only some properties they represent.

(β) Models are proxy-driven. The structure of the model corresponds to the proxies it represents.

(γ) Models represent only what has been observed and not what is false in each possibility in contrast to fully explicit models (also representing what is false).

(δ) The more proxies that are considered, and the richer those models are, the more accurate the world view is. (ϵ) Additionally, we use pragmatic reasoning schemata, e.g. A causes B ; B prevents C ; therefore, A prevents C .

The model themselves illustrate then concepts. Therefore, sources support concepts and conceptions what inverts the mapping (*G mapping* instead of *g* mapping).

Let us now consider the theory-oriented research. The frame for empirical research is similar to communication frames in Subsection 2.5. We neglect inverse modelling [Men89] although it is an important approach to science and it has been reconsidered and generalised under various other names, e.g. [ASG13, Noa09, SV05, BST06, TT13]. Data science approaches have been considered in [KT17].

So, we arrive with the hexagon in Figure 7.

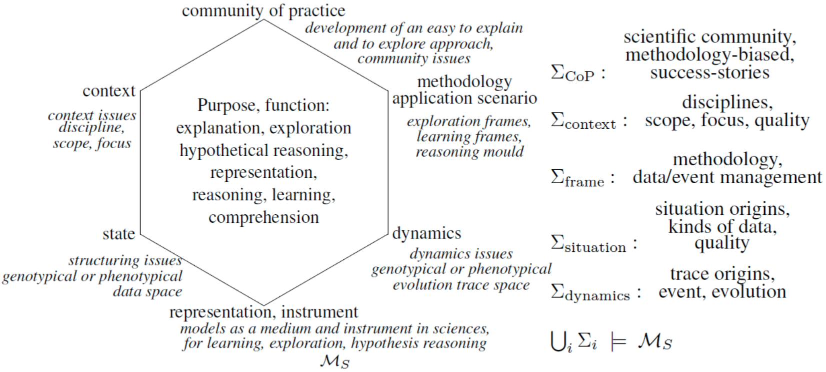

community of practice
development of an easy to explain and to explore approach, community issues
context
context issues discipline, scope, focus
methodology application scenario
exploration frames, learning frames, reasoning mould
Purpose, function: explanation, exploration hypothetical reasoning, representation, reasoning, learning, comprehension
state
structuring issues genotypical or phenotypical data space
dynamics
dynamics issues genotypical or phenotypical evolution trace space
representation, instrument
models as a medium and instrument in sciences, for learning, exploration, hypothesis reasoning
$\mathcal{M}_S$
Σ_{CoP} : scientific community, methodology-biased, success-stories
Σ_{context} : disciplines, scope, focus, quality
Σ_{frame} : methodology, data/event management
$\Sigma_{\text{situation}}$: situation origins, kinds of data, quality
Σ_{dynamics} : trace origins, event, evolution
$\bigcup_i \Sigma_i \models \mathcal{M}_S$

Figure 7: Models are used in natural, social, and other sciences as enhancements and contributions to sciences and as instruments: science contribution, explanation, exploration, learning, comprehension, intellectual absorption, simulation, and reasoning sce- nario

Models function as instruments within the science. They are vehicles for investigation, for analysis, for discovery of alternatives, for prognosis, for exploration, for explanation, for intellectual absorption, for learning, for understanding, for scoped and focused comprehension, for representation of certain aspects, for discussion with partners within their background, for quick illustration, etc. They are supported by various kinds of reasoning. It seems that this variety is rather broad. If we however orient our investigation on the scenarios then we discover that the model utilisation scenarios determine the function of the model. At the same time, the background with the grounding and basis strikes through. Models are biased by their foundations, by their development and utilisation methods, their communities of interest, and their context. A specific context is the school of thought [Bab03, Fle11]. The concept space determines what could the content and the scope of a model. The MMM compendium [TN15] illustrates that models, the approach for to model, and modelling share a good number of common approaches.

2.4. Conceptual Models

Conceptual models are widely used in Computer Science and more specifically in Computer Engineering. In Computer Science and Computer Engineering, one main scenario is (1) the model-based construction of systems beside (2) the explanation and exploration of an application, (3) description of structure and behaviour of systems, and the (4) prognosis of system properties. Model-based construction might include conceptualisation. The application scenarios mainly follows the description-prescription frame. The model is used as a description of its origin and as a prescription of the system to be constructed. The notion of conceptual model is not commonly agreed however[8]. In a nutshell, a *conceptual model* is an language-determined enhancement of a model by concepts from a concept(ion) space.

The conceptual modelling method uses a canonical style of model development and utilisation. Models are instruments in perception and utilisation scenarios. They function is explicitly defined, e.g. models for design and synthesis. The scenario can incorporate a decision point that stops after understanding the perception and domain-situation models or that designs and synthesises the conceptual model after a preparation phase. The last stage support then evaluation and acceptance of the model.

[8] We know almost threescore different notions what shows the wider controversy about this notion[Tha18a]. E.g., Wikiquote (see [Wik17]) lists almost 40 notions. Facetted search for the term “conceptual model” in DBLP results in more than 5.000 hits for titles in papers (normal DBLP search also above 3.400 titles)

So, Figure 8 displays the more specific way of conceptual modelling for information systems. The IS com- munity with its actors shares an IT orientation. It might however be in conflict with the business users. They reason in a different way and are often using a local-as-viewpoint approach. The global-as-design approach might not provide an appropriate support. The model development and utilisation becomes canonical after the choice of the enabling language and the modelling method. The origin models such as the perception and domain-situation models follow the style accepted in these communities. The global-as-design approach must then provide appropri- ate aggregations and derivations for support of local viewpoints. The community also shares the assumption of strict separation of specification into syntax and semantics with the firstness paradigm [KL13, Pei98] for structures and the secondness [Cas55] of functions and views. The model to be developed inherits all the paradigms, assumptions, biases, conceptualisations, cultures, background theories, etc.

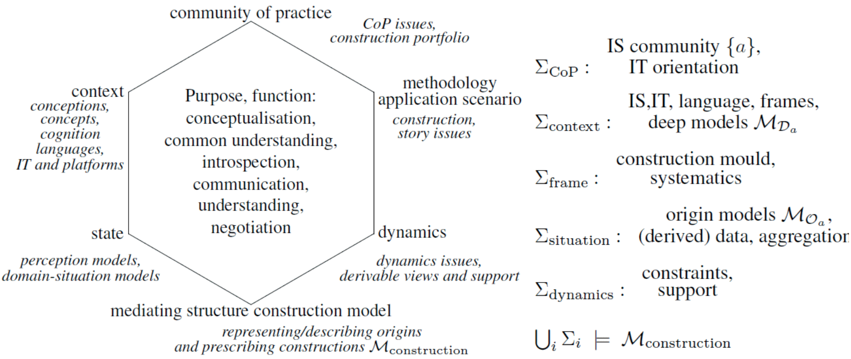

community of practice
CoP issues, construction portfolio
context
conceptions, concepts, cognition languages, IT and platforms
Purpose, function: conceptualisation, common understanding, introspection, communication, understanding, negotiation
methodology application scenario
construction, story issues
state
perception models, domain-situation models
dynamics
dynamics issues, derivable views and support
mediating structure construction model
representing/describing origins and prescribing constructions $\mathcal{M}_{\text{construction}}$
Σ_{CoP} : IS community $\{a\}$, IT orientation
Σ_{context} : IS,IT, language, frames, deep models $\mathcal{M}_{\mathcal{D}_a}$
Σ_{frame} : construction mould, systematics
$\Sigma_{\text{situation}}$: origin models $\mathcal{M}_{\mathcal{O}_a}$, (derived) data, aggregatior
Σ_{dynamics} : constraints, support
$\bigcup_i \Sigma_i \models \mathcal{M}_{\text{construction}}$

Figure 8: Conceptual models for IS structuring

A typical example for conceptual modelling is entity-relationship modelling. [Tha18b] observed a large number of paradigms, postulates, specific modelling cultures, commonsense, practices, and assumptions such as global-as- design (with derivate ion of local viewpoints), Salami slice typing (for homogenisation of object structure within a class), set semantics (instead of multi-set semantics that is used for implementation), uniqueness of names within a schema, hidden implementation assumptions, specific styles for

> *model composition one must follow, well- formedness conditions, etc. Some approaches add also requirements such as strict binarisation of all relationship types.*

The notion of conceptualisation, conceptual models, and concepts are far older than considered in Computer Science. The earliest contribution to models and their conceptualisations we are aware of is pre-socratic philosophy and especially the work by Heraclitus [Leb14].

2.5. Models for Communication and Human Interaction

Human communication heavily uses models. They are often not called models. Some models might be metaphors or prototypes. Other models might be incomplete or not really coherent or consistent. They are however used for exchange of opinions among users. Models function in communication scenario as a medium. The communication itself determines the role and thus the function and therefore the purpose of the model. Models represent in this case a common understanding of the communication partners. They are biased by these partners. Communication is based on some common understanding about the topic that is under consideration. Partner have already agreed on some background. They use this agreement within their communication. This agreement is based a common reflection and some common model. This model is taken for granted and not further discussed in communication. So, partners agree on some background or deep model. Typically, deep models [KT17, Tha17b] are not explicitly communicated. We need however an understanding of a theory of deep model and return to it in the next Section. The model is used for a shared understanding, for sense making, for reflection, for derivation of open issues, and for negotiation. Human interaction is typically context-biased and agent-oriented. The same utterance might have many meanings depending on the context, the receiver, and the receiver-sender relationship. A classical easy to understand model is the topographical model of the Königsberg bridge problem [MV10]. The topography abstracts from any non-essential details while adding some abstractions for the identification of the bridges and the teritorries.

The hexagon in Figure 9 shows the differences between models in Mathematical Logics or sciences and communication model. The main difference is the explicit community dependence of such models. Each of the partners or agents *a* has some understanding of the world. This understanding is the main ingredient of a personal model that we call below perception model. The perception model also reflects the setting of the agent, especially the orientation and the priming. The communication might also be based on some common understanding, i.e. on a situation model. The situation model represents the common world view, shared knowledge and beliefs, and shared opinion. The modelling methods is governed by communication and human interaction. So, we might base the frame on the dialogue and interaction frame. Models play a different role. They are used for common understanding. Typical specific models for human interaction are metaphors [Lak87].

Our second case shows the differences and also commonalities between Mathematical Logics and human interaction. The model must suffice all hidden agreements within the community of practice, the context, and the specific scope and focus taken by the agents. Therefore, the logics becomes now more advanced. Mathematical Logic as the opposite is oriented on general laws and thus not oriented on one model but rather on a family of models.

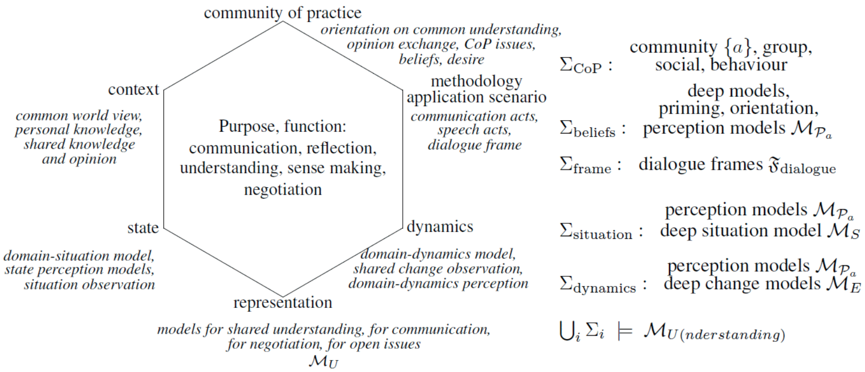

community of practice
orientation on common understanding, opinion exchange, CoP issues, beliefs, desire
context
common world view, personal knowledge, shared knowledge and opinion
methodology application scenario
communication acts, speech acts, dialogue frame
Purpose, function: communication, reflection, understanding, sense making, negotiation
state
domain-situation model, state perception models, situation observation
dynamics
domain-dynamics model, shared change observation, domain-dynamics perception
representation
models for shared understanding, for communication, for negotiation, for open issues
$\mathcal{M}_U$
Σ_{CoP} : community $\{a\}$, group, social, behaviour
Σ_{beliefs} : deep models, priming, orientation, perception models $\mathcal{M}_{\mathcal{P}_a}$
Σ_{frame} : dialogue frames $\mathfrak{F}_{\text{dialogue}}$
$\Sigma_{\text{situation}}$: perception models $\mathcal{M}_{\mathcal{P}_a}$ deep situation model $\mathcal{M}_S$
Σ_{dynamics} : perception models $\mathcal{M}_{\mathcal{P}_a}$ deep change models $\mathcal{M}_E$
$\bigcup_i \Sigma_i \models \mathcal{M}_{U(nderstanding)}$

Figure 9: Models in human interaction: development of common understanding, exchange of opinion, communication, reflection, negotiation; context on the basis of commonalities in world views as deep models; scenario based on communication acts

Communication is often based on models that refer to some common understanding of humans. Models in communication might become situation-domain models which describe a common understanding of the phenomena in an area of interest, e.g. structuring of data in a business application case.

2.6. Lessons Learned with the Case Studies

We may now summarise the experience we gained:

- We realise by these case studies that there exists a common framework to models, to the activities of modelling and to modelling as a systematics reflection, for development of models, and for utilisation of models.
- Models are used to represent certain issues. They are more focused and must serve its purpose. The purpose and the focus determine which kind of adequacy is appropriate.
- Models do not exist on their own. They represent something in the world. The world under consideration depends again on the modelling frame. In most cases, mental models and perception or situation models are the origins which are reflected by the model.
- The justification must be given in a way that can be accepted by its community of practice. Models are developed by some members of this community and are utilised by some – may be other – members of this community of practice. So, models must be satisfying. Therefore, we need an explicit understanding of the sufficiency and thus quality of the given model.
- Models are composed of models that reflect their background and of models that represent specific states and situations within from one side and specific dynamics.
- Models are used as instruments in certain scenarios. They have a number of specific functions in these scenarios.
- Models are typically multi-models, i.e. an association of models which are reflecting specific sides of the same issue depending on the viewpoint that is actually considered. Since such models must be coherent we may bundle them within a model suites [DT10, Tha10].

- Model development and model utilisation typically follow canonical stories. An example is mathematical modelling that consists of a six-step procedure. similar procedures can be observed for most sciences that start with a research question, initialise a certain research agenda or problem solving program or schedule, adapt elements to be used to this program, and then solve a problem. Solution-faithfulness is assumed as a hidden quality characteristics beyond the problem invariance. Modelling is typically based on some specific method or methodology, e.g. the mathematical method. These methods are a mould for the modelling process itself, e.g. a pattern, template, stereotype, work-holding attachment, and an appliance. The method itself follows a macro-model.
- Modelling is still a big challenge to science and has a lot of lacunas. The biggest lacunas seem to be the missing support for combined model-based reasoning. Conceptual modelling uses a specific kind of layered model-based reasoning with changing reasoning methods depending on the stage of model development and model utilisation, e.g. in greenfield development of conceptual models: settlement of the context and the method, transfer of mentalistic concepts to codified ones with a concept expression language, transfer of domain-situation models to raw conceptual models, language-backed negotiation and agreement on a number of conceptual models that allow reflexion of different viewpoints, maturation of these conceptual models, and proper documentation. The reasoning method changes according to the stages. The integration of all these reasoning methods into a holistic one is not required.

3. Towards a General Theory of Models

3.1. Deep Models and the Modelling Matrix

The context and methodology layer determines the set-up of the model. It is often taken for granted and as given. It makes modelling more economical and also more reliable. A number of quality characteristics can be thus satisfied without any further consideration. Model development is typically based on an explicit and rather quick description of the ’surface’ or normal model and on the mostly unconditional acceptance of this set-up. In reality, this setting becomes an essential part of the model. We call it deep model [Tha18b]. It directs the modelling process and the surface or normal model. Modelling itself is often understood as development and design of the normal model. This approach has the advantage that the deep model can be used as a basis for many models.

The set-up is the modelling matrix behind the model. It consists of the grounding for modelling (paradigms, postulates, restrictions, theories, culture, foundations, conventions, authorities), the outer directives (context and community of practice), and basis (assumptions, general concept space, practices, language as carrier, thought community and thought style, methodology, pattern, routines, commonsense) of modelling. It uses a collection of undisputable elements of the background as grounding and additionally a disputable and adjustable basis which is commonly accepted in the given context by the community of practice.

The modelling matrix is often given as a stereotype one should follow while developing the normal model. Adequacy and dependability of is partially already defined by such stereotypes. The stereotype of a modelling process is based on a general modelling situation.

Stereotypes determine the model kind, the background and way of modelling activities. They persuade the activities of modelling.

3.2. The Five Concerns and a General Approach to Modelling

The case studies led us to the conclusion that there is a common three-layer setting in modelling:

(1) **Community and scenario setting:** The community governs the function that a model has to serve according to their issues and scenario.

Community of practice and application cases: The community of practice has its needs and desires. It faces a number of application cases. The application case consists of tasks that should be accomplished. These tasks form the community portfolio. The application cases can be solved by a model, i.e. the model functions as an instrument. Community members determine which model functions best. The community agrees on the issues for modelling.

(2) **Guiding settings:** The deep model and the matrix is commonly agreed according to the setting in the first layer.

Context: Modelling has its implicit and sometimes also its explicit context. Knowledge and disciplinary schools of thought and understanding are considered to be fixed. In a similar form, the background is fixed. This context forms the deep model that underpins the entire modelling process. A typical element of the deep model is the school of thought.

Modelling methodology and application mould: Modelling follows typically practices that are accepted within the community of practice. These practices are often stereotyped. The methods that are used for model development

(3) **Origins and targets:** Members of the community form their personal perception models and share their domain- situation model that characterises states and dynamics in the application domain that is of interest. These models are the origins on which the normal model is formed as an extension of the deep model.

The final result is a model that combines the normal and the deep models. The representation of the final model must not show all details of the deep model.

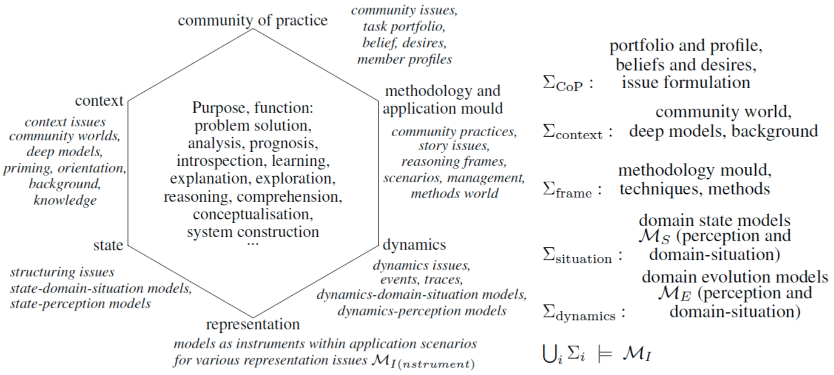

community of practice
community issues, task portfolio, belief, desires, member profiles
context
context issues community worlds, deep models, priming, orientation, background, knowledge
Purpose, function: problem solution, analysis, prognosis, introspection, learning, explanation, exploration, reasoning, comprehension, conceptualisation, system construction …
methodology and application mould
community practices, story issues, reasoning frames, scenarios, management, methods world
state
structuring issues state-domain-situation models, state-perception models
dynamics
dynamics issues, events, traces, dynamics-domain-situation models, dynamics-perception models
representation
models as instruments within application scenarios for various representation issues $\mathcal{M}_{I(nstrument)}$
Σ_{CoP} : portfolio and profile, beliefs and desires, issue formulation
Σ_{context} : community world, deep models, background
Σ_{frame} : methodology mould, techniques, methods
$\Sigma_{\text{situation}}$: domain state models $\mathcal{M}_S$ (perception and domain-situation)
Σ_{dynamics} : domain evolution models $\mathcal{M}_E$ (perception and domain-situation)
$\bigcup_i \Sigma_i \models \mathcal{M}_I$

Figure 10: The five concerns for models as a kernel for a theory of models and of modelling

This general setting takes us back to the rhetorical frame[9] and its generalisation to the W∗H specification frame- work [DT15]: In our case, the model ("what") incorporates the meaning of parties (semantical space; "who") during a discourse ('when') within some application with some purpose ("why") based on some modelling language.

We thus distinguish between *five grounding and driving perspectives* to models:

- *Community perspective:* The community has intentionally set-up its application cases, its interests, its desires and its portfolio. The community communicates, knows languages, explains, recognizes, accept the grounding behind the models, has been introduced to the basis and is common with it. Models are used by, developed by and for, and gain a surplus value for a community of practice. They may have a different shape, form, and value for community members. They must, however, be acceptable for its community. Typical specialisations of this concern are 'by whom', 'to whom', 'whichever', and 'worthiness'.
- *Purpose, function, goal perspective:* Models and model development serve a certain purpose in some utilisation scenarios. The model has to function in these scenarios and should thus be of certain quality. At the same time it is embedded into the context and is acceptable by a community of practice with its rules and understandings. We answer 'why' and 'for which reason' questions.
- *Product perspective:* Models are products that are requested, have been developed, are delivered according to the first perspective, are potentially applicable within the scenarios, and have their merits and problems. Typ- ical purpose characteristics are answers to 'how-to-use', 'why', 'whereto', 'when', for which reason" and
- 'wherewith' (carrier, e.g., language) questions.
- *Engineering perspective:* Models are mastered within an engineering process based on some approaches to mod- elling activities and to uti-

[9] It relates back to Hermagoras of Temnos or Cicero more than 2000 years ago., i.e. they are characterised through "who says what, when, where, why, in what way, by what means" (Quis, quid, quando, ubi, cur, quem ad modum, quibus adminiculis). The Zachman frame used in software engineering and computer science is only at its best a reinvention of this frame.

lisation of models. Modelling is a systematically performed process that uses meth- ods, techniques, preparations, and experience already gained in former modelling processes. The modelling method is typically given in a canonical form. It guides and steers the model development and the model utilisation processes. This guidance can be derived from the scenarios in which the model functions.

- *Background and context perspective*: Model development and utilisation is a systematic, well-founded process that allows one to reason on the capacity and potential of the model, to handle adequacy and dependability of models in a proper way, and the reason on the model and its origins that it represents. A modelling culture also answers the by-what-means question beside providing the background. The background is typically con- sidered to be given and not explicitly explained. It consists of an undisputable grounding and of a disputable and adjustable basis. The context clarifies on which basis and especially on which grounding the model has been developed and must be restricted in its utilisation. Additional context characteristics are answers to questions about the 'whereat', 'whereabout', 'whither', and 'when'.

Models are typically laden by these grounding and driven perspective. They are the hidden part of the deep model. In daily practice, modelling is mainly modelling `beside or additional' to the deep model. The modelling matrix is also taken for granted. That means, modelling is mainly normal modelling that incorporates and unconditionally accepts the perspectives.

3.3. Model-Based Reasoning

The observation depicted in Figure 6 drives us to a multi-model approach. We build models in situations, concepts and theories in dependence on their function and purpose. The same situation-concept-theory may be the basis for a variety of models. A typical multi-model approach is the consideration of models in Physics. Models should thus be considered to be the third dimension of science [BFA+ 16, TN15, TTF16]. Disciplines and also human communication, human interaction, and human collaboration have developed a different understanding of the notion of model, of the function of models in scientific research and communication. Models are often considered to be artifacts. In reality, they are however instruments that are used with a certain intention. Models might also be perception models that incorporate mentalistic concepts [Jac04]. Models are used in various *utilisation scenarios* such as construction of systems, verification, optimization, explanation, and documentation. *In these scenarios* they *function* as *instruments* and thus satisfy a number of properties.

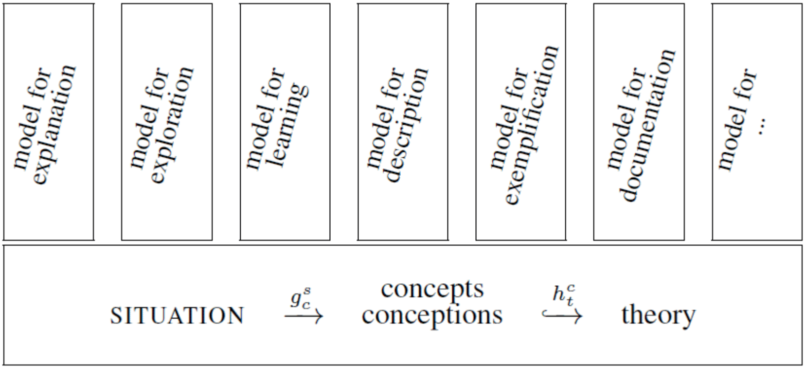


Figure 11: Models as specific representations of situations, concept(ion)s, and theories

Model-based reasoning [Bre10, Mag14] is enhances classical reasoning such as reasoning mathematical calculi or logical derivation. There are several kinds of reasoning that are more appropriate and widely used:

Evidence-based modelling and reasoning is one of the main approaches for quantitative models. Models only represent acceptable possibilities. Each model captures a distinct set of possibilities to which the current de- scription refers. Possibilities are consistent with the premises and the knowledge gained so far what makes them intrinsically uncertain because they mirror only some properties they represent. In investigative and quantitative modelling, models can be proxy-driven where the the structure of the model corresponds to the proxies it represents. They might also include abstractions such as negation which must be then stratified. Propositional evidence-based reasoning is based on monotone functions and specific interpretations for log- ical connectives. Models represent in this case only what has been observed and not what is false in each possibility what is different from fully explicit models which represent what is false. The more proxies are considered the richer those models are, the more accurate the world view is. Evidence-based modelling and reasoning uses pragmatic reasoning schemata, e.g. A causes B; B prevents C; therefore, A prevents C. The calculus may use several implication forms, e.g. deterministic conclusions (A cause B to occur: given A then B occurs) and ordered sets of possibilities (A enables B to occur: given A then it is possible for B to occur).

Hypothetical and investigative modelling considers different assumptions in order to see what follows from them, i.e. reasons about alternative possible worlds (i.e. states of the world), regardless of their resemblance to the actual world. Potential assumptions with their possible world conclusions and assertions are supported by a number of hypotheses (allowing to derive them). It is often combined with abductive reasoning. Evidence against hypothesis is performed by testing its logical consequences, i.e. exploring different alternative solu- tions in parallel to determine which approach or series of steps best solves a particular problem.

Causal reasoning and modelling is a specific variant of inductive reasoning and justification-backed truth main- tenance with assertions (beliefs, background) and justifications within some context (current beliefs, justifications, arguments). It establishes the presence of causal relationships among events based on methods of agreement, difference, concomitant variation, and residues. It uses assumptions and thus avoids inconsistent sets ('nogood' environment). The environment consists of a set of assumptions, premises, assumed state- ments, and derived statements for the world view. Justifications (e.g. data-supported) represent cause. Hypotheses are not derived from evidence but are added to evidence. They direct the search for evidence. They are tested by modus tollens ($(H \rightarrow I) \wedge \neg I \Rightarrow \neg H$).

Network reasoning uses models that are expressed as networks. Nodes carry justification (arguments) and status (in, out, believed, relevant, necessary, …). Edges, hyperedges, or directed edges have an antecedent (support nodes) and conclusions. They may also be non-monotonic and enable backtracking for dependencies (causal- ity, chronological, space, etc. Labels also express the degree of consistency and believability. Queries can be expressed as subgraphs and are evaluated by query embedding into the network.

Model-based reasoning is an interactive and iterative process that helps to digest a theory and to develop the theory. Therefore, model-based reasoning integrates many reasoning approaches, e.g. deduction, induction, abduction, Solomonoff induction, non-monotonic reasoning, and restrospective reasoning. Model refinement might also be based on inverse modelling approaches. Facets of the last one are inductive learning, data mining, data analysis, generic modelling, universal applications, systematic modelling, and pattern-based reasoning.

3.4. Towards Powerful Methodological Moulds

The hexagon picture and the consideration of the variety of different (reasoning) techniques might lead to the impression that a general treatment of models and a methodological support is infeasible. Sciences and humans have however developed their specific approaches and overcome the challenges of this complexity. We will illustrate resolution of complexity by two methods: Layered treatment and generic modelling. Both approaches are based on the separation of a model into a deep or core model and a normal model. A typical example of a methodology is the mathematical modelling method [BT15, GKBF13, vDGOG09, Pol45] (see Subsection 2.2). The CRISP cycle (data selection according to generic model, data preprocessing, data transformation, data mining, model development, interpretation/evaluation) [BBHK10] and classical investigation cycles (define issues and functions of the model, hypothetically predict model properties, experiment, (re)define model, apply and validate the model against the situation) are typical methodologies. Similar methodologies are known for data mining [Jan17], data analysis [BBHK10], and systematic mathematical problem solving [Pod01]. They use a variety of reasoning techniques and layer their application of these techniques according to the stage that is currently under consideration. These modelling methods and methodologies are used similar to moulds that are commonly used in manufacturing.

Data mining [Jan17], inverse modelling [RSS+ 10], and generic modelling [TTFZ14] start with a generic model.

A set of associated models (called model suite) is the result of a modelling process. We may develop a singleton model or a model suite. Figure 12 displays a variant that starts with an initialisation and setting of the modelling process. The initialisation is based on the issues that are important for the community of practice, the tasks that are on the agenda, and the injection of the context. The community of practice aims at completion of tasks from its portfolio and is bound by profiles of their members what also includes beliefs and desires shared in this community. At the same time, the methodology for modelling is already chosen. That means, the upper dimensions in Figure 10 governs the entire modelling process. A similar approach can be declared for model redevelopment model evolution instead of model development from scratch (greenfield modelling). The result of the first layer is a deep model and a matrix.

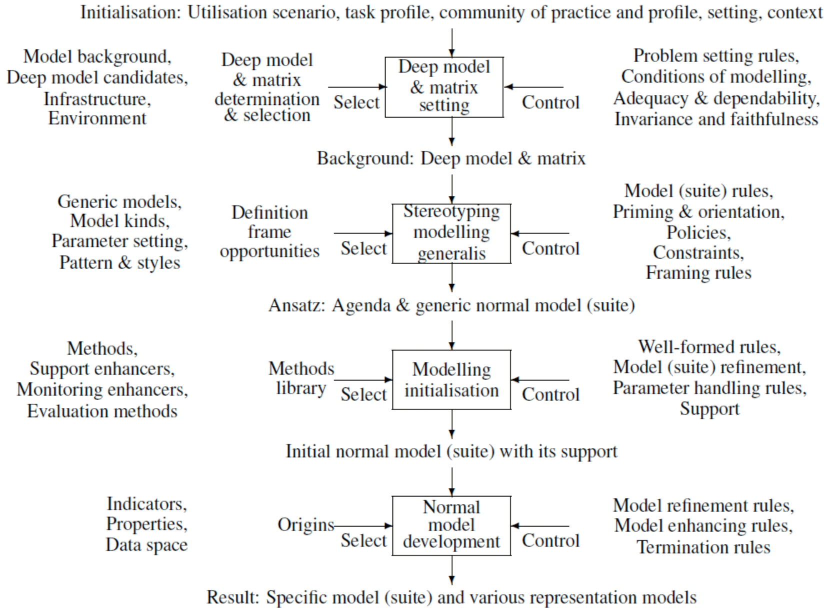

Initialisation: Utilisation scenario, task profile, community of practice and profile, setting, context
Model background, Deep model candidates, Infrastructure, Environment
Deep model & matrix determination & selection
Select
Deep model & matrix setting
Control
Problem setting rules, Conditions of modelling, Adequacy & dependability, Invariance and faithfulness
Background: Deep model & matrix
Generic models, Model kinds, Parameter setting, Pattern & styles
Definition frame opportunities
Select
Stereotyping modelling generalis
Control
Model (suite) rules, Priming & orientation, Policies, Constraints, Framing rules
Ansatz: Agenda & generic normal model (suite)
Methods, Support enhancers, Monitoring enhancers, Evaluation methods
Methods library
Select
Modelling initialisation
Control
Well-formed rules, Model (suite) refinement, Parameter handling rules, Support
Initial normal model (suite) with its support
Indicators, Properties, Data space
Origins
Select
Normal model development
Control
Model refinement rules, Model enhancing rules, Termination rules
Result: Specific model (suite) and various representation models

Figure 12: Layered model (suite) development (None-iterative form, greenfield variant)

The second layer or stage uses some kind of most general and refinable model as the initial model. A generic model [BST06, TF16] is a general model which can be used for the function within a given utilisation scenario and which is not optimally adapted to some specific origin collection. It is tailored in next steps to suit the particular purpose and function. It generally represents many origins under interest, provides means to establish adequacy and dependability of the model, and establishes focus and scope of the model. Modelling is often based on some experience. This experience can be systematically collected within a number of libraries. Libraries and collections are used for collecting the most appropriate setting and model. This selection is controlled or governed by rules, restrictions, conditions, and properties. The main results of the second layer are generic models and an agenda for the next modelling steps.

The third layer sets the environment for the development of the normal model. This environment prepares model development on the basis of the generic models and under inclusion of the deep model. The section of methods might also include the selection of parts and pieces from the context, e.g. from the background and especially from theories and knowledge. The fourth layer results then in the development of a normal model that can be neatly combined with the deep model. Representation models are developed for different members of the community of practice and for different functions the model must fulfill in the utilisation scenario.

This development process is often cut down to the fourth layer assuming the results of the first, second, and third layer as already given. This kind of implicitness has often been assumed for language utterance. The government and binding approach [Cho82, BST06] made the two-step generation of sentences explicit: we intentionally prepare the deep model and then express ourselves by an explicit statement which is build similar to a combination of a normal model and of a cutout of the deep model.

4. Conclusion

A collection of modelling approaches has been presented in [TN15]. It seems that the variety of modelling approaches, the different utilisation of model, the broad span of underpinning theories, the variety of models themselves do not allow to develop a common setting for models. We often met the claim that models used in social and natural sciences, in mathematics, in logics and in daily life are so different that a common treatment cannot exist. From the first side, logicians provided a specific understanding of models that is easy and formally to handle. They inspired model research and the notion of model, e.g. [Bal16, Kas03, Mah09, Mah15, Sta73, Ste66, Ste93]. This notion has mainly been based on properties that a model should satisfy: mapping, truncation, and pragmatic properties as phenomenalistic characterisation of the notion. From the second side, models in all sciences have been used as an artifact for solution of problems, e.g. [BT15, Her84, vDGOG09, vN55]. The model notion has been enhanced by amplification, distortion, idealisation, carrier, added value, and purpose-preservation properties. From the third side, language- and concept-based foundations of models have been developed in philosophy of science and linguistics [Bla¨15, Bur15, Cas55, KL13, Lat15, Pei98]. From the fourth side, models in engineering [BFA+ 16, LH15, TD16, TTF16] are instruments for system construction. From the sixth side, models are also in- struments in human interaction. They are used as metaphors, for communication, for brief reference, for depiction, as prototype, etc. For instance, the question whether a picture or a photo is a model depends on their utilisation in some interaction scenarios. We thus may conclude that a common science and culture of modelling cannot exist.

The main claim in this paper is however that a common treatment of models in science and human interaction can be developed. We base our foundational framework on a separation of concern. This separation into five governors for models provides a common treatment of models and model utilisation. We base our framework on the observation that not all concerns are considered at the same time. So, we can use some kind of stepwise procedure for model development.

Utilisation of models as instruments in scenarios is the main driving property that distinguishes something from a model. The model functions in scenarios such as communication, reflection, understanding, negotiation, expla- nation, exploration, learning, introspection, theory development, documentation, illustration, analysis, construction, description, and prescription.

How the model functions has been illustrated in the case of model-based reasoning. Model-based reasoning goes far beyond model methods used in classical first-order predicate logics or mathematics. We use the layering approach also for model methods since the development of a general reasoning method is far beyond the horizon.

The meta-models of modelling concerns in Figures 3, 4, 7, 8, 9, 10 support the layered modelling method in Figure 12. Instead, we could separate the layers into communities and their application scenario, into background and methodology setting, into situation and theory setting, into origin calibration, and model delivery layers.

This paper has been centered around models, theories, communities, context, methodologies, state, and dynamics at the same level of abstraction. Model-driven development and architecture [MMR+ 17, SV05] is an orthogonal approach to this paper. It distinguishes abstraction layers for models (M1), model frames (M2) [as meta-models], model frameworks (M3) [as meta-meta-models], and model framework setting (M4). The data/information and traces/events abstraction layer (M0) underpins models. Our approach has been mainly oriented on M1. We envision that the general M0-M1-M2-M3-M4 architecture can be integrated into our approach as well.

References

[ASG13] B.F. Albdaiwi, B. Szalkai, and V.I. Grolmusz. Finding combinatorial biomarkers for type 2 diabetes in the CAMD database. In *European Biophysics Journal with Biophyscs Letters*, volume 42, pp. S.195–S195. Springer, NY, USA, 2013.

[Bab03] B.E. Babich. From Fleck's Denkstil to Kuhn's paradigm: conceptual schemes and incommensurability. *International Studies in the Philosophy of Science*, 17(1):75–92, 2003.

[Bal82] W. Balzer. *Empirische Theorien: Modelle - Strukturen - Beispiele*. Vieweg-Teubner, 1982.

[Bal16] W. Balzer. *Die Wissenschaft und ihre Methoden*. Karl Alber, 2016.

[BBHK10] M.R. Berthold, C. Borgelt, F. Ho¨ ppner, and F. Klawonn. *Guide to intelligent data analysis*. Springer, London, 2010.

[BFA+ 16] M. Bichler, U. Frank, D. Avison, J. Malaurent, P. Fettke, D. Hovorka, J. Krämer, D. Schnurr, B. Müller, L. Suhl, and B. Thalheim. Theories in business and information systems engineering. *Business & Information Systems Engineering*, pp. 1–29, 2016.

[Blä15] C. Blättler. Das Modell als Medium. In *Wissenschaft und Kunst der Modellierung: Modelle, Modellieren, Modellierung*, pp. 107–137. De Gryuter, Boston, 2015.

[Bre10] J.E. Brenner. The logical process of model-based reasoning. In L. Magnani, W. Carnielli, and C. Pizzi, editors, *Model-based reasoning in science and technology*, pp. 333–358. Springer, Heidelberg, 2010.

[BST06] A. Bienemann, K.-D. Schewe, and B. Thalheim. Towards a theory of genericity based on government and binding. In *Proc. ER'06, LNCS 4215*, pp. 311–324. Springer, 2006.

[BT15] R. Berghammer and B. Thalheim. Methodenbasierte mathematische Modellierung mit Relationenalgebren. In *Wissenschaft und Kunst der Modellierung: Modelle, Modellieren, Modellierung*, pp. 67–106. De Gryuter, Boston, 2015.

[Bur15] T. Burkard. Der Blick des Philologen. Modelle 'Literatur als Text' in der Klassischen Philologie. In *Wissenschaft und Kunst der Modellierung: Modelle, Modellieren, Modellierung*, pp. 175–217. De Gryuter, Boston, 2015.

[Cas55] E. Cassirer. *The Philosophy of Symbolic Forms*, volume 1–3. Yale University Press, New Haven, 1955.

[CH04] S. Chadarevian and N. Hopwood, editors. *Models - The third dimension of science*. Stanford University Press, Stanford, California, 2004.

[Cho82] N. Chomsky. *Some concepts and consequences of the theory of government and binding*. MIT Press, 1982.

[DT10] A. Dahanayake and B. Thalheim. Co-evolution of (information) system models. In *EMMSAD 2010*, volume 50 of *LNBIP*, pp. 314–326. Springer, 2010.

[DT15] A. Dahanayake and B. Thalheim. W∗H: The conceptual model for services. In *Correct Software in Web Applications and Web Services*, Texts & Monographs in Symbolic Computation, pp. 145–176, Wien, 2015. Springer.

[Fle11] L. Fleck. *Denkstile und Tatsachen, edited by S. Werner and C. Zittel*. Surkamp, 2011.

[GKBF13] G. Greefrath, G. Kaiser, W. Blum, and R. Borromeo Ferri. Mathematisches Modellieren – Eine Einführung in theoretische und didaktische Hintergründe. In Mathematisches Modellieren für Schule und Hochschule, pp 11-37, Springer, 2013.

[Gra07] J. Gray. eScience: A transformed scientific method. Technical report, Talk given Jan 11, 2007. Edited by T. Hey, S. Tansley, and K. Tolle. http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB eScience.ppt, Microsoft Research Publications, 2007.

[Her84] H. Hertz. *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*. Akad. Verl.-Ges. Geest und Portig, Leipzig, 2. Aufl., Nachdr. der ausg. Edition, 1984.

[Jac04] R. Jackendoff. *Foundations of languages: Brain, meaning, grammar, evolution*. Oxford University Press, 2004.

[Jan17] K. Jannaschk. *Infrastruktur für ein Data Mining Design Framework*. PhD thesis, Christian-Albrechts University, Kiel, 2017.

[Kas03] R. Kaschek. *Konzeptionelle Modellierung*. PhD thesis, University Klagenfurt, 2003. Habilitationsschrift.

[KL13] B. Kralemann and C. Lattmann. Models as icons: modeling models in the semiotic framework of peirce's theory of signs. *Synthese*, 190(16):3397–3420, 2013.

[KT17] Y. Kropp and B. Thalheim. Data mining design and systematic modelling. In *Proc. DAMDID/RCDL'17*, pp. 349–356, Moscov, 2017. FRC CSC RAS.

[Lak87] G. Lakoff. *Women, fire, and dangerous things - What categories reveal about the mind*. The University of Chicago Press, Chicago, 1987.

[Lat15] C. Lattmann. Die Welt im Modell. Zur Geburt der systematischen Modellierung in der Antike. In *Wissenschaft und Kunst der Modellierung: Modelle, Modellieren, Modellierung*, pp. 307–327. De Gryuter, Boston, 2015.

[Leb14] A.V. Lebedev. *The Logos Heraclitus - A reconstruction of thoughts and words; full commented texts of fragments (in Russian)*. Nauka, 2014.

[LH15] J. Leibrich and P.A. Höher. Modelle in der Kommunikationstechnik. In *Wissenschaft und Kunst der Modellierung: Modelle, Modellieren, Modellierung*, pp. 329–345. De Gryuter, Boston, 2015.

[Mag14] L. Magnani, editor. *Model-Based Reasoning in Science and Technology*. Springer, 2014.

[Mah09] B. Mahr. Information science and the logic of models. *Software and System Modeling*, 8(3):365–383, 2009.

[Mah15] B. Mahr. Modelle und ihre Befragbarkeit - Grundlagen einer allgemeinen Modelltheorie. *Erwägen Wissen-Ethik (EWE)*, Vol. 26, Issue 3:329–342, 2015.

[Men89] W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory*, volume 45 of *International Geophysics*. Academic Press Inc., 1989.

[MMR+ 17] H.C. Mayr, J. Michael, S. Ranasinghe, V.A. Shekhotsov, and C. Steinberger. Model centered architecture. In *Conceptual Modeling Perspectives*, pp. 85–104, Cham, 2017. Springer.

[Mu¨ l16] R. Müller. Model history is culture history. From early man to cyberspace. http://www.muellerscience.com/ENGLISH/model.htm, 2016. Assessed Oct. 29,2017.

[MV10] B. Mahr and W. Velminski. Denken in Modellen. Zur Lösung des Königsberger Brückenproblem. In *Mathesis & Graphe´ , Leonhard Euler und die Entfaltung der Wissenssysteme*, pp. 85–100. Akademie-Verlag, Berlin, 2010.

[Noa09] K. Noack. Technologische und methodische Grundlagen von SCOPELAND. White paper, www.scopeland.de, 2009.

[Pei98] C.S. Peirce. What is a sign? In Peirce Edition Project, *The essential Peirce: selected philosophical writings*, vol. 2, pp. 4 – 10. Indiana University Press, Bloomington, Indiana, 1998.

[Pod01] A.S. Podkolsin. *Computer-based modelling of solution processes for mathematical tasks (in Russian)*. ZPI at Mech-Mat MGU, Moscov, 2001.

[Pol45] G. Polya. *How to solve it: A new aspect of mathematical method*. Princeton University Press, Princeton, 1945.

[Rei84] H. Reichel. *Structural induction on partial algebras*. Mathematical research, 18. Akademie-Verlag, Berlin, 1984.

[RSS+ 10] J. Rückelt, V. Sauerland, T. Slawig, A. Srivastav, B. Ward, and C. Patvardhan. Parameter optimization and uncertainty analysis in a model of oceanic co2-uptake using a hybrid algorithm and algorithmic differentiation. *Nonlinear Analysis B Real World Applications*, 11(5):3993–4009, 2010.

[ST08] K.-D. Schewe and B. Thalheim. Semantics in data and knowledge bases. In *SDKB 2008*, LNCS 4925, pp. 1–25, Berlin, 2008. Springer.

[Sta73] H. Stachowiak. *Allgemeine Modelltheorie*. Springer, 1973.

[Ste66] W. Stegmüller. Eine modelltheoretische Präzisierung der Wittgensteinschen Bildtheorie. *Notre Dame Journal of Formal Logic*, 7(2):181–195, 1966.

[Ste93] W. SSteinmüller *Informationstechnologie und Gesellschaft: Einfu¨hrung in die Angewandte Informatik*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1993.

[SV05] T. Stahl and M. Völter. *Model-driven software architectures*. dPunkt, Heidelberg, 2005. (in German).

[Tar56] A. Tarski. *Logic, semantics, metamathematics*. Oxford at the Claredon Press, 1956.

[TD16] B. Thalheim and A. Dahanayake. Comprehending a service by informative models. *T. Large-Scale Data- and Knowledge-Centered Systems*, 30:87–108, 2016.

[TF16] M. Tropmann-Frick. *Genericity in Process-Aware Information Systems*. PhD thesis, Christian-Albrechts University of Kiel, Technical Faculty, Kiel, 2016.

[Tha10] B. Thalheim. Model suites for multi-layered database modelling. In *Information Modelling and Knowledge Bases XXI*, volume 206 of *Frontiers in Artificial Intelligence and Applications*, pp. 116–134. IOS Press, 2010.

[Tha14] B. Thalheim. The conceptual model ≡ an adequate and dependable artifact enhanced by concepts. In *Information Modelling and Knowledge Bases*, volume XXV of *Frontiers in Artificial Intelligence and Applications, 260*, pp. 241–254. IOS Press, 2014.

[Tha17a] B. Thalheim. Conceptual modeling foundations: The notion of a model in conceptual modeling. In *Encyclopedia of Database Systems*. Springer US, 2017.

[Tha17b] B. Thalheim. General and specific model notions. In *Proc. ADBIS'17*, LNCS 10509, pp. 13–27, Cham, 2017. Springer.

[Tha17c] B. Thalheim. Model-based engineering for database system development. In *Conceptual Modeling Perspectives*, pp. 137–153, Cham, 2017. Springer.

[Tha18a] B. Thalheim. Conceptual model notions - a matter of controversy; conceptual modelling and its lacunas. *EMISA International Journal on Conceptual Modeling*, February: 9–27, 2018.

[Tha18b] B. Thalheim. Normal models and their modelling matrix. In *Models: Concepts, Theory, Logic, Reasoning, and Semantics*, Tributes, pp. 44–72. College Publications, 2018.

[TN15] B. Thalheim and I. Nissen, editors. *Wissenschaft und Kunst der Modellierung: Modelle, Modellieren, Modellierung*. De Gruyter, Boston, 2015.

[TT13] M. Tropmann and B. Thalheim. Mini story composition for generic workflows in support of disaster management. In *DEXA 2013*, pp. 36–40. IEEE Computer Society, 2013.

[TTF16] B. Thalheim and M. Tropmann-Frick. Models and their capability. In C. Beierle, G. Brewka, and M. Thimm, editors, *Computational Models of Rationality*, volume 29 of *College Publications Series*, pp. 34–56. College Publications, 2016.

[TTFZ14] B. Thalheim, M. Tropmann-Frick, and T. Ziebermayr. Application of generic workflows for disaster management. In *Information Modelling and Knowledge Bases*, volume XXV of *Frontiers in Artificial Intelligence and Applications, 260*, pp. 64–81. IOS Press, 2014.

[vDGOG09] C. von Dresky, I. Gasser, C.P. Ortlieb, and. *Mathematische Modellierung: Eine Einführung in zwölf Fallstudien*. Vieweg, 2009.

[vN55] J. von Neumann. Method in the Physical Sciences. In *The Unity of Knowledge*, pp. 157–164. Doubleday & Co., New York, 1955.

[Wec92] W. Wechler. *Universal Algebra for Computer Scientists*. Springer-Verlag, Berlin, 1992.

[Wik17] Wikiquote. Conceptual model. https://en.wikiquote.org/wiki/Conceptual model, 2017. Assessed Nov. 21, 2017.

Bernhard Thalheim
Christian-Albrechts University at Kiel, Department of Computer Science,
D-24098 Kiel
bernhard.thalheim@email.uni-kiel.de

STUDIA METODOLOGICZNE
NR 39 • 2019, 49–71
DOI: 10.14746/sm.2019.39.2

Axel Gelfert

Cultures of Modelling: Rudolf Peierls on 'Model-Making in Physics'

Abstract. The philosophical debate about scientific models has, over the past thirty years or so, reached a high degree of sophistication. Yet, in spite of efforts to seek common ground with scientific practice, there remains the suspicion that philosophical accounts are sometimes too 'free-floating', in that they do not adequately reflect scientists' views (and actual uses) of models. The present paper deals with one such scientific perspective, due to physicist Sir Rudolf Peierls (1907-1995). Writing thoroughly from the perspective of a theoretician with a deep appreciation for experimental physics, Peierls, in a series of papers, developed a taxonomy of scientific models, which – in spite of some inevitable arbitrariness – exhibits surprising points of convergence with contemporary philosophical accounts of how scientific models function. The present paper situates Peierls's approach within the philosophical and scientific developments of his time, engages (in an immersive way) with his proposed taxonomy, and argues that Peierls's views – and others like them – warrant the recent philosophical shift from a focus on model-based representation to non-representational (e.g., exploratory) uses and functions of models.

Keywords: scientific models, modeling, scientific representation, exploration.

1. Introduction

Scientific models and the activity of modelling in science have, in recent years, attracted considerable attention from philosophers of science. Sophisticated philosophical accounts have been proposed regarding how models represent their targets and allow us to infer knowledge about them, and a plethora of case studies from the various special sciences have been worked out, many of which engage with the cutting edge of contemporary science. What has sometimes been neglected, however, is the perspective of scientists themselves. To be sure, there are laudable exceptions, notably Daniela Bailer-Jones's analysis of scientists' thoughts on scientific models [Bailer-Jones, 2003]. Yet how the widespread adoption of scientific modelling across the sciences 'adds up', so as to shape the production of new knowledge, remains a philosophically neglected question. At the risk of disappointing my readers, I must be up-front and admit that the current paper will not fill this lacuna. Any attempt to do so would require a breadth of coverage and a level of detailed analysis that would be impossible within the constraints of a single paper. What I will attempt, instead, is to show, by way of example, how scientific models have become the preferred way for scientists to deal with, and reflect on, a range of worthy goals: representing reality, reducing complexity, getting a grasp on novel and elusive phenomena, deriving potential explanations, exploring constraints and theoretical structures, implementing approximations, and studying limiting cases. Specifically, I will be engaging with the work of Sir Rudolf Peierls (1907-1995), who was at the centre of many important twentieth-century developments in physics, without ever achieving the pop-science stardom of some of his contemporaries (notably Richard Feynman). Two papers, written in the 1980s after his retirement, explicitly discuss different model types and their functions; while these lack philosophical rigour, they provide an interesting glimpse into the epistemic culture of modelling in physics, as experienced by one of its prime exponents.

The rest of this paper is organized as follows. Section 2 discusses some of the history of scientific models, focussing on the shift from emphasizing mechanical models to a more inclusive notion of 'model' that accommodates, among others, analogical reasoning in physics. It also dicusses Mary Hesse's influential mid-twentieth century text 'Models and Analogies in Science' (1963), which marks the beginning of a rapid growth of philosophical interest in models and their role in science. Section 3 surveys some of the theoretical tensions that afflict any philosophical attempts to come to a global characterization of what models are and how they function in inquiry. Section 4 summarizes in some detail the seven-fold taxonomy of 'model types' proposed by Rudolf Peierls in a semi-popular article published in 1980 as 'Model-Making in Physics' (and reprised in his 1987 'Models, Hypotheses and Approximations'). Peierls's perspective is that of a practicing scientist who suspends his immediate research agenda and reflects on the broader direction of physics; he is not primarily concerned with the finer philosophical points concerning models and their functions. Taking his remarks at face value, then, requires immersing oneself in the – often 'hands-on' and outcome-oriented – epistemic culture of scientists using models for a variety of purposes. Such an immersive approach is rewarded, however, by insights into the practice of model-making and its guiding values, such as its recognition of pluralism – first and foremost, the realization 'that different models serve quite different purposes, and they vary in their nature accordingly' [Peierls, 1980, p. 3]. The fifth and final section relates the material presented thus far to recent attempts to shift philosophical attention from model-based representation to non-representational uses and functions of models. Some of these attempts have coalesced under the label of 'exploratory modelling' [Gelfert, 2016, pp. 71-100]; on this view, models – beyond representing actual targets – can probe modal and theoretical structure (e.g., by considering various counterfactual, e.g. higher- or lower-dimensional scenarios), generate potential explanations, or provide 'proofs-of-principle'. Furthermore it is argued that models often play a regulative role in scientific inquiry more generally, by giving direction to prospective research programmes and setting them on course for future successes.

2. Historical background: from analogies to models

When looking at the history of models in science, one may feel tempted to survey the history of science from the vantage point of our current understanding of the term 'scientific model' (which itself is far from uniform) and look for episodes that appear to fit with one's preferred definition of what constitutes a scientific model. Yet such an approach would hardly do justice to the varied history of the term 'model' in scientific discourse – a task which is also beyond the scope of this paper. It will nonetheless be instructive to look at the (surprisingly recent) emergence of 'model talk' in science, and in physics in particular.

It seems safe to say that systematic self-reflection on the uses and limitations of models in physics did not begin in earnest until some time in the nineteenth century. While methodological reflection and sophisticated analyses of the status of hypotheses, theories, and observations can be found throughout the history of science, including in its early stages, these did not coalesce into a systematic discussion of the role and significance of models in scientific inquiry. In philosophy of science, the recognition that central models are central to the pursuit of science did not set in until even more recently. Only from the middle of the twentieth century onwards did philosophers of science shift their focus from theories to models – which, until then, had often been regarded as playing a merely auxiliary role in applying fundamental theories to specific situations. One important transformation that contributed to the remarkable rise of scientific models in physics from the nineteenth century onwards, and to their proliferation in the twentieth century and beyond, was the shifting of emphasis from mechanical models (i.e., real or imagined mechanical 'stand-ins' for real target systems) to a far more inclusive notion of 'model' (as reflected in twentieth-century expressions such as the 'standard model' in particle physics).

Consider Pierre Duhem's endorsement of the use of *analogy* in physics. The idea of analogy derives its utility from the thought that relations in one domain resemble those in what may be an otherwise entirely separate domain, such that *A* is related to *B* (where *A* and *B* belong to one domain) like *C* is related to *D* (where these belong to the other domain). Whether such resemblance is merely formal or is underwritten by material similarity is of secondary importance for our purposes; at any rate, Duhem conceives of analogy primarily as a relation between sets of statements, more specifically between one theory and another:

> Analogies consist in bringing together two abstract systems; either one of them already known serves to help us guess the form of the other not yet known, or both being formulated, they clarify the other. There is nothing here that can astonish the most rigorous logician, but there is nothing either that recalls the procedures dear to ample but shallow minds. [Duhem, 1954, p. 97]

A good example is Christiaan Huygens's proposal, in 1678, of his wave theory of light. In developing his theory, Huygens was guided by an analogy with the theory of sound waves: the relations between the various properties and qualities of light are *like* those of sound waves, as described by acoustic theory. Understood in this way, analogy is, for Duhem, an entirely legitimate tool for studying one domain on the basis of our (more secure) knowledge of quite another domain. Sound waves, in our contemporary scientific vernacular, provided Huygens with a good *theoretical model* for how light propagates and behaves in various settings.

This contrasts with Duhem's forceful rejection of *mechanical models* as a way of expounding a new theory (rather than, say, merely illustrating it). Taking a textbook presentation of Maxwell's theory of electromagnetism as his target of choice, Duhem strikes a polemical tone in his dismissal of what he takes to be an undue reification of theoretical relationships into mechanical processes:

> Here is a book intended to expound the modern theories of electricity and to expound a new theory. In it there are nothing but strings which move round pulleys which roll around drums, which go through pearl beads, which carry weights; and tubes which pump water while others swell and contract; toothed wheels which are geared to one another and engage hooks. We thought we were entering the tranquil and neatly ordered abode of reason, but we find ourselves in a factory. [Duhem, 1954, p. 7]

It is important to be clear about precisely what Duhem is criticizing. Given his earlier endorsement of analogical reasoning, Duhem cannot be categorically opposed to the idea of relating one domain to another (qualitatively different) one. Instead, what the quoted passage mocks is a *style* of reasoning – one in which the desire to visualize physical processes in purely mechanical terms masks the theoretically more ambitious task of understanding them in their own right. His hostility is thus directed at mechanical models only—as is also clear from the contrast implicit in the title of the chapter ('Abstract Theories and Mechanical Models') from which the quoted passage is taken—and extends neither to 'theoretical models' nor, necessarily, to other contemporary uses of the term 'scientific model' in physics. Whereas mechanical models encourage the hasty identification of the entities being visualized ('pulleys', 'drums', 'pearl beads', 'toothed wheels', etc.) with the (unknown) actual physical processes, analogy makes it possible for us to make inferences about one domain on the basis of knowledge about another, while at the same time acknowledging their qualitative difference.

A first sketch of a philosophical account of how analogies can underwrite the use of models in science was presented by Mary Hesse in her influential 1963 essay *Models and Analogies in Science*, which is explicitly conceived of as a dialogue between a 'Duhemist' and his opponent, the 'Campbellian' (after the English physicist Norman Robert Campbell, 1880-1949). The dialogue begins with the Campbellian attributing to the Duhemist the following view: 'I imagine that along with most contemporary philosophers of science, you would wish to say that the use of models or analogues is not essential to scientific theorizing and that […] the theory as a whole does not require to be interpreted by means of any model.' The Duhemist, after conceding that 'models may be useful guides in suggesting theories', replies as follows: 'When we have found an acceptable theory, any model that may have led us to it can be thrown away.' The Campbellian, by contrast, insists: 'I, on the other hand, want to argue that models in some sense *are* essential to the logic of scientific theories.' [Hesse, 1963, pp. 8-9] What is at stake in this dispute, then, is both the ontological question of what, essentially, models are in the first place and, importantly, also the extent to which they are admissible in inquiry.

Hesse's own analogical account of scientific models begins by drawing a three-fold distinction between 'positive', 'negative', and 'neutral' analogies. Consider the billiard ball model of gases, which portrays gases as being composed of tiny elastic 'billiard balls' that fill up a given volume and sometimes bounce off each other, in a way that is meant to account for properties such as pressure and temperature. Some characteristics are shared between the (imagined) billiard balls and the target system consisting of gas atoms: for example, the momentum that can be ascribed to individual constituents and the phenomenon of collision between them. This set of shared characteristics constitutes the *positive* analogy, whereas properties we know to belong to billiard balls, but not to gas atoms—such as colour—constitute the *negative* analogy of the model. Yet the positive and negative analogy together do not exhaust the set of all properties, as there will typically be properties of the model for which it is as yet unclear whether they (also) apply to its target system. These constitute what Hesse calls the *neutral* analogy of the model.

It is the neutral analogy that injects an exploratory element into the process of inquiry, since it offers the prospect of gaining new insights into the target system by studying the model in its place—a prospect that might ordinarily have seemed slim: 'If gases are really like collections of billiard balls, except in regard to the known negative analogy, then from our knowledge of the mechanics of billiard balls we may be able to make new predictions about the expected behaviour of gases.' [Hesse 1963, p. 10]

The 1950s and 1960s—that is, exactly the period during which Hesse and other philosophers of science began to consider models in their own right—were a period of rapid growth in physics, much of which was driven by the development of ever more ambitious and sophisticated models in physics. In particle physics, the 'standard model' was beginning to be conceived—even if, arguably, it served more as a framework for theorizing than as a representation of any one target system in particular; quantum theory had gained sufficient maturity to also enter more applied subdisciplines—leading, amongst other developments, to the formation of solid-state physics as a separate subdiscipline and to the development of a myriad of quantum many-body models.[1] The other sciences, too, witnessed a growing reliance on models, driven, not least, by mathematical models in disciplines ranging from biology to economics. This helped prepare the ground for further scientific and conceptual explorations concerning models and their role in inquiry—a development that continues unabated to this day.

3. The ontology of models and the practice of modelling

The proliferation of models across the sciences makes it difficult to give a comprehensive and uncontroversial answer to the ontological question of what, in general, a scientific model is. Disagreement on general terms such as 'knowledge', 'theory', or 'model' is, of course, part and parcel of the philosophical enterprise, yet there is an unavoidable trade-off between achieving

[1] For the emergence of solid-state physics as a recognized stand-alone subdiscipline, see especially [Weart, 1992]; for a philosophical survey of the variety of quantum many-body models in condensed matter physics, see [Gelfert 2015].

conceptual clarity through stipulation and retaining adequate scope in relation to the issues that initially motivated our philosophical inquiry. In the present case, what generated philosophical interest in scientific models in the first place were the perceived growth of 'model talk' among scientists and the growing presence of models in scientific practice. In their attempts to make sense of what scientists call 'models', and of how they use them, philosophers of science created their own sprawling taxonomies, as is evident from this list of model-types, found in the *Stanford Encyclopedia of Philosophy*: 'Probing models, phenomenological models, computational models, developmental models, explanatory models, impoverished models, testing models, idealized models, theoretical models, scale models, heuristic models, caricature models, didactic models, fantasy models, toy models, imaginary models, mathematical models, substitute models, iconic models, formal models, analogue models and instrumental models' [Frigg & Hartmann, 2012]. In light of this dazzling diversity, it is perhaps no surprise that Nelson Goodman, as early as in his 1968 *Languages of Art*, voices the following lament: 'Few terms are used in popular and scientific discourse more promiscuously than "model"'. [Goodman, 1968, p. 171] If this was true of science and popular discourse in the late 1960s, it is no less true today.

The great variety of models employed in scientific practice makes vivid just how central the use of models is to contemporary science and, perhaps increasingly, to the self-image of scientists how rely on them. As John von Neumann once put it: 'The sciences do not try to explain, they hardly even try to interpret, they mainly make models.' [von Neumann, 1961, p. 492] It might, however, also lead one to ask whether it is at all reasonable to look for a unified philosophical account of models. Given the vast range of things we call 'models', and the divergent uses to which they can be put, a one-size-fits-all answer to the question 'What is a model?' may simply seem out of reach. One reaction has been to try to assimilate models to theories, thereby treating them as entirely auxiliary and subordinate. On this view, models may be, as Richard Braithwaite put it, 'the most convenient way of thinking about the structure of the theory' [Braithwaite, 1968, p. 91], but they are just that: ways of thinking about an underlying theory. Even more sternly, Rudolf Carnap urged his readers 'to realize that the discovery of a model has no more than an aesthetic or didactic or at best a heuristic value, but is not at all essential for a successful application of the physical theory' [Carnap, 1969, p. 210].

Another standard reaction to the puzzling diversity of what constitutes a scientific model has been to argue that, as Gabriele Contessa puts it, 'if all scientific models have something in common, this is not their *nature* but their *function*' [Contessa, 2010, p. 194]. Amongst functional characterizations of models, a further distinction can be drawn between instantial and representational views. According to the former, models have their function in virtue of instantiating the axioms of a theory, where the latter is understood in terms of linguistic statements. By contrast, on the representational *view*, 'language connects not directly with the world, but rather with a model, whose characteristics may be precisely defined'; the model makes contact with the world only inasmuch as there is a 'similarity between a model and designated parts of the world' [Giere, 1999, p. 56]. Generally speaking, proponents of the instantial view regard models as primarily being in the business of 'providing a means for interpreting formal systems', whereas those who favour the representational view consider models to be 'tools for *representing the world*' [Giere, 1999, p. 44]. The representational view, in turn, can be construed by either highlighting the informational aspects of models or their pragmatic role in inquiry. The basic idea of the former is 'that a scientific representation is something that bears an objective relation to the thing it represents, on the basis of which it contains information regarding that aspect of the world' [Chakravartty, 2010, p. 198]; by contrast, the *pragmatic* view of model-based representation holds that models represent their targets in virtue of the cognitive uses to which human reasoners put them.

The turn to pragmatic (or 'practice-oriented') aspects of scientific models has been a fairly recent development. It acknowledges that models are the outcome of a process of model construction which is itself responsive to the context of inquiry. On this view, the question 'What is a model?' simply cannot be answered satisfactorily without a proper consideration of the activity of *modelling*, which, according to pragmatic theorists of models, is characterized by 'piecemeal borrowing' [Suárez & Cartwright, 2008, p. 63] from a range of representational resources. Thinking of models as standing in purely objective (e.g., informational) relations to one another, and to their target systems, would overlook the ineliminable role of beliefs, intentions, and cognitive interests of on the part of model users, as well as of the material constraints that come with the heterogeneous components that, typically, make up any real-world model of a phenomenon or target system. Shifting attention away from models and the abstract relations they stand in, towards modelling as a complex activity pursued by human agents, also involves – as Tarja Knuuttila puts it – a shift away from 'the model-target dyad as a basic unit of analysis' [Knuuttila, 2010, p. 142] towards a 'triadic' picture that acknowledges the equal importance of model, target, and user.

4. Rudolf Peierls and the culture of 'model-making'

Given the relative recency of the aforementioned turn towards scientific practice in philosophical accounts of models, it is perhaps prescient that Rudolf Peierls, in 1980, published a paper titled 'Model-Making in Physics', in which he speaks of a 'model-making habit' and attributes to physicists a tendency 'to use models of various kinds to aid their understanding of complicated physical situations' [Peierls, 1980, p. 3]. At the time the paper was published, Peierls was in his early seventies and had already spent six years in retirement; his breezy presentation of various models in physics is, therefore, less of a summary of state-of-the-art scientific modelling than a reflection on the proliferation of models in physics since the middle of the twentieth century.

Peierls's own career is closely linked to many of the main developments in twentieth-century physics. Born in 1907 in Berlin, Peierls studied at the universities of Berlin, Munich, and Leipzig, with stints in Switzerland and the Soviet Union, before escaping the deteriorating political situation in Germany by emigrating to Britain in 1933, where he eventually became a citizen in 1940. This allowed him to take up war work and led to his joining the Manhattan Project, though he remained critical of the pursuit of nuclear weapons (and later campaigned against their proliferation). His scientific work was unusually diverse, with Peierls being described, by the mathematician Herbert S. Green, as 'a highly competent, though not a notably creative mathematician; his principal interest was clearly in making calculations which would lead to a deep understanding of physical phenomena, and he was adept at finding approximations which gave trustworthy numerical results' [Green, 1999]. Unlike some of his more famous contemporaries, Peierls did not seek the public limelight, and his choice of research topics, too, reflects a decidedly 'middle-of-the-road' preference for soundness and applicability; this, I would argue, makes him a better representative than most for the bulk of research activity that constitutes post-war twentieth-century physics.

In 'Model-Making in Physics', Peierls adopts very much the perspective of a practicing scientists who suspends his immediate research agenda and instead reflects on the broader shape of his discipline. He does not explicitly set out to seek common ground between, say, physics and philosophy of science; neither does he promise a unified theory of how scientific models work. While he makes it his goal 'to examine the nature and purpose of […] in some detail', he immediately acknowledges 'that different models serve quite different purposes, and they vary in their nature accordingly' [Peierls, 1980, p. 3]. This in itself is noteworthy since it shifts, almost effortlessly, the emphasis from the ontological question 'What is a model?' to the more pragmatic question of how models achieve their varied functions. Before summarizing Peierls's answer, however, it is worth emphasizing that he never intended his paper to be an up-to-date contribution to the philosophical debate. As a result, he makes no effort to engage with whatever philosophical debate of scientific models had developed by the time the paper was published – of which, by 1980, there had been a considerable amount. All the references are entirely to other papers in physics, and most of them refer to case studies Peierls is using for illustrative purposes, not to discussions of how models are being used in physics in general. While one might lament the lack of engagement with the extant philosophical literature, along with a number of conceptual infelicities on Peierls's part that a closer engagement with philosophy of science might have prevented, his text nonetheless deserves to be taken seriously. This is why, in what follows, I have decided to take Peierls's text at face value, treating it as an expression of a certain epistemic culture of model-based physics, and immersing myself in it, rather than taking him to task for, say, eliding various philosophical distinctions.

Peierls's taxonomy of models lacks hierarchy and systematicity, and he readily acknowledges that his choice 'of the categories, and the assignment of specific models to them, is of course very subjective'[2] (3), and that there is bound to be disagreement about individual assignments, yet not about 'the width of the spectrum' of cases considered. Peierls's choice of the first of the seven 'types' of models he distinguishes – '*Hypothesis ("Could be true")*' – already makes clear that he has no truck with extant philosophical distinctions, given that hypotheses are not usually lumped together with models. Yet, speaking as a scientific practitioner, Peierls notes that 'hypotheses are often called models', and any taxonomy of uses of models (as well as uses of the term 'model') had better comment on its relation to hypotheses. And, to be sure, models are often invoked by hypotheses that 'consist of a tentative explanation of a phenomenon'. Examples would be early models of the atom, such as the textbook models put forward by J.J. Thomson and Ernest Rutherford, which 'amount really to statements about the nature of the Universe which may or may not be correct' (4). Peierls's second type – '*Phenomenological model ("Behave as if…")*' – is more in line with established taxonomies in philosophy of science, though he is vague on the question of precisely what it takes for a model to count as 'behaving as if' it were the real target system. According to Peierls, in a phenomenological model 'a physical phenomenon [is] accounted for by a certain mechanism, but there is insufficient evidence to convince us that this is the correct explanation' (5). For Peierls, whether or not a model counts as phenomenological, appears to be less a matter of "saving the phenomena" (while remaining agnostic about what underlying processes might have brought about an observed phenomenon), than a matter of uncertainty about whether the proposed underlying mechanisms are, in fact, realized or not. Phenomenological models, then, are regarded as characteristic of research in its early stages. As a 'very characteristic example' (6), he discusses Pierre Weiss's model of ferromagnetism, which posits that elementary magnets contribute to the magnetization of a substance not only via their response to an external magnetic field, but also via 'a "molecular field" proportional to the number of magnets already aligned'. This simple model helped explain the existence of a transition temperature, the Curie temperature, 'at which the

[2] For the remainder of this section, numbers in round parentheses refer to the corresponding page number in [Peierls, 1980].

spontaneous magnetization goes continuously, but very steeply, to zero'. While later work revealed the model to be inadequate in a number of ways, it is 'still useful if we want a quick orientation on the likely behaviour of a ferromagnet in unfamiliar circumstances'. Phenomenological models, beyond merely reproducing observed phenomena and (sometimes) identifying potential causal mechanisms, also importantly play a way in *regulating and guiding inquiry* – a topic I shall return to in the final section.

The next two types of models – '*Type 3: Approximation (Something is very small, or very large)*' and '*Type 4: Simplification (Omit some features for clarity)'* – are again labelled in a somewhat misleading manner, in that they really refer to different (though potentially co-existing) methodological approaches. When Peierls refers to them as 'models', this is best understood elliptically as referring to models generated by the (predominant) use of one or the other. 'Approximation' – which, judging from the parenthetical characterization as 'Something is very small, or very large', includes the consideration of limiting cases – is required whenever no closed solutions exist to the model equations and is considered, by Peierls, to be an 'art' that is 'much more subtle than that of solving an equation exactly' (7). The example discussed by Peierls is that of linear response models, which describe how a target system responds to an infinitesimal disturbance. Once again he notes that, even where (by some measure) 'better' models exist, linear approximations have their legitimate uses: e.g., 'the stability of a system depends on the sign of its linear response coefficients for various possible disturbances' (8). At the same time, one must take great care to also respect the inevitable limits on when a system's response can be modelled as linear. For example, when calculating the electronic shell structure of the atom, higher-order calculations are important in order to lift degeneracy in the energy levels – contenting oneself with the linear approximation would miss out on key aspects of the physics of the atom. When to deploy the right sorts of approximations, Peierls argues, is a matter of 'judgement and experience'; for the purpose of model-making and using models successfully, we cannot abstract away from the context of inquiry.

'*Simplification*', or the omission, for clarity, of some (known) features of the target system ('*Type 4*' [9]), is likewise an important approach in the construction of models. Peierls's preferred example here is Peter Debye's model for the specific heat of solids. Debye essentially proposed a formula that interpolates between the (known) behaviour of solids at near-zero temperatures (when only a diminishing number of low-frequency vibrations of the solid are possible) and the maximum frequency of lattice vibrations (which, in reality, is determined by the lattice structure of the crystal, but which Debye chose 'so as to get the total number of modes right' [10], i.e. the maximum number of possible excitations for the total number of atoms in a crystal). This resulted in a model that is surprisingly useful even at intermediate temperatures. Yet the success of any given model also creates pitfalls, in particular for those who lose track of the model's limitations. One such case for the Debye model is the thermal behaviour of beryllium. For beryllium, the difference between the 'predicted' Debye curve and actual measurements – that is, the curve that results from subtracting one from the other – had a hump-like structure, which some researchers – incorrectly – interpreted 'as suggesting a transformation in this substance'. Yet this interpretation overlooked entirely that the Debye curve merely interpolates between two known constraints; its numerical values in-between make no claim to realism.

The fifth type, '*Instructive model (No quantitative justification, but gives insight*'), according to Peierls, achieves 'even greater simplification' at the cost of moving 'even further away from a realistic description', while 'still retaining enough similarity with the true situation to help understanding something about its nature' (13). This characterization, which Peierls acknowledges is 'less sharp than previous dividing lines', is admittedly vague: How much similarity with the true situation is 'enough', and how specific to the target system must our 'understanding something about its nature' really be? Yet, often what is needed at a certain point in inquiry is not numerical accuracy, or even qualitative similarity, but an estimation of, say, the expected order of magnitude of a process or phenomenon. This, Peierls argues, is just what such models are good at providing, for example in such cases as the mean free path model, transport phenomena, and conductivity: 'For general orientation [such a] model is so useful that it is being used all the time even by those who are familiar with is weaknesses and pitfalls', yet it stands to reason that there is, in turn, a very real risk of the model being used unthinkingly, by those who are less able to spot potential pitfalls.

Peierls's sixth type of model, '*Type 6: Analogy (Only some features in common)*', echoes Mary Hesse's account of scientific models as analogies (discussed in a previous section of this paper), though it is unlikely that Peierls would have closely engaged with Hesse's work. Peierls posits that, in many situations, we 'learn something about a physical system from the study of a simpler system which does not resemble it in all essentials, but has some of its typical features' (14). Interestingly, he quickly moves to the discussion of scientific examples, most of which – Debye's model for phonon scattering, the Ising model, the London model of superconductor – could have easily been fitted under one of the other categories. Had Peierls aimed for a more systematically ambitious taxonomy, he might have chosen analogy as an overarching framework that subsumes the various types of models – just as Hesse did – rather than as a separate category, or 'type', in its own right. Nonetheless, what is significant is Peierls's recognition that even models that are known to be fundamentally flawed – such as Ising's model of ferromagnetism, which ignores many aspects of quantum many-body systems that are known to be of the utmost importance in determining the collective behaviour (including phase transitions) in solids – can still function as a source of insight: 'Nevertheless much can be learnt from the model.' (15) Importantly, 'variants of the Ising model have served as a proving ground for methods and approximations in this field' (15). Once again, Peierls ranks pragmatic utility, and the way in which models – in spite of their flaws – can keep scientific research programmes progressive higher than their representational accuracy in absolute terms, as it were.

The final type of model, '*Type 7: Gedanken experiments (Mainly to disprove a possibility)*' covers thought experiments such as the Carnot cycle, Maxwell's demon, Heisenberg's gamma-ray microscope, and the Einstein-Podolsky-Rosen paradox. While again a non-standard usage of the term 'model', Peierls's decision to include thought experiments among the various types of models is nonetheless instructive. He notes that, by considering thought experiments, it is often possible to derive constraints on what is, and isn't, possible. Thus, in thermodynamics, considering the Carnot cycle 'can, for example, place limits on the efficiency obtainable from an engine working in a given temperature range' (16); the idea, then, is not that the

Carnot cycle could, or should, be brought about in the real world, but rather that any real-world process must respect certain constraints that are brought into sharp focus – albeit counterfactually – by the corresponding thought experiment. Sometimes, thought experiments do not establish (im)possibilities, but – like models used for instruction – serve illustrative purposes. This, Peierls argues, is the case with Heisenberg's γ-ray microscope, which imagines attempting to see an electron in a microscope: in order to achieve a sufficiently good resolution, one would need to use radiation of sufficiently short wavelengths – i.e. γ-rays – yet any encounter of an electron with such radiation would cause an uncontrollable change in its momentum, leading to an unavoidable trade-off between locating the electron and measuring its momentum – Heisenberg's famous uncertainty relation. Heisenberg's thought experiment 'was not used to prove anything, because the uncertainty relation could be deduced directly from the formalism of quantum mechanics' (16), but it helped physicists 'understand the nature of the new principle without recourse to the mathematical formalism' (17).

Peierls's seven model 'types' span different levels of inquiry, ranging from general methodological approaches – approximation and simplification – to specific contexts (e.g., instruction), uses (e.g., in order to 'save the phenomena'), and formats of representation (e.g., 'statements about the nature of the Universe which may or may not be correct'). One may deem his taxonomy haphazard and unsystematic – yet, arguably, no more so than the messy and disunified character of scientific practice. Indeed, it could be argued that the fact that Peierls foregoes any strongly normative stance – except for the recurring injunction to remain aware of the limitations of one's models – simply reflects the epistemic culture associated with 'model-making' in physics. Some models – because of their wide applicability, or because they feature prominently in shared curricula of physics education – constitute common ground, to which even proponents of competing research programmes can jointly retreat. Others are themselves hotly contested. And even where there is in-principle agreement on the utility of a particular approach, trade-offs between different desiderata – simplicity, numerical accuracy, generality, etc. – are the norm rather than the exception. As noted above in relation to the Debye model, some researchers place considerable (even excessive) faith in the realism of their models; others adopt a thorough-going instrumentalist attitude.[3] Yet, as in the case of the messy world of experimental physics, '[w]hen added together, these goings-on in a particular domain form what one might call an *epistemic culture*' [Knorr-Cetina, 1991, p. 107]. Usually, the term 'epistemic culture' has been associated with specific sub-disciplines in science. Karin Knorr-Cetina famously characterized the epistemic culture of particle physics in terms of its intensely collaborative research environment and its extreme reliance on computational methods, leading simultaneously to a 'relative loss of the empirical' and 'a loss of epistemic status of the individual' [Knorr-Cetina, 1991, p. 120], as compared with other subdisciplines in physics and other branches of science. One might worry that referring to model-making in physics as an 'epistemic culture' creates an illusion of unity and masks the great diversity of domains across which, as we have seen, models in physics are being deployed. At the same time, it has often been observed that models in physics 'travel' from one discipline to another: every physics student is familiar with the ubiquity of the harmonic oscillator equations in various seemingly unrelated branches of physics, and even models that were originally intended for highly specific target systems, such as the Ising model of ferromagnetism, have over time been applied to a wide range of phenomena, from spin glasses to systems of neurons. Without a shared body of tacit knowledge about when, and how, to deploy models successfully – without, that is, a shared culture of modelling – it would seem difficult to explain such mobility of models across disciplinary boundaries.

5. Representation, exploration, and regulation

Our discussion so far, like much of philosophical discourse on scientific models, has – almost unreflectively – used the idiom of *representation* in connection with how models function. It is indeed commonplace to find models characterized exclusively as 'tools for *representing the world*' [Giere, 1999, p. 44] and to measure the success of a given model by how well it matches physical reality. When a model is tailored to specific phenomena and is intended to

[3] Preferences for realism or instrumentalism among scientists are often fleeting; on this point, see [Gelfert, 2005].

represent specific target systems, such a view has a lot going for it. Who, in all seriousness, would doubt that scientists often help themselves to models as 'stand-ins' for real target systems which scientists know of, but which they cannot directly access or give a complete characterization of? The ability of scientific models to enable us to draw inferences about actual target systems is surely one of the great attractions of using models in scientific inquiry.

Yet it takes only a moment's reflection to realize that representing actual target system – or representation, *simpliciter* – is just one way in which models are being applied in science. One might think that this is because of the inevitable use of abstraction, idealization, and approximation in the construction of models, all of which render model descriptions strictly speaking false. Yet representations obviously do not need to be completely accurate or complete in order to represent their targets. Partial representation, and even misrepresentation, are entirely compatible with models being 'stand-ins' for actual target systems. After all, for a model to represent its target, we should not require that it be a perfect, or even a particularly good, representation.[4] What we should require, though, is that, for something to be a representation, it should have a target in the actual world – put crudely, it should be doing some *representing*. This is not to deny that there are some hard ontological questions that a full account of model-based representation ought to be able to address. How much can a putative representation get wrong about its target, while still being considered a representation of said target – rather than a failed, or vacuous, attempt at representing (as, arguably, in the case of phlogiston)? Furthermore, there have been sophisticated attempts to widen the range of admissible targets, e.g., by including fictions among them. Notwithstanding pronouncements to the effect that 'there is absolutely no difference in kind between fictional and real-object representation – other than the existence or otherwise of the target' [Suárez, 2004, p. 770], even such an inclusive approach does not adequately capture the more overtly non-representational uses of models.

In recent years, there has been a growing recognition that the activity of *scientific modeling* goes well beyond the task of deriving representations of real-world target systems. Recent papers in this vein speak of the activity

[4] For a critique of the ideal of perfection in scientific modelling, see [Teller, 2001].

of 'modeling without models' [Levy, 2015] or of 'models in search of targets' [Gelfert, 2018]. As Arnon Levy notes:

> When a model is proposed it might not be clear at first what target it is tied to, and there might be a period in which the right target is sought. But later, assuming the model is retained, this issue is usually clarified. [Levy, 2015, p. 796]

A similar sentiment was evident in Peierls's discussion of different model-types: Recall how Peierls discusses Weiss's model of ferromagnetism – which posited the existence of 'microscopic magnets' even in the absence of a theory of what might constitute these – as 'very characteristic' of early-stage modeling; similarly, in his discussion of the Ising model of ferromagnetism, he noted the tenuous nature of assigning target systems:

> The model is unrealistic for ferromagnetism, because if the atomic spin is greater than ½ it has more than two orientations, and if it is ½, quantum effects are not negligible. The model is slightly more realistic for alloys, except that in metallic alloys the interaction between atoms is in part mediated by the conduction electrons, and such an interaction is by no means limited to nearest neighbours. [Peierls, 1980, p. 15]

As it turns out, for a model to function as a tool of scientific inquiry, it need not – at least not initially – refer to any real-world target system in particular. Even in the absence of a uniquely identifiable target, 'much can be learnt from the model' [Peierls, 1980, p. 15].

It would be wrong, however, to think of this indeterminacy with respect to a model's target as solely due to initial confusion or ambiguity in the early stages of inquiry, which always must be – and eventually will be – rectified. Tailoring a model to a specific target may be a promising strategy when we have already developed a good grasp of what constitutes, and what causal factors might contribute to, the phenomena in question. Yet in the absence of comprehensive theoretical knowledge – that is, in the context of *exploratory research* – the varied tasks of stabilizing phenomena, delineating their causal substrate, separating signal from noise, etc. are hardly straightforward. Sometimes it may not even be desirable to prematurely focus on certain target systems at the expense of others. Much of scientific modelling serves the purpose of exploring theoretical relationships, establishing 'no-go' theorems (thereby exploring the modal structure of potential phenomena), providing proofs of principle (which need not, in fact, be instantiated) – in all of these cases, focussing on the goal of representing specific real-world target systems would run the risk of impeding, rather than furthering, our understanding of the science involved, as a closer look at exploratory modelling in science reveals.

Recent case studies of exploratory modelling in science have shown that, under conditions of exploratory research, models can function in a variety of ways: as starting points for future inquiry, as proofs of principle, as sources of potential explanations, and as a tool for reassessing the suitability of the target system. (See also [Gelfert, 2016, pp. 71-100].[5]) This list is meant to be neither exhaustive nor mutually exclusive; as already mentioned, sometimes a fruitful line of future inquiry can be identified by looking at a range of potential target systems and considering whether any of them (or any aspect of them) are captured by a given (e.g. mathematical) model. Interestingly, this is just what Peierls states with respect to the Ising model. Having noted that the Ising model's failure to successfully represent does not preclude learning from it, he continues as follows:

> For example the Onsager solution for the two-dimensional case [of the Ising model] demonstrates that the specific heat is not merely discontinuous at the critical point […], but tends to infinity as the critical point is approached either form below or from above. This helped in the development of the theory of phase transitions, which by now is a very sophisticated branch of statistical mechanics. Several variants of the Ising model have served as a proving ground for methods and approximations in the field. [Peierls, 1980, p. 15]

In other words, Peierls explicitly recognizes that models can serve as 'proofs of principle' and explore the structure of underlying theoretical frameworks, not *in virtue of*, but *instead of* representing actual target systems.

There is another way in which the utility of models in physics extends beyond their representational success. Models often 'give stability to scientific practice'

[5] The framework of 'exploratory modelling' has since also been fruitfully applied to case studies from bioengineering [Poznic, 2016], mathematical biology [Gelfert, 2018], astronomy [Wilson, 2017], and mesoscopic physics [Shech and Gelfert, this issue].

[Gelfert, 2015, p. 224] by serving a regulative purpose. As mentioned earlier, the same model templates – e.g. in the case of mathematical models, the same equations – may travel across disciplinary boundaries, adding to the cohesion of the scientific enterprise. Scientists themselves acknowledge that models are not only used to derive predictions about specific target systems, but, as Peierls puts it with regard to the Weiss model, are 'useful if we want a quick orientation on the likely behaviour of a ferromagnet in unfamiliar circumstances' [Peierls, 1980, p. 8]. Deploying models at crucial junctures, whether in order to gain a 'quick orientation' or contribute to the development of a serviceable theory, can be an effective way of giving direction to a given process of inquiry and set it on course for future successes. Making decisions about when to rely on models, and which model to use, is 'where judgement and experience comes in', specifically that of 'the experienced physicist' [Peierls, 1980, 8]. Who is to be credited with the requisite experience is determined, in part, by the recurring norms and directives associated with the prevailing epistemic culture. Criteria of model choice themselves evolve from the collective repetition of the acts of choosing (and, where appropriate, dismissing) models; over time, they become what may be called 'normative-directival complexes' [Moraczewski, 2014, p. 41], which feed into, and in turn are shaped by, scientists' collective practices of model-making.

There will always remain the unavoidable risk of mistaking one kind of model for another, and even the best model – if ever there could be such a thing – may, on occasion, lead us into error. But, as Peierls puts it in a later paper, 'one also has to guard against the opposite mistake, of being too timid in learning something from an approach whose basis is not formally established' [Peierls, 1987, p. 95]. Yet such is the fallible nature of all inquiry, and exploring the structure of the empirical world around us requires us to take a chance. Science is not for the timid.

References

D. Bailer-Jones, "Scientists' Thoughts on Scientific Models," *Perspectives on Science*, vol. 10, no. 3, pp. 275-301.

R.B. Braithwaite, *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*, Cambridge: Cambridge University Press, 1968.

R. Carnap, „Foundations of logic and mathematics," in *Foundations of the Unity of Science* (vol. 1, eds. O. Neurath, R. Carnap, and C. Morris), Chicago: The University of Chicago Press, 1969, pp. 139-214.

A. Chakravartty, „Informational versus functional theories of scientific representation," *Synthese,* vol. 217, no. 2, pp. 197-213, 2010.

G. Contessa, „Editorial introduction to special issue," *Synthese,* vol. 2010, no. 2, pp. 193-195, 2010.

P. Duhem, *The Aim and Structure of Physical Theory*. (Transl. P.P. Wiener), Princeton: Princeton University Press, 1914/1954.

R. Frigg, S. Hartmann, „Models in science," *Stanford Encyclopedia of Philosophy*, 2012. [Online]. Available: plato.stanford.edu/entries/models-science/. [Accessed 10 February 2018].

A. Gelfert, "Mathematical rigor in physics: putting exact results in their place," *Philosophy of Science*, vol. 72 (2), 2005, pp. 723-738.

A. Gelfert, "Between rigor and reality: many-body models in condensed matter physics," in *Why More is Different: Philosophical Issues in Condensed Matter Physics and Complex Systems* (eds. B. Falkenburg and M. Morrison), Heidelberg: Springer, 2015, pp. 201-226.

A. Gelfert, *How to Do Science With Models: A Philosophical Primer*, Cham: Springer, 2016.

A. Gelfert, "Models in search of targets: exploratory modelling and the case of Turing patterns," in *Philosophy of Science: Between the Natural Sciences, the Social Sciences, and the Humanities* (European Studies in Philosophy of Science, vol. 9; eds. A. Christian, D. Hommen, N. Retzlaff, and G. Schurz), Heidelberg: Springer, 2018, pp. 245-269.

R. Giere, „Using models to represent reality," in *Model-based Reasoning in Scientific Discovery* (eds. L. Magnani, N. Nersessian, and P. Thagard), New York: Plenum Publishers, 1999, pp. 41-57.

N. Goodman, *Languages of Art*, Indianapolis: Bobbs-Merrill, 1968.

H.S. Green, "Review of *Selected scientific papers of Sir Rudolf Peierls*," *Mathematical Reviews*, MR1632685 (99e:01035), 1999.

M. Hesse, *Models and Analogies in Science*, London: Sheed and Ward, 1963.

K. Knorr-Cetina, "Epistemic cultures: forms of reason in science," *History of Political Economy*, vol. 23, pp. 105-122, 1991.

T. Knuuttila, "Some consequences of the pragmatist approach to representation," in *EPSA Epistemology and Methodology of Science* (eds. M. Suárez, M. Dorato, and M. *Rédei)*, Dordrecht: Springer, 2010, pp. 139-148.

A. Levy, "Modeling without models," *Philosophical Studies,* vol. 172, no. 3, pp. 781-798, 2015.

K. Moraczewski, *Cultural Theory and History: Theoretical Issues*, Poznań: WNS, 2014.

R. Peierls, "Model-making in physics," *Contemporary Physics*, vol. 21, no. 1, pp. 3-17, 1980.

R. Peierls, "Models, hypotheses and approximations," in *New Directions in Physics: The Los Alamos 40th Anniversary Volume* (eds. N. Metropolis, D.M. Kerr, and G.-C. Rota), San Diego: Academic Publishers, 1987, pp. 95-105.

M. Poznic, “Modeling organs with organs on chips: scientific representation and engineering design as modeling relations,” *Philosophy & Technology*, vol. 29, no. 4, pp. 357–371, 2016.

E. Shech and A. Gelfert, “The exploratory role of idealizations and limiting cases in models”, *Studia Metodologcizne* (this issue).

M. Suárez, “An inferential conception of scientific representation,” *Philosophy of Science,* vol. 71, no. Proceedings, p. 767–779, 2004.

M. Suárez and N. Cartwright, “Theories: tools versus models,” *Studies in History and Philosophy of Modern Physics,* vol. 39, no. 1, pp. 62-81, 2008.

P. Teller 2001, “Twilight of the perfect model model,” *Erkenntnis*, vol. 55, pp. 393-415.

J. von Neumann, „Method in the physical sciences," in *Collected Works Vol. VI. Theory of Games, Astrophysics, Hydrodynamics and Meteorology*, ed. A.H. Taub, Oxford, Pergamon Press, 1961, pp. 491-498.

S.R. Weart, “The solid community,” in *Out of the Crystal Maze: Chapters from the History of Solid-State Physics* (eds. L. Hoddeson and E. Braun, J. Teichmann, and S. Weart), New York: Oxford University Press, 1992, pp. 617-669.

K. Wilson, “The case of the missing satellites,” *Synthese* (Online First, 2017, https://doi.org/10.1007/s11229-017-1509-6).

Axel Gelfert
Institute of History and Philosophy of Science, Technology, and Literature
Technische Universität Berlin
axel@gelfert.net

STUDIA METODOLOGICZNE
NR 39 • 2019, 73–98
DOI: 10.14746/sm.2019.39.3

Steven Hecht Orzack, Brian McLoone

Modeling in Biology: looking backward and looking forward

Abstract. Understanding modeling in biology requires understanding how biology is organized as a discipline and how this organization influences the research practices of biologists. Biology includes a wide range of sub-disciplines, such as cell biology, population biology, evolutionary biology, molecular biology, and systems biology among others. Biologists in sub-disciplines such as cell, molecular, and systems biology believe that the use of a few experimental models allows them to discover biological universals, whereas biologists in sub-disciplines such as ecology and evolutionary biology believe that the use of many different experimental and mathematical models is necessary in order to do this. Many practitioners of both approaches misunderstand best practices of modeling, especially those related to model testing. We stress the need for biologists to better engage with best practices and for philosophers of biology providing normative guidance for biologists to better engage with current developments in biology. This is especially important as biology transitions from a "data-poor" to a "data-rich" discipline. If 21st century biology is going to capitalize on the unprecedented availability of ecological, evolutionary, and molecular data, of computational resources, and of mathematical and statistical tools, biologists will need a better understanding of what modeling is and can be.

Keywords: biology, model building, model testing, philosophy of biology, subdisciplines of biology.

"Biology" is the study of nature. It dates as far back as our ancestors paid attention to the benefits and hazards of their surroundings. A possible vestige of this study may be the common but not universal fear of snakes and spiders, which could be due to long-ago observations that some are dangerous (Rakison 2018).

More formal biology is less ancient but still has a long history. Aristotle was a biologist among other things and his contributions to the discipline are profound (Egerton 1975, 2001; Balme 1987; Gotthelf 1999, 2012; Leroi 2014). His study of nature was not just passive observation; for example, he may have sequestered fish in order to study their foraging behavior (Tipton 2008).

Today, biology is organized into a wide range of sub-disciplines such as cell biology, ecology, evolutionary biology, molecular biology, systems biology, and population biology. These have divisions as well, such as experimental population biology and theoretical population biology. There are also sub-disciplines that focus on particular kinds of organisms, such as botany, entomology, mammalogy, microbiology, ornithology, virology, and zoology. Use of these descriptors depends on context. A biologist might say to another biologist "My colleague is a microbiologist but I am an ecologist", whereas in conversation with the public (s)he might say "My colleague and I are biologists".

1. Sub-disciplinary differences in the way biologists create, understand, and use models

We now discuss sub-disciplinary differences in modeling practices and in the institutional context in which modeling occurs. In doing so, we use this "folk" definition enunciated by Barbour (1974, p. 6): "[a model is] a symbolic representation of selected aspects of behavior of a complex system for particular purposes…". We believe that this captures the loose definition of model used by many biologists (see also Lewontin 1968); Frigg and Hartmann (2012) review more nuanced philosophical considerations of what models are and how they are used in science.

Models in biology can be computational, experimental, mathematical, or verbal. Even within these categories, there is variety. For example, an experimental model can be an actual organism (e.g., Tickoo and Russell 2002) or a physical representation of an organism (e.g., Colbert 1962). Mathematical models can be deterministic or stochastic (e.g., Bartlett 1956).

Most biochemists, cell biologists, molecular biologists, neurologists, pharmacologists, and systems biologists use experimental models involving one or a few types of organisms. A few of the almost innumerable examples of such "microcosms" as models are the use of cells in culture to investigate a cellular process (e.g., Szostak, Orr-Weaver, Rothstein, & Stahl, 1983) and in-vitro analysis of an enzyme and substrate to investigate the enzyme's in-vivo activity (e.g., Tcherkez et al., 2013; Torres, Mateo, Melendez-Hevia, & Kacser, 1986). Mathematical models are rare in these sub-disciplines although some are very influential (e.g., Hodgkin and Huxley 1952; Kacser and Burns 1973; Byrne 2010).

Models in these sub-disciplines are usually conceived of as providing insight into "the" biology of the cell, enzyme, or pathway studied. This typological conception is based upon beliefs that the model is "about" large classes of organisms and not specifically about the species used, that differences among species are "noise", and that related species would provide only redundant information. These beliefs sustain careers dedicated to specific experimental models. Almost all research in molecular biology involves less than ten species, with the mouse *Mus musculus* and the zebrafish *Danio rerio* being the "universal" vertebrates (Dooley and Zon 2000; Sharpless and DePinho 2006), the cress *Arabidopsis thaliana* being the "universal" plant (Woodward and Bartel 2018), the bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae* being the "universal" microorganisms (Orr-Weaver et al. 1981; Lee and Lee 2003), and the worm *Caenorhabditis elegan*s and the fly *Drosophila melanogaster* being the "universal" invertebrates (Rankin et al. 1990; Tickoo and Russell 2002). In contrast, there are thousands of species currently studied by ecologists and evolutionary biologists. There is no universal model. The variety of species involved ranges from viruses (e.g., Bull, 2006) to whales (e.g., Alexander et al., 2016).

Microcosms are also used as models in, say, ecology and evolutionary biology. These are often experimental investigation of a trait in the laboratory or in the field; there are many thousands of examples, some dating back to the beginnings of these disciplines (e.g., Beal 1885; Lutz 1915; Sturtevant 1915). Increasingly, some of the same microcosms used are those used in biochemistry, cell biology, and molecular biology (Jarosz and Dudley 2017; Kawecki et al. 2018). However, an evolutionary biologist studying cells would investigate the influence of, say, natural selection on the rate of cell division, whereas the cell biologist would study the mechanics of cell division. This distinction is often viewed by biologists as being a distinction between a focus on "why?" and a focus on "how?"; see Mayr (1961), Tinbergen (1963), Ariew (2003), Laland et al. (2011), Bateson and Laland (2013), and Nesse (2013).

The beliefs accompanying models in ecology and evolutionary biology are usually that differences among species are "signal" not "noise" and that different species provide unique information. One reason for the focus on differences is that they are causally central in evolutionary processes such as natural selection (cf., Murray 1991). These beliefs sustain careers dedicated to different kinds of organisms (e.g., herpetologists, mammalogists, and ornithologists, many of whom study one or a few species, which are often chosen because they are *not* studied by others).

Ecologists and evolutionary biologists also traffic in more mathematical models, especially those that are viewed as having broad relevance, than do cell, molecular, and systems biologists. Many ecologists and evolutionary biologists view results derived from mathematical models as providing conceptual insights that are central to their work, even if it is experimental. Examples in ecology of such mathematically-derived model insights are (Levins 1968*a*; MacArthur 1970); examples in evolutionary biology include Fisher (1930), Wright (1931), and Haldane and Jayakar (1963).

What underlies these sub-disciplinary differences in regard to the influence of mathematical models? It is not as though cellular and molecular processes are inherently un-amenable to mathematical analysis. The focus on the experimental microcosms in cell, molecular, and systems biology is largely a result of the influence of Thomas Hunt Morgan and of Jacques Loeb. Each did early widely-influential 20th century work in embryology, genetics, and physiology and advocated for a biology in which the use of experimental models is paramount (Allen 1978; Pauly 1987; Brush 2002). They trained or mentored many biologists who gained substantial influence and who in turn trained many more biologists (so much so that many thousands of current biologists are their scientific "descendants"; the first author is one of them).

A focus on experimental microcosms licenses career and institutional investment in work to master the techniques needed to yield interpretable results. These influences act to diminish the use of mathematical models. The notion that broad insights are to be gained via the investigation of "universal" experimental models is paramount.

Sub-disciplinary differences in modeling are increasingly influenced by institutional structure. Some important institutions for biological research and training in the United States have few if any faculty members doing research in ecology and evolutionary biology; these include Brandeis University, California Institute of Technology, Johns Hopkins University, Massachusetts Institute of Technology, Northwestern University, Rockefeller University, and the Salk Institute for Biological Sciences. At several, a "Department of Biology" became solely focused on cell and molecular biology by the elimination of other sub-disciplines. We know of no examples in which a "Department of Biology" became solely focused on ecology and evolution by the elimination of other sub-disciplines. At other institutions, sub-disciplines have not been eliminated but the "Department of Biology" has fissioned into the "Department of Cell and Molecular Biology" and the "Department of Ecology and Evolutionary biology" (or name variants thereof). Both kinds of transformation sustain if not strengthen sub-disciplinary differences in modeling.

2. Sub-disciplinary similarities in the way biologists view the process of model building

Despite differences in types of model typically used, sub-disciplines of biology are similar in two ways. The first is that most practitioners have an aversion to discussion of the nature and practice of modeling. This is especially true of normative guidelines for modeling. Such an activity is often deemed to be "unscientific" and a "waste of time", an underlying trope being that "facts and experiments mean something, philosophizing does not". This attitude may in part be due to a lack of exposure to such material during training in biology, which often has little or no exposure to mathematics and statistics (cf., Bialek and Botstein 2004; Marshall and Durán 2018). By contrast, even undergraduate training in physics includes topics such as relativity and quantum mechanics that necessitate some exposure to "philosophizing" about modeling and about what observations mean.

The second way in which the sub-disciplines are similar is that few researchers have called for increased attention to what modeling is and should be.

In cell, molecular, and systems biology, such calls include Gunawardena (2014), Torres and Santos (2015), Bartocci and Lió (2016), Sztul (2017), and Medina (2018). Most appeal for more use of computational and mathematical models. These calls appear to be motivated by the massive increase in the availability of data concerning the genomic, metabolomic, proteomic, and transcriptomic "levels" of the organism. For example, just *one* study (Telenti et al. 2016) reports on 2.8 x 10^{13} nucleotides sequenced from 10,545 humans and documents 150 million sequence variants. Another motivation is the belief that the new data make it possible to understand biological "complexity". These calls are not just for the use of computers to help collect and store data. They are calls that the increased use of computational and mathematical models is *required* to provide new kinds of answers (e.g., Berro 2018; Wallmeier 2018). Whether or not this is true, it is too early to tell whether these calls will be widely-heeded and to what extent this increased use provides answers that would be much more difficult or impossible to attain via the use of experimental models.

In ecology and evolutionary biology, calls for increased attention to what modeling is and should be have a longer history. Prominent among these are Holling (1964) and Levins (1966), who claimed that generality (termed "breadth" by Holling), precision, and realism are three desired attributes of a model in population biology (and of models in general). Both authors discussed the tradeoff among these attributes, with Holling claiming that modifying a model so as to increase any two, say, generality and realism, need not decrease precision, and Levins claiming that such a decrease is unavoidable. A necessary tradeoff implies that there are three types of models, which differ in which attribute is sacrificed in order to increase the other two. Levins viewed models with more generality and realism and less precision as most desirable. He did not define how generality, realism, and precision of a model can be assessed or demonstrate that they exhibit a necessary tradeoff. Levins further claimed that (p. 422) "Our truth is the intersection of independent lies", i.e., we can regard the common prediction arising from multiple "independent" models as "truth".

More than fifty years have passed since the publication of Levins' article and it has received over 1900 Google Scholar citations as of early 2019; there are far fewer citations of Holling's article, which was not cited by Levins. What can we conclude from this large number of citations of Levins' article? Does it mean that his claims have improved modeling in biology and that they should continue to provide guidance?

The answer to this question is unclear. Levins' article has often been cited in ways that suggest misunderstanding of or disagreement with his claims (see Orzack and Sober 1993, who reviewed all citations of his article up to 1993, and Orzack 2005). For example, Armstrong (1988) claimed that his model of an ecological community *is* general, realistic, and precise. Most often claims about model attributes are ambiguous because they are not anchored in explicit comparisons of models (as in "model x is more realistic than model y"). Many citing authors describe their model as being the type preferred by Levins, which at least naively can be taken to imply belief in the existence of his trichotomy of models, but what features of their model make it this type are not specified. The character of most citations of Levins' paper suggests (but does not prove) that they are mainly an effort to provide "quasi-philosophical" support for a model apart a demonstration that it provides biological insights.

Much of this history reflects the attitudes of biologists a generation or more ago. It is still conceivable that Levins' claims are relevant to modeling in the 21st century. Do biologists believe this to be true? One way to judge this is to assess the recent biological literature as recorded in the Google Scholar database. For example, it lists approximately 20,800 articles published in 2017 that contain "biology" and "modeling" in their abstract or title. Of these, 78 cite Levins' article. These citations occur almost always in the context of population and evolutionary biology, even though his claims apply to any natural science and his article was published in a general science journal. For example, as of March 2018, Google Scholar lists 366 post-2013 articles that have "biology" and "modeling" in the abstract and which cite Levins' article. The majority of these articles present original biological research; just two do not concern population or evolutionary biology (Shirsat et al. 2015; Ho et al. 2018). (Other citations are by philosophers discussing Levins' ideas or are by other kinds of scientists).

We do not know whether non-citation is mostly the result of scientists being unaware of Levins' article or that they deem it irrelevant or wrong. (It is even possible that non-citation occurs because it is regarded as self-evidently correct.) What is clear is that few biologists (much less other scientists) cite Levins' article and that almost all biologists that do cite it study population or evolutionary biology. Perhaps their non-citing colleagues are ignorant; perhaps they are enlightened.

The apparently very small constituency of Levins' article among biologists raises questions about its constituency among philosophers of biology and biologists concerned with the philosophical implications of modeling in biology. Critical assessment of his article began with Orzack and Sober (1993). Some commentators believe that Levins' article provides important insights and normative claims about modeling (Wimsatt 1987, 1981; Godfrey-Smith 2006; Plutynski 2006; Weisberg 2006*a*, *b*, 2007, 2013; Weisberg and Reisman 2008; Matthewson and Weisberg 2009; Goldsby 2013). Others are skeptical (Orzack and Sober 1993; Orzack 2005; Orzack 2012; Odenbaugh and Alexandrova 2011; Justus 2012). A main point of debate is Levins' claim about tradeoffs. Orzack and Sober (1993) showed that one can increase the generality and realism of a model without a decrease of its precision (see responses by Levins 1993 and Matthewson and Weisberg 2009).

This debate appears to have done little to help biologists develop better models, perhaps in part because the debate mainly involves philosophers of biology and journals typically not read by biologists (e.g., Biology & Philosophy). Almost all citations by biologists are at most acknowledgements that there is a debate (e.g., Martínez del Rio 2008) although some biologists engage with its substance (e.g., Slobodkin 1994). Perhaps some have been spurred to create models with increased generality, realism, and precision as compared to previous models after reading Orzack and Sober's demonstration that tradeoffs are not inevitable. If so, they have done this without citing Levins (1966) or Orzack and Sober (1993).

Many of the philosophers engaged in the debate over Levins' claims appear to have limited knowledge about the practice of modeling in biology. Instead, they appear to derive their knowledge about modeling from Levins' article. They also appear to be unaware of the very small constituency that Levins' claims have among biologists and of how much biology has changed since the 1960s (see above), which reduces the practical relevance of Holling's and Levins' claims. Better knowledge of current biology would likely improve the philosophical and normative content of the debate.

We expect that the lack of connection between practitioners of biology and those who could provide useful conceptual and normative guidance will remain if not increase, given the ongoing avalanche of ecological and molecular data, which underlies claims that the 21st century is the century of biology (Venter and Cohen 2004; National Research Council 2009) and that there will be breakthroughs in, for example, the treatment of diseases and the remediation of environmental degradation. This possibility might cause biologists to be more open to normative guidelines derived from more formal consideration of what modeling is and should be. However, this appears not to be true as of yet.

The future usefulness to biologists of Holling's and Levins' claims is unclear. Perhaps they will provide normative guidance, even if this only amounts to better awareness of how model attributes such as generality, realism, and precision relate to one another. That said, their claims are rooted in the doing of the data-poor biology of fifty years ago. Their claims might be justified as an attempt to understand complex systems in the relative absence of data. Whatever relevance their insights still have, current biologists have abundant data and resources needed to analyze them such as computers, databases, and statistical and mathematical tools. Biologists and philosophers of biology hoping to provide much-needed normative guidance to biologists using models must pay attention to the realized character of current data and tools, not those of the 1960s.

One consequence of the availability of data is that the goal of biological modeling is increasingly the identification of a model that makes a *non-robust* prediction. In particular, biologists often seek a model is tailored to fit the specific biology under investigation and is not necessarily useful outside of this domain. Predictions are not derived from an ensemble of models (cf. Levins 1966, p. 423). There appears to be a diminishing potential constituency for Levins' claim that the identification of such robust model predictions is a good way to discover biological truth. This search for models that are non-robust has been aided by the development of the Akaike Information Criterion, which can be used as a basis for choosing which model from a set of plausible models has the greatest support from the data (Akaike 1974; Burnham and Anderson 2002; Johnson and Omland 2004).

The search for a model with a non-robust prediction occurs in many sub-disciplines of biology. It has become especially common in ecology and evolutionary biology. For example, the determination of a phylogeny or evolutionary "tree" for a group of species given data on the trait expressed by each species now routinely involves finding the instantiated evolutionary model that makes the data most likely (e.g., Felsenstein 2004; Lemey et al. 2009). The investigator often does not even choose the uninstantiated model of the trait's evolution and so it is unclear in what sense (s)he is aware of potential tradeoffs among models.

As cell, molecular, and systems biology encompass more and more data, there has also been more explicit formulation of computational and mathematical models that make non-robust predictions. For example, Altan-Bonnet and Germain's (2005) mathematical model of signaling in the immune system (see their Figures 2B and S6) is based upon a "complete" representation of the signaling pathway underlying the T-cell antibody response. This representation, which includes hundreds of reaction steps, is based upon experimental work to elucidate the pathway. As such, the predictions of this model are non-robust in as much as it is intended to exactly represent the biology under investigation. There is no search for a prediction that is common to several models. As in the case of experimental models in these sub-disciplines of biology, the expectation is that the predictions are relevant to many organisms, not just those from which the experimental data were derived.

3. Causal models, data models, and algorithmic models

The search for non-robust predictions is part of a broader move by scientists to focus on algorithms and data as hypothesis "generators". This tendency has been addressed by Breiman (2001) who distinguished between the "data modeling culture" and the "algorithmic modeling culture". In the former, the scientist chooses an underlying statistical model that (s)he believes generated the data and then uses it to make inferences about the observed data (e.g., whether the arithmetic average of an observed sample of data has a low or high probability of occurrence given random sampling of data generated by the chosen statistical model). In the latter, the scientist generates "black box" predictions from the data, without recourse to an initial choice of an underlying statistical model.

Breiman's distinction is couched in terms of statistics and so it does not describe what one could call the "causal modeling culture" in biology, in which idealized features of a biochemical pathway, or an organism, or a population, or a ecosystem, etc., are used to create the model. For example, when an ecologist creates a model of population dynamics, (s)he chooses whether or not there is a single population, whether or not a population is finite, whether or not two sexes are present, whether or not individuals mate randomly, whether or not the number of offspring produced is finite, whether or not there is environmental variation, and so forth (see example below). Statistical analysis of data does not inform the choice of model alternatives offered (although the choice of a particular alternative on offer is sometimes informed by statistical analysis). Hypothesis testing need not be the goal of the model. When done, it requires assumptions extrinsic to the assumptions used to create the model. In contrast, hypothesis testing of "statistical" and "algorithmic" models is intrinsic to the model (either because an assumption is made about the error distribution or the data are used to generate the distribution).

This distinction is underscored by the fact that some "causal" model assumptions could never be supported by data. For example, it is never true that, say, a population has infinite size, that individuals have an infinite number of offspring, that the environment is constant, etc. Nonetheless, one or more of these assumptions have long been used in the creation of "causal" models in ecology, many of which have yielded important insights.

In sub-disciplines dominated by experimental models, the "causal modeling culture" has led to some of the great discoveries of 20th century biology. For example, Meselson and Stahl (1958) created two distinct causal models in order to identify the correct mechanism for the replication of the DNA helix. However, data models play an increasing role in these sub-disciplines. For example, mutations in a DNA sequence are often identified as being potentially disease-causing solely by the strength of their statistical association with the disease and not by their consequences for the function of the associated protein or of the associated biochemical pathway (Balding 2006; Edwards et al. 2013).

In sub-disciplines in which mathematical models are widely used, the use of "causal" models is common. For example, in evolutionary biology, there is a well-known model that is intended to explain the female-biased sex ratios found in some insect populations (Hamilton 1967). The model predicts that a female-biased sex ratio is the most evolutionarily-beneficial or "optimal" when a finite number of fathers and mothers comprise a "local" mating group. (Other assumptions are also required.) This prediction is based upon idealized biology that can never be true, e.g., that a mother produces an infinite number of offspring. (The prediction changes when offspring number is assumed to be finite, see Nagelkerke 1996). Hamilton's model is not a "data" model or an "algorithmic" model; the prediction of the model is not derived from statistical analysis of the relationship between the sex ratio and the number of fathers and mothers in a mating group.

Other important causal models in evolutionary biology have led the reconciliation of the genetic mechanisms underlying traits that vary discretely and those that vary continuously (Fisher 1918) and to the development of population-genetic models that include deterministic and stochastic evolutionary forces (e.g., Wright 1931). In turn, both of these led to the development of a highly influential but still controversial "synthetic" causal model that connects short-term and long-term evolution (e.g., see Laland et al. 2014; Wray et al. 2014).

The "causal modeling culture" and "data modeling cultures" are very different. Bringing them together is not just a matter of "adding" standard assumptions about sampling error to causal models (see below). Unfortunately, the important distinctions between the two cultures are often misunderstood. For example, Gunawardena (2014) confounds the two in his overview of modeling in biology when he writes (p. 6) that "Judging from some of the literature, we seem to forget that a model does not predict the data to which it is fitted: the model is chosen to fit them." Similarly, Nijhout et al. (2015) write (p. 2) that "… the 'model' is not a fixed object, but continually evolves through testing it against data and revising it accordingly." In fact, causal models as defined here are *very rarely* formed and re-formed by data in the direct sense these authors describe (although observed data may inspire the creation of a causal model). Their descriptions concern the "data modeling culture" although the models they discuss are products of the "causal modeling culture".

Although the "causal modeling culture" and the "data modeling culture" have long histories in biology, they are problematic in important ways. An important problem for causal models can be illustrated in the context of the optimality model in evolutionary biology mentioned above. Consider a deviation between the optimal sex ratio predicted by Hamilton's model and an observed sex ratio. How large can such a deviation be and still allow us to conclude that the model provides a causal explanation of the data? The deviation cannot automatically be assumed to be due just to sampling error because it has causal implications. All other things being equal, the deviant observed sex ratio has a lower evolutionary fitness than does the optimal sex ratio and thereby should not be observed in the population. Is the observed sex ratio optimal but not correctly specified by an incorrectly-formulated model? Or is the model correctly formulated but the optimal sex ratio cannot evolve in the population studied? Neither can be assumed to be true a priori.

These central issues concerning how to understand the relationship between causal model predictions and data are almost never explored. The consequence has been inferential ambiguity because there are no agreed-upon standards by which success and failure are judged. The criteria used are often private and apparently arbitrary. For example, the *same* test of the predictions of an optimality model has engendered these opposite assessments: "there is a striking correspondence between theory and data" and "there is a great deal of scatter around the quantitative prediction" (see Orzack 2014 for details). These statements were based solely on visual inspection of the predictions and data, which is known to be strongly influenced by the graphical presentation (Cleveland and McGill 1987). Each of these assessments licensed opposite conclusions as to whether the observed data indicate that the species possesses an optimal trait. Sometimes a "qualitative" test of model predictions is used in which a predicted trend (e.g., downward) is compared with the observed trend of the data. This is not inherently problematic, although it leaves unresolved what causal conclusions one should draw from concordance or discordance between the predicted and observed trends. These ambiguities in regard to the means by which the correspondence between data and causal model predictions is interpreted have received almost no attention by biologists; they are addressed in the context of testing optimality models by Orzack (1990), Orzack et al. (1991), and Orzack and Sober (1994).

The "data modeling culture" is also problematic but in part for different reasons. As noted by Breiman (2001), the assumptions routinely made about the model underlying the data are often of unknown validity at best. In addition, the criteria by which associations are judged to be "significant" or not are arbitrary and may often lead to incorrect conclusions (Ioannidis 2005).

We note that the distinction we make between "causal models" and "data models" is not meant to imply that the former are the only way to gain causal understanding. Shipley (1999) and Spirtes et al. (2000) present a method for making causal inferences from "data models". See also Pearl (2009). Their important motivation is that one often is confronted with the need to make causal inferences when controlled experiments are difficult or impossible. Their method is little-used in biology and wider implementation is required before we can assess how often it produces biologically-meaningful causal insights. If it often does so (despite important problems with their statistical approach, see Karlin et al. 1983; Freedman 1987, 1997), their method will be important, if only because it could partially reduce the need for experimental intervention to infer causation.

The "causal modeling" and "data modeling" cultures continue to play prominent and often useful roles in modeling in most if not all sub-disciplines of biology. That said, as Breiman (2001) noted, there is an increasing "algorithmic modeling culture". An important manifestation of this culture is the use of machine learning in the analysis of data. Here, automated procedures are used to make predictions from data. The perceived advantage of these methods is that they allow the investigator to forego much of the hard work needed to construct, analyze, and validate a causal model and or a statistical model. The use of machine learning is nicely illustrated by Olden et al. (2008) who used three different approaches (classification and regression trees, artificial neural networks, and evolutionary algorithms) to investigate the causes of heterogeneity in the number of fish species found in over 8,000 freshwater lakes in Canada. The results of all approaches suggest that important determinants of the number of species are the amount of precipitation, the length of the shoreline, and the area of the lake. Their analyses exemplify the power of the "algorithmic modeling culture", especially that a huge amount of data have been encompassed in the analyses, the analyses are carried out with software that is readily available, a causal model has not been created "from scratch", and testable predictions about causes have been specified. As Olden et al. (2008) note, these methods have their weaknesses (such as the potential of over-fitting; see their Table 1). That said, these methods allow biologists to tackle analyses that were previously un-addressable. The importance of this cannot be exaggerated.

In this context, it is worth noting that Levins' claims about model building were in part intended to counter the view that ecologists need to focus on data analysis and numerical simulation so as to understand "systems ecology" (Levins 1968*b*; Palladino 1991). Levins claimed that this approach could not lead to causal understanding because (p. 421) there are "too many parameters to measure", the "equations are insoluble analytically", and any predictions would "have no meaning for us". However, none of these is necessarily true for ecological models, even those in systems ecology. Of course, there are models for which there are "too many parameters to measure". But for some, there are sufficient data to allow all parameters to be estimated. Of course, there are models for which "equations are insoluble analytically". But for some, the equations are soluble or can be solved via analytical approximation. Of course, there could be model predictions that "have no meaning for us". But for some, predictions have meaning (or are eventually understood). Despite all of these possible difficulties, there is no basis for a claim that a particular approach to modeling could not result in causal understanding. It is Levins' absolute claim that lacks substantiation. He correctly identified potential difficulties but that does not inform how often they occur. In addition, Levins failed to mention that these difficulties may arise in the approach to model building that he advocates.

It is also worth noting that our ability to solve some of these difficulties is not static. For example, machine learning analyses in many sub-disciplines in biology demonstrate that prediction and causal understanding of many-dimensional "complex" systems can be attained even when many but not all parameters can be measured and when likely-relevant equations are insoluble (e.g., see Shan et al. 2006; Tarca et al. 2007; Wernick et al. 2010; Kampichler et al. 2010; Touw et al. 2012; Sommer and Gerlich 2013; Schrider and Kern 2015, 2016, 2018; Dumancas et al. 2017; Dietze et al. 2018; Flassig and Schenkendorf 2018; Ghosal et al. 2018).

Although Levins' claim is false as description of what is necessarily true about models, it does serve as a reminder to not view any approach to data analysis and causal understanding as "automatic". After all, different machine learning approaches can yield different results (e.g., Olden et al. 2008; Kampichler et al. 2010). Biological judgment will always be needed.

A development related to the "algorithmic modeling culture" is the computational "reverse engineering" of biological systems. Here, data and "candidate" symbolic representations of the dynamics are iteratively combined so as to ultimately generate "the" equations underlying the dynamical system (Bongard and Lipson 2007; Schmidt and Lipson 2009; Brunton et al. 2016). This approach to model generation underscores how far biological modeling can go beyond apparent limitations, such as those identified by Levins. It has remarkable potential, even it is not "automatic" and must be applied with judgment.

4. Whither Biology?

It is unclear as to whether there will be "movement" to close the gaps among sub-disciplines of biology in regard to the culture of modeling and also to close the gap between biologists and philosophers and others interested in biological modeling. In the former case, it is encouraging that sub-disciplines that have very different histories in regard to modeling at least have in common a search for models with non-robust predictions (although it is unlikely that the disparate groups of biologists understand that this is a common goal). In the latter case, if biologists and philosophers of biology are to provide normative guidance so as to improve the practice of biological modeling, they must situate their guidance in the context of current practice of biology. The biology of today is not the biology of twenty years ago, much less fifty years ago.

We have described some of the extraordinary changes that all sub-disciplines in biology have undergone because of unprecedented increases in data and computational resources. As daunting as it can seem to biologists to assimilate new data and techniques, it is likely that the most important task faced by biologists is being open to change in regard to what biological "complexity" is conceived to be.

Complexity is a badge of honor for many biologists. In many sub-disciplines, biologists use it to represent the notion that systems under investigation, whether they be biochemical pathways, organs, or ecosystems, are tangled inscrutable webs of interactions. The trope is that we can only barely understand the simplest aspects of such systems and that we will *always* fail in regard in our attempts to provide complete understanding. This attitude commonly coexists with the notion that the only way to achieve partial understanding is a reductionist approach that involves "disassembling" the system. This approach is regarded as necessary but never sufficient to understand complexity.

In fact, this combination of attitudes is not inherently problematic and will continue to be fruitful. That said, we emphasize the need to be open to new understanding of what biological complexity actually is. It is telling that natural systems that could be construed as similarly highly complex and inscrutable are understood in different ways by different kinds of scientists. Phenomena such as cellular metabolism and energy flow in ecosystems are "complex" to biologists. Phenomena such as climate dynamics are "complex" to physicists, meteorologists, and geophysicists. Yet, biologists view the complexity as a manifestation of a unknowable multitude of interactions of roughly equivalent magnitude, whereas physical scientists view the complexity as a manifestation of a multitude of interactions but dominated by just a few (e.g., Ditlevsen and Johnsen 2010; Cimatoribus et al. 2012). In effect, physical scientists view the complexity as *simple*. This is sometimes justified by claims about separation of time scales, with some processes being "fast" enough that they only add noise to the "slow" low-dimensional drivers of the system (cf., Ditlevsen and Johnsen 2010, p. 2). Correct or not, the concordance perceived by physical scientists between the dimensionality of the system and the dimensionality of the analysis is desirable at least in terms of inferential consistency. The discordance in this regard on the part of biologists is problematic and they would do well to eliminate it. In some instances, this might mean embracing complexity in the full sense (and abandoning low-dimensional tools) and in others by viewing complexity as simple and taking simple models seriously as providers of casual explanation (e.g., Reynolds 1987).

The contrast between biologists and physical scientists in attitudes towards the nature of complexity suggests that biologists may make substantial conceptual progress by being open to the possibility that complexity can be tractable. It is promising in this context to note that biologists studying "complex" systems they view as dynamically complex and physical scientists studying "complex" systems they view as dynamically simple often use the *same* tools from deterministic dynamical systems theory (e.g., see Strogatz 2018). The promise of this overlap is that it may help biologists change their understanding of how to model and understand biological complexity.

Acknowledgements

We thank an anonymous reviewer, Bruce Glymour, Elliott Sober, and Krzysztof Nowak-Posadzy for assistance and comments.

References

Akaike, H. 1974: A new look at the statistical model identification. IEEE Transactions on Automatic Control AC-19:716–723.

Alexander, A., D. Steel, K. Hoekzema, S. Mesnick, D. Engelhaupt, I. Kerr, R. Payne, and C.S. Baker. 2016: What influences the worldwide genetic structure of sperm whales (Physeter macrocephalus)? Molecular Ecology 25:2754–2772.

Allen, G.E. 1978: Thomas Hunt Morgan: The Man and His Science. Princeton University Press, Princeton.

Altan-Bonnet, G., and R.N. Germain. 2005: Modeling T cell antigen discrimination based on feedback control of digital ERK responses. PLoS Biology 3:e356.

Ariew, A. 2003: Ernst Mayr's "ultimate/proximate" distinction reconsidered and reconstructed. Biology & Philosophy 18:553–565.

Armstrong, R.A. 1988: The effects of disturbance patch size on species coexistence. Journal of Theoretical Biology 133:169–184.

Balding, D.J. 2006: A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7:781–791.

Balme, D.M. 1987: The place of biology in Aristotle's philosophy. Pp. 9–20 *in* A. Gotthelf and J.G. Lennox, eds. Philosophical Issues in Aristotle's Biology. Cambridge University Press, Cambridge.

Barbour, I.G. 1974: Myths, Models, and Paradigms: A Comparative Study in Science and Religion. Harper & Row, New York.

Bartlett, M.S. 1956: Deterministic and stochastic models for recurrent epidemics. Pp. 81–109 *in* J. Neyman, ed. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability Volume 2: Contributions to Probability Theory. Vol. 4. University of California Press, Berkeley.

Bartocci, E., and P. Lió. 2016: Computational modeling, formal analysis, and tools for systems biology. PLOS Computational Biology 12:e1004591.

Bateson, P., and K.N. Laland. 2013: Tinbergen's four questions: an appreciation and an update. Trends in Ecology & Evolution 28:712–718.

Beal, W.J. 1885: The viability of seeds. Proceedings of the Society for the Promotion of Agricultural Science 5:44–46.

Berro, J. 2018: "Essentially, all models are wrong, but some are useful"—a cross-disciplinary agenda for building useful models in cell biology and biophysics. Biophysical Reviews:10(6): 1637–1647.

Bialek, W., and D. Botstein. 2004: Introductory science and mathematics education for 21st-century biologists. Science 303:788–790.

Bongard, J., and H. Lipson. 2007: Automated reverse engineering of nonlinear dynamical systems. Proceedings of the National Academy of Sciences 104:9943–9948.

Breiman, L. 2001: Statistical modeling: the two cultures (with comments and a rejoinder by the author). Statistical Science 16:199–231.

Brunton, S.L., J.L. Proctor, and J.N. Kutz. 2016: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences 113:3932–3937.

Brush, S.G. 2002: How theories became knowledge: Morgan's chromosome theory of heredity in America and Britain. Journal of the History of Biology 35:471–535.

Bull, J.J. 2006: Optimality models of phage life history and parallels in disease evolution. Journal of Theoretical Biology 241:928–938.

Burnham, K.P., and D.R. Anderson. 2002: Model Selection and Multi-model Inference: A Practical Information-theoretic Approach. Springer, New York.

Byrne, H.M. 2010: Dissecting cancer through mathematics: from the cell to the animal model. Nature Reviews Cancer 10:221–230.

Cimatoribus, A.A., S.S. Drijfhout, V.N. Livina, and G. Van der Schrier. 2012: Dansgaard-Oeschger events: tipping points in the climate system. Clim Past Discuss 8:4269–4294.

Cleveland, W.S., and R. McGill. 1987: Graphical perception: The visual decoding of quantitative information on graphical displays of data. Journal of the Royal Statistical Society. Series A (General) 150:192–229.

Colbert, E. 1962: The weights of dinosaurs. American Museum Novitates 2076:1–16.

Committee on a New Biology for the 21st Century. 2009: A New Biology for the 21st Century. The National Academies Press, Washington, DC.

Dietze, M.C., A. Fox, L.M. Beck-Johnson, J.L. Betancourt, M.B. Hooten, C.S. Jarnevich, T.H. Keitt, M.A. Kenney, C.M. Laney, L.G. Larsen, H.W. Loescher, C.K. Lunch, B.C. Pijanowski, J.T. Randerson, E.K. Read, A.T. Tredennick, R. Vargas, K.C. Weathers, and E.P. White. 2018: Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences 115:1424–1432.

Ditlevsen, P.D., and S.J. Johnsen. 2010: Tipping points: early warning and wishful thinking. Geophysical Research Letters 37:L19703.

Dooley, K., and L.I. Zon. 2000: Zebrafish: a model system for the study of human disease. Current Opinion in Genetics & Development 10:252–256.

Dumancas, G.G., I. Adrianto, G. Bello, and M. Dozmorov. 2017: Current developments in machine learning techniques in biological data mining. Bioinformatics and Biology Insights 11:1–4.

Edwards, S.L., J. Beesley, J.D. French, and A.M. Dunning. 2013: Beyond GWASs: illuminating the dark road from association to function. The American Journal of Human Genetics 93:779–797.

Egerton, F.N. 1975: Aristotle's population biology. Arethusa 8:307–330.

———. 2001: A history of the ecological sciences, part 2: Aristotle and Theophrastos. Bulletin of the Ecological Society of America 82:149–152.

Felsenstein, J. 2004: Inferring phylogenies. Sinauer Associates, Sunderland, MA.

Fisher, R.A. 1918: The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh 52:399–433.

———. 1930: The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Flassig, R.J., and R. Schenkendorf. 2018: Model-based design of experiments: where to go? Ninth Vienna Internatioal Conference on Mathematical Modelling:875–876.

Freedman, D.A. 1987: As others see us: A case study in path analysis. Journal of Educational and Behavioral Statistics 12:101–128.

———. 1997: From association to causation via regression. Advances in Applied Mathematics 18:59–110.

Frigg, R., and S. Hartmann. 2018: *Models in Science*. Stanford Encyclopedia of Philosophy. Downloaded from https://plato.stanford.edu/archives/sum2018/entries/models-science/%7D Metaphysics Research Lab, Stanford University.

Ghosal, S., D. Blystone, A.K. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar. 2018: An explainable deep machine vision framework for plant stress phenotyping. Proceedings of the National Academy of Sciences 115:4613–4618.

Godfrey-Smith, P. 2006: The strategy of model-based science. Biology & Philosophy 21:725–740.

Goldsby, M. 2013: The "structure" of the "strategy": Looking at the Matthewson-Weisberg trade-off and its justificatory role for the multiple-models approach. Philosophy of Science 80:862–873.

Gotthelf, A. 1999: Darwin on Aristotle. Journal of the History of Biology 32:3–30.

———. 2012: Teleology, First Principles, and Scientific Method in Aristotle's Biology. Oxford University Press, Oxford.

Gunawardena, J. 2014: Models in biology: "accurate descriptions of our pathetic thinking." BMC Biology 12:1–11.

Haldane, J.B. S., and S.D. Jayakar. 1963: Polymorphism due to selection of varying direction. Journal of Genetics 58:237–242.

Hamilton, W.D. 1967: Extraordinary sex ratios. Science 156:477–488.

Ho, P.-Y., J. Lin, and A. Amir. 2018: Modeling cell size regulation: From single-cell level statistics to molecular mechanisms and population level effects. Annual Review of Biophysics 47:251-271.

Hodgkin, A.L., and A.F. Huxley. 1952: A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of Physiology 117:500–544.

Holling, C.S. 1964: The analysis of complex population processes. The Canadian Entomologist 96:335–347.

Ioannidis, J.P. A. 2005: Why most published research findings are false. PLOS Medicine 2:e124.

Jarosz, D.F., and A.M. Dudley. 2017: Meeting report on experimental approaches to evolution and ecology using Yeast and other model systems. G3: Genes|Genomes|Genetics 7:3237–3241.

Johnson, J.B., and K.S. Omland. 2004: Model selection in ecology and evolution. Trends in Ecology & Evolution 19:101–108.

Justus, J. 2012: The elusive basis of inferential robustness. Philosophy of Science 79:795–807.

Kacser, H., and J.A. Burns. 1973: The control of flux. Symposium of the Society of Experimental Biology 27:65–104.

Kampichler, C., R. Wieland, S. Calmé, H. Weissenberger, and S. Arriaga-Weiss. 2010: Classification in conservation biology: a comparison of five machine-learning methods. Ecological Informatics 5:441–450.

Karlin, S., E.C. Cameron, and R. Chakraborty. 1983: Path analysis in genetic epidemiology: a critique. American Journal of Human Genetics 35:695–732.

Kawecki, T.J., R.E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M.C. Whitlock. 2018: Experimental evolution. Trends in Ecology & Evolution 27:547–560.

Laland, K.N., K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller. 2011: Cause and effect in biology revisited: is Mayr's proximate-ultimate dichotomy still useful? Science 334:1512–1516.

Laland, K.N., T. Uller, M.W. Feldman, K. Sterelny, G.B. Müller, A.P. Moczek, E. Jablonka, and J. Odling-Smee. 2014: Does evolutionary theory need a rethink? Yes, urgently. Nature 514:161–164.

Lee, P.S., and K.H. Lee. 2003: Escherichia coli—a model system that benefits from and contributes to the evolution of proteomics. Biotechnology and Bioengineering 84:801–814.

Lemey, P., M. Salemi, and A.-M. Vandamme. 2009: The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge.

Leroi, A.M. 2014: The Lagoon: How Aristotle Invented Science. Viking Penguin, New York.

Levins, R. 1966: The strategy of model building in population biology. American Scientist 54:421–431.

———. 1968a: Evolution in Changing Environments: Some Theoretical Explorations. Princeton University Press, Princeton.

———. 1968b: Ecological engineering: theory and technology. The Quarterly Review of Biology 43:301–305.

———. 1993: A response to Orzack and Sober: Formal analysis and the fluidity of science. Quarterly Review of Biology 68:547–555.

Lewontin, R.C. 1968: Biological models. Pp. 342–246 *in* P.P. Wiener, ed. Dictionary of the History of Ideas Volume 1. Scribner's, New York.

Lutz, F.E. 1915: Experiments with Drosophila ampelophila concerning natural selection. Bulletin of the American Museum of Natural History 34:605–624.

MacArthur, R.H. 1970: Species packing and competitive equilibrium for many species. Theoretical Population Biology 1:1–11.

Marshall, J.A., and P. Durán. 2018: Are biologists getting the mathematical training they need in college? Biochemistry and Molecular Biology Education 46:612–618.

Martínez del Rio, C. 2008: Metabolic theory or metabolic models? Trends in Ecology & Evolution 23:256–260.

Matthewson, J., and M. Weisberg. 2009: The structure of tradeoffs in model building. Synthese 170:169–190.

Mayr, E. 1961: Cause and effect in biology. Science 134:1501–1506.

Medina, M.Á. 2018: Mathematical modeling of cancer metabolism. Critical Reviews in Oncology / Hematology 124:37–40.

Meselson, M., and F.W. Stahl. 1958: The replication of DNA in Escherichia coli. Proceedings of the National Academy of Sciences 44:671–682.

Murray, B.G. J. 1991: Sir Isaac Newton and the evolution of clutch size in birds: a defense of hypothetico-deductive method in ecology and evolutionary biology. Pp. 343–180 *in* J.L. Casti and A. Karlqvist, eds. Beyond Belief: Randomness, Prediction and Explanation in Science. CRC Press, Boca Raton, Florida.

Nagelkerke, C.J. 1996: Discrete clutch sizes, local mate competition, and the evolution of precise sex allocation. Theoretical Population Biology 49:314–343.

Nesse, R.M. 2013: Tinbergen's four questions, organized: a response to Bateson and Laland. Trends in Ecology & Evolution 28:681–682.

Nijhout, H.F., J.A. Best, and M.C. Reed. 2015: Using mathematical models to understand metabolism, genes, and disease. BMC Biology 13:1–10.

Odenbaugh, J., and A. Alexandrova. 2011: Buyer beware: robustness analyses in economics and biology. Biology & Philosophy 26:757–771.

Olden, J.D., J.J. Lawler, and N.L. Poff. 2008: Machine learning methods without tears: a primer for ecologists. The Quarterly Review of Biology 83:171–193.

Orr-Weaver, T.L., J.W. Szostak, and R.J. Rothstein. 1981: Yeast transformation: a model system for the study of recombination. Proceedings of the National Academy of Sciences 78:6354–6358.

Orzack, S.H. 1990: The comparative biology of second sex ratio evolution within a natural population of a parasitic wasp, Nasonia vitripennis. Genetics 124:385–396.

———. 2005: What, if anything, is "The strategy of model building in population biology?" - A comment on Levins (1966) and Odenbaugh (2003). Philosophy of Science 72:479–485.

———. 2012: The philosophy of modelling or does the philosophy of biology have any use? Philosophical transactions of the Royal Society of London. Series B 367:170–180.

———. 2014: A commentary on "The Formal Darwinism Project": there is no grandeur in this view of life. Biology & Philosophy 29:259–270.

Orzack, S.H., and E. Sober. 1993: A critical assessment at Levins's The strategy of model building in population biology (1966). Quarterly Review of Biology 68:533–546.

———. 1994: Optimality models and the test of adaptationism. The American Naturalist 143:361–380.

Orzack, S.H., E.D. Parker, Jr, and J. Gladstone. 1991: The comparative biology of genetic variation of conditional sex ratio behavior in a parasitic wasp, Nasonia vitripennis. Genetics 127:583–599.

Palladino, P. 1991: Defining ecology: ecological theories, mathematical models, and applied biology in the 1960s and 1970s. Journal of the History of Biology 24:223–243.

Pauly, P.J. 1987: Controlling Life: Jacques Loeb & The Engineering Ideal in Biology. Oxford University Press, New York.

Pearl, J. 2009: Causality. Cambridge University Press, Cambridge.

Plutynski, A. 2006: Strategies of model building in population genetics. Philosophy of Science 73:755–764.

Rakison, D.H. 2018: Do 5-month-old infants possess an evolved detection mechanism for snakes, sharks, and rodents? Journal of Cognition and Development 19:456–476.

Rankin, C.H., C.D.O. Beck, and C.M. Chiba. 1990: Caenorhabditis elegans: a new model system for the study of learning and memory. Behavioural Brain Research 37:89–92.

Reynolds, C.W. 1987: Flocks, herds and schools: a distributed behavioral model. Computer Graphics 21:25–34.

Schmidt, M., and H. Lipson. 2009: Distilling free-form natural laws from experimental data. Science 324:81–85.

Schrider, D.R., and A.D. Kern. 2015: Inferring selective constraint from population genomic data suggests recent regulatory turnover in the human brain. Genome Biology and Evolution 7:3511–3528.

———. 2016: S/HIC: robust identification of soft and hard sweeps using machine learning. PLOS Genetics 12:e1005928.

———. 2018: Supervised machine learning for population genetics: a new paradigm. Trends in Genetics 34:301–312.

Shan, Y., D. Paull, and R.I. McKay. 2006: Machine learning of poorly predictable ecological data. Ecological Modelling 195:129–138.

Sharpless, N.E., and R.A. DePinho. 2006: Model organisms: The mighty mouse: genetically engineered mouse models in cancer drug development. Nature Reviews Drug Discovery 5:741–754.

Shipley, B. 1999: Testing causal explanations in organismal biology: causation, correlation and structural equation modelling. Oikos 86:374–382.

Shirsat, N.P., N.J. English, B. Glennon, and M. Al-Rubeai. 2015: Modelling of mammalian cell cultures. Pp. 259–326 *in* M. Al-Rubeai, ed. Animal Cell Culture. Springer, Cham, Switzerland.

Slobodkin, L.B. 1994: The connection between single species and ecosystems. Pp. 35–87 *in* D.W. Sutcliffe, ed. Water Quality and Stress Indicators in Marine and Freshwater Ecosystems: Linking Levels of Organisation (Individuals, Populations, Communities). Freshwater Biological Association, Ambleside.

Sommer, C., and D.W. Gerlich. 2013: Machine learning in cell biology–teaching computers to recognize phenotypes. Journal of Cell Science 126:5529–5539.

Spirtes, P., C. Glymour, and R. Scheines. 2000: Causation, Prediction, and Search. MIT Press, Cambridge.

Strogatz, S.H. 2018: Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering. CRC Press, Boca Raton, Florida.

Sturtevant, A.H. 1915: Experiments on sex recognition and the problem of sexual selection in Drosoophilia. Journal of Animal Behavior 5:351–366.

Szostak, J.W., T.L. Orr-Weaver, R.J. Rothstein, and F.W. Stahl. 1983: The double-strand-break repair model for recombination. Cell 33:25–35.

Sztul, E. 2017: How can biological modeling help cell biology? Cellular Logistics 4:e1404780.

Tarca, A.L., V.J. Carey, X. Chen, R. Romero, and S. Drăghici. 2007: Machine Learning and Its Applications to Biology. PLOS Computational Biology 3:e116.

Tcherkez, G.G. B., C. Bathellier, H. Stuart-Williams, S. Whitney, E. Gout, R. Bligny, M. Badger, and G.D. Farquhar. 2013: D2O solvent isotope effects suggest uniform energy barriers in Ribulose-1,5-bisphosphate Carboxylase/Oxygenase catalysis. Biochemistry 52:869–877.

Telenti, A., L.C. T. Pierce, W.H. Biggs, J. di Iulio, E.H. M. Wong, M.M. Fabani, E.F. Kirkness, A. Moustafa, N. Shah, C. Xie, S.C. Brewerton, N. Bulsara, C. Garner, G. Metzker, E. Sandoval, B.A. Perkins, F.J. Och, Y. Turpaz, and J.C. Venter. 2016: Deep sequencing of 10,000 human genomes. Proceedings of the National Academy of Sciences 113:11901–11906.

Tickoo, S., and S. Russell. 2002: Drosophila melanogaster as a model system for drug discovery and pathway screening. Current Opinion in Pharmacology 2:555–560.

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

Tipton, J.A. 2008: Aristotle's observations of the foraging interactions of the red mullet (Mullidae: Mullus spp) and sea bream (Sparidae: Diplodus spp). Archives of Natural History 35:164–171.

Torres, N. V, and G. Santos. 2015: The (mathematical) modelling process in biosciences. Frontiers in Genetics 6:1–9.

Torres, N. V, F. Mateo, E. Melendez-Hevia, and H. Kacser. 1986: Kinetics of metabolic pathways. A system in vitro to study the control of flux. Biochemical Journal 234:169–174.

Touw, W.G., J.R. Bayjanov, L. Overmars, L. Backus, J. Boekhorst, M. Wels, and S.A. F.T. van Hijum. 2012: Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? Briefings in Bioinformatics 14:315–326.

Venter, J.C., and D. Cohen. 2004: The Century of Biology. New Perspectives Quarterly 21:73–77.

Wallmeier, H. 2018: Quantitative biology-a perspective for the life sciences' way into the future. Journal of Business Chemistry 15:78–92.

Weisberg, M. 2006a: Forty years of 'The strategy': Levins on model building and idealization. Biology & Philosophy 21:623–645.

———. 2006b: Robustness analysis. Philosophy of Science 73:730–742.

———. 2007: Three kinds of idealization. The Journal of Philosophy 104:639–659.

———. 2013: Simulation and Similarity: Using Models to Understand the World. Oxford University Press, Oxford.

Weisberg, M., and K. Reisman. 2008: The robust Volterra principle. Philosophy of Science 75:106–131.

Wernick, M.N., Y. Yang, J.G. Brankov, G. Yourganov, and S.C. Strother. 2010: Machine learning in medical imaging. IEEE Signal Processing Magazine 27:25–38.

Wimsatt, W.C. 1981: Robustness, reliability and overdetermination. Pp. 324–163. *in* R. Brewer and B. Collins, eds. Scientific Inquiry and the Social Sciences. Jossey-Bass, San Francisco.

———. 1987: False models as means to truer theories. Pp. 33–55 *in* M.N. Nitecki and A. Hoffman, eds. Neutral Models in Biology. Oxford University Press, Oxford.

Woodward, A.W., and B. Bartel. 2018: Biology in bloom: A primer on the Arabidopsis thaliana model system. Genetics 208:1337–1349.

Wray, G.A., H.E. Hoekstra, D.J. Futuyma, R.E. Lenski, T.F.C. Mackay, D. Schluter, and J.E. Strassmann. 2014: Does evolutionary theory need a rethink? No, all is well. Nature 514:161–164..

Wright, S. 1931: Evolution in Mendelian populations. Genetics 16:97–159.

Steven Hecht Orzack
Fresh Pond Research Institute
Cambridge, MA 02140
orzack@freshpond.org

Brian McLoone
National Research University
Higher School of Economics
Moscow, Russia
and
Fresh Pond Research Institute
Cambridge, MA 02140
brianbmcloone@gmail.com

STUDIA METODOLOGICZNE
NR 39 • 2019, 99–121
DOI: 10.14746/sm.2019.39.4

Piotr Swistak

The Curious Case of Formal Theory in Political Science: How Did It Emerge as a Discipline, Why Its Existence Is a Sign of a Failure, and Why the Science of Politics Is Not Possible Without It?

Abstract. American political science has evolved a subfield which is commonly referred to as *formal theory.* Political scientists identify themselves as specializing in formal theory, departments advertise faculty positions in formal theory and put together formal theory subfields that offer undergraduate and graduate curricula. The roots of the field can be traced to Thomas Hobbes. Hobbes' message, however, seems to have been utterly ignored by the social science. William Riker's second launch of "Hobbesian advice", in 1950's and 60's, proved more successful and put the field of formal theory on the map of political science. Yet, the very existence of the formal theory field can be seen as the failure of both Hobbes and Riker. There seems to be a continuing need for teaching social scientists why they should construct a proper science and how they should do it. This paper is an attempt to meet this need. I believe that the future science of politics will have to follow in the footsteps of Hobbes and Riker. And so will other social sciences. My point in the paper is not new; the way I make it, is.

Keywords: formal theory, axiomatic theory, formal theory field in political science.

1. Introduction

The "formal theory" label used by political scientists has always struck me as odd. Possibly because I could not imagine a field of "formal theory" evolving, for instance, in physics. In fact, I do not know of any other discipline of science in which a form of a theory became a label for a field of inquiry. Clearly, the very use of a qualifier “formal” implies that such theories are special, possibly rare, and that the common understanding and use of the notion of theory is different. For anyone interested in the general methodology of science, this singularity of political science may be interesting to note and to reflect on. Hence, one of my objectives is to bring this case to the attention of those who may find it worthwhile from the perspective of philosophy of science or general methodology.

My other objective is to provide an easy to read text on (formal) theory construction for social scientists. In my experience there is a widespread need for a basic reflection on the form and the benefits of a properly constructed theory in all social sciences with a possible exception of economics.[1] The absence of short and simple texts on the subject reinforces this problem. My goal is to reach those who would want to learn on their own as well as those who would want their students to better understand why this form of theory construction is important to use and necessary to accept as an integral part of any scientific endeavor.

My third, and last, objective is to engage yet a different group of scholars. Many philosophers and mathematicians will find the idea of writing about theory construction as antiquated and trivial. This is understandable. Yet, many in this group work within well formulated theories and rarely, if ever, face a problem of theory construction. Even fewer would ever deal with a case in which they observe an empirical phenomenon and have to formalize it from scratch. This is a very different problem than working with deductions and one that can be sufficiently difficult and fascinating to attract mathematicians like John von Neumann to the social sciences. Still, I know of a good number of mathematicians and other scientists who seem to dismiss social science as a domain where any interesting deductive work has been or can be done. Famously, Stanisław

[1] In my experience it is very common for students graduating with a political science degree not to have seen a formal theory nor to have reflected on the benefits of having one. This is less frequent, though still not uncommon, among graduate students and faculty members in political science. The same can be said about sociologists, psychologists and anthropologists. Economics is the only exception since it routinely exposes students to formal theories, like preference theory and expected utility theory, that are fundamental to modern economic theory. These claims derive from my many interactions with students and faculty from all disciplines of social science associated with dozens of universities worldwide. Barring a close-to-zero probability of being exposed to a very skewed sample, my observation should have a dose of generality to it.

Ulam has once challenged Paul Samuelson (Samuelson 1969, p.9) to name "one proposition in all of the social sciences which is both true and non-trivial." Some thirty years later Samuelson provided an answer to this challenge.[2] We are less likely, perhaps, to witness such a challenge today, since the underlying attitude towards economics has changed, though I think a challenge of naming a non-trivial theory in political science, or sociology, is as likely to come from a mathematician today as it was back when Ulam has posed his challenge to Samuelson. For those who cannot name a "true and non-trivial proposition" in political science I would like to offer a few very simple examples.

In short, regardless of the reader's academic discipline this paper may have something useful to offer to everyone. The value for political scientists may be in realizing the significance of the proper formulation of a problem. I hope they will see that "formal theory" is more than a subfield of political science, or a short-lived fad, it is rather the only proper way to go about theory construction if we ever hope to turn political science into a cumulative body of knowledge. This point should also be useful for other social scientists including economists some of whom get frustrated with dismal predictive power of "formal" theories in economics and turn, unwisely so, against the very scientific method used to construct these theories. Some mathematicians may take my bait and choose to look up my answers to the Ulam challenge. Perhaps they will start constructing, and solving, their own puzzles of human behavior. Finally, anyone with interest in the philosophy of science, or general methodology, may find the very existence of the formal field in political science, and some historical details behind it, to be an interesting and useful anecdote with some important lessons to learn or to teach.

I begin the paper by recalling Hobbes' advice from *Leviathan* that, ironically, was used by Wilder, an influential mathematician in 1950s, to open a chapter on axiomatic theory. The case of a prominent mathematician who brings up Hobbes whose advice was utterly ignored by political science all the while

[2] Samuelson later wrote about Ulam's testing question: "This was a test I always failed. By now, some thirty years later (…) an appropriate answer occurs to me: The Ricardian theory of comparative advantage." (Samuelson 1969, p.9)

Leviathan remained a required reading for all in the discipline[3] makes for an anecdote that was difficult for me to pass. More importantly, though, I find it important for the social scientists to remember that the proper form of a theory has been argued for and pointed to by Hobbes centuries before it started being used in the modern social science. Someday, perhaps, Hobbes' main legacy would be seen as that of a protagonist of a scientific social science. Possibly this was also on Wilder's mind when he tuned to Hobbes for an opening quote.

Hobbes' advice on theory construction fell on deaf ears among social scientists and it took some three hundred years to change that. The change came through some major developments in economics; in political science it was largely driven by the efforts of William Riker and the so called Rochester School. I will later describe the intellectual roots and the significance of the Rochester School for modern political science. While Riker and his students have effectively changed modern political science the very existence of the "formal theory" field proves that the main massage of Hobbes and the Rochester School is still lost on the discipline. The day this message settles in there will be no need to single out formal theory lest make it a special field of political science. While nearly all political scientists have heard of Prisoners' Dilemma and many can write it down and explain, only a handful would fully understand the details of the model including the properties of payoffs.[4] People use products of a proper theory construction but do not understand their basic properties. The very purpose of a proper form of science is completely lost on them. For this reason, in the last part of the paper I will go back to Hobbes' point one more time. To make the point obvious I will use some of the simplest examples I could have

[3] Hobbes' *Leviathan* is one of the most important works in the history of social science and one of fundamental importance for political science. The reason, however, is the importance of Hobbes' explanation of the emergence of state rather than his singular effort to follow Euclid's theory construction. A short comment preceding excerpts from *Leviathan* in the *Princeton Readings in Political Thought* (Cohen 1996) are typical in that regard.

[4] I can name scores in the discipline who thought that using identical payoffs for different players means that we assume that they have identical utilities and requires interpersonal utility comparisons, or those who thought that negative payoffs signify displeasure while positive pleasure, or finally those whom I have asked to tell me what is the difference between ordinal and cardinal payoff only to find out they have never heard of either. Some people in this set are prominent scholars, many use games in their research. This, by no measure, can indicate a success of Hobbes or Riker.

thought about. Even if I fail to convince anyone to change the ways they practice social science I hope to be convincing enough to make them see that the standard practice of social science cannot possibly be a long term equilibrium. It may take decades or centuries but at some point Hobbes will not only get his message through but will effectively change the social science.

2. The message of Hobbes that social science chose to ignore

Perhaps the most striking aspect of formal theories in the social sciences is their ubiquitous absence rather than occasional presence. For most social scientists a theory corresponds, roughly, to a statement "*x* affects *y*" like in "the level of education affects the level of income." The relation between *x* and *y* would be typically assumed to be of the form $y = bx + a$, and a and b are parameters estimated from the data. The more "complex" theories would assume that *y* is affected by more variables x, x', x'' etc, albeit in the same linear fashion: We won't see anyone conjecture $y = \int x^3 e^{-2^{(x-3)}} dx$, for instance, since how could we possibly come up with this functional form if we don't have a set of more fundamental principles, a theory, to derive it from? Most statistical tools require that we specify a function, or a class of functions, before we do any data analysis. The functions used in such statistical analyses are commonly understood as theories. Linear functions of the general form have been, by far, the most common type of "theories" generated in the social science. This literally is what most social scientists understand by "theory." It is embarrassingly uncommon for them to think of a theory the way their colleagues in science think about theories, that is, as built of undefined terms (the elementary building blocks), assumptions (axioms) defining their properties, and deductive consequences (theorems) of these assumptions. For this reason, I find it perversely amusing to see Raymond Wilder, a leading mathematician of his days, opening an essay on theory construction with a quote from Thomas Hobbes' "Leviathan" originally published in 1651:

> "The errors of definitions multiply themselves according as the reckoning proceeds; and lead men into absurdities, which at last they see but cannot avoid, without reckoning anew from the beginning."[5]

[5] Hobbes (1651, 1981 edition).

A pure mathematician using a quote from a father of political science and one of the great classics of the social science to tag an explanation of the modern form of scientific theory presents an awkward dilemma for the social science: Has Hobbes' advice, realized with a great success by mathematics and hard science, been blatantly ignored by most of the mainstream social theory for centuries, and all that while "Leviathan" has never stopped to be one of the most widely assigned and read pieces of social science? Well, yes. And it puts social sciences in an embarrassing situation, especially that Hobbes' advice is hardly controversial.

Hobbes' point, to put it in the crudest way, is that science[6] cannot be built on ill defined, or non-defined, concepts. This sounds rather obvious. If we make inferences using concepts that we don't understand—and how can we understand something that is ill defined? —we can only make a bad problem worse.[7] For sure we won't be able to explain these concepts to others, which is to say that if we don't know what we are talking about, it would be unreasonable to expect someone else to understand it. The first problem leads to absurd conclusions, the second prevents an accumulation of knowledge; either and both make the development of science impossible. This is a general observation and one that should apply to any endeavor aspiring to be called "science."

Hobbes might have felt uneasy about preaching the obvious as well. After all, he was merely repeating points that have been made twenty centuries ago by Aristotle and then implemented by Euclid in "Elements." "Elements" were, in fact, Hobbes' blueprint for constructing a scientific argument, so much so that he called it "the only science that it hath pleased God hitherto to bestow on mankind."[8] Hobbes organization of "Leviathan" is meant to

[6] I use the term "science" in its most general meaning, not one that merely restricts it to "physical science."

[7] In Hobbes' words "(…) a man that seeketh precise truth, had need to remember what every name he uses stands for; and to place it accordingly; or else he will find himself entangled in words, as a bird in lime-twigss; the more he struggles, the more belimed."

[8] Indeed, Euclid's "Elements" have been the only axiomatic theory in existence for many centuries. The modern form of axiomatic theory is typically credited to Hilbert and his 1899 axiomatization of Euclidean geometry. For a fascinating account of the history Hobbes accidental inspiration by "Elements" see Macpherson (1985.)

follow the construction of "Elements."[9] In fact, "Of Man," the first book of "Leviathan," is not only an implementation of a properly constructed theory but it also constitutes a proof that science—and indeed any reasonable form of human communication—cannot be practiced in any other way. Hobbes' implementation of Euclid's blueprint was so good that some three and a half centuries later "Leviathan" still stands out in the social sciences—if we clean up the antiquated language and shrink the long sentences—as an example of clarity and rigor. This is as much a praise of Hobbes as it is a criticism of the social sciences.

The fact that a scientific theory has to be constructed in a proper way was obvious to Hobbes and by 1950's it was obvious to anyone working in mathematics and in physical sciences and to a growing number of economists especially those with background in mathematics or in physics[10]. This must have made Raymond Wilder[11] conclude in 1952 that this form of theory construction "is acknowledged now as a fundamental part of the scientific method in every realm of human endeavor" (p.1622.) Wilder has been wrong when he wrote these words and he would have been wrong today. Most social scientists practice "science" that lacks "a fundamental part of the scientific method."

Wilder was right on a different count though. The tide was about to turn. In 1950's the change has already been sweeping economics and soon it was about to hit political science.

[9] It has been pointed to me by one of the reviewers that Hobbes had contacts with Descartes and Galileo, hence Euclid was hardly the only source of his convictions. My objective here is not, however, to provide a comprehensive historical account.

[10] By now to be accepted into a major graduate program in economics, an undergraduate degree in mathematics, or a coursework that comes close to getting one, is almost a prerequisite.

[11] Raymond Louis Wilder (1896-1982) was a leading mathematician of his times. His 1952 "Introduction to the Foundations of Mathematics" includes a chapter on theory construction, "The Axiomatic Method." It was later reprinted as an independent essay in "The World of Mathematics." More than half a century after its first publication "The Axiomatic Method" remains an excellent introductory writing on the subject. (Wilder spent most of his academic career in the Department of Mathematics at the University of Michigan, Ann Arbor; he was a member of the National Academy of Sciences and served as the President of the American Mathematical Society and the Mathematical Association of America.)

3. William Riker and the emergence of the formal theory field in political science

While the early uses of formal theory in political science can be traced to Black (1948), Downs (1957), Shapley and Shubik (1954) and a few others, there is no doubt that a subfield of political science that goes under the name of formal theory, rational choice theory or positive political theory (these names are often seen as interchangeable) is a work of one man. His name is William Riker. For anyone who witnessed the emergence and the growth of this field, "formal theory" feels largely synonymous with William Riker. Riker, in a sense, succeeded at implementing Hobbes' advice for political science.

Critical for Riker's success was the timing of his efforts. By 1950's mathematicians and logicians have refined the notion of a formal theory, which in these disciplines is better known as an axiomatic theory, and it became a standard of theory construction for an increasing number of scientists including a growing number of economists. In 1944 game theory has been launched by von Neumann and Morgenstern's "Theory of Games and Economic Behavior." It quickly became a fad among mathematicians and a small, but influential, group of social scientists. John Nash and Lloyd Shapley were among the earliest followers of the fad. In 1950's they were both graduate students in mathematics at Princeton and they both chose to work on game theory under Albert Tucker. Their doctoral theses were to lay the foundations of non-cooperative and cooperative game theory. (Their solution concepts are known as Nash equilibrium and Shapley value. John Nash and Lloyd Shapley were both awarded Nobel Prizes for these contributions.) About the same time Albert Tucker, came up with a name and an interpretation of a game that was brought to his attention by Merrill Flood and Melvin Dresher.[12] The name he proposed was Prisoner's Dilemma. Prisoner's Dilemma went on to become the single most important and most influential game in game theory. Finally, 1951 saw the publication of Kenneth Arrow's "Social Choice and Individual Values" which contained what is now known as Arrow's Theorem. The result

[12] This story has been recounted in many writings. One of the more recent ones is (Holt 2004).

gave rise to social choice theory and became an important factor in awarding Arrow a Nobel Prize. These developments have proved to be transformative for all social sciences but they were particularly consequential for the science of politics. At a closer look the core of most of them was par excellence political.

Game theory is a general theory of interactions and as such it is relevant for all social sciences. Of the two branches of game theory, cooperative and non-cooperative, cooperative game theory was created as a theory of coalition formation. Coalition formation is, of course, the domain of political science which made cooperative game theory impossible to ignore by political scientists. The relevance of non-cooperative games soon became equally obvious. The Prisoner's Dilemma hit at the very core of political science by providing a general model for a class of problems that require solution by government. The game's equilibrium came to be seen as the argument for the necessity of state and a proof that political science cannot be subsumed by economics. Jon Elster made this point explicitly clear by defining politics as "the study of ways of transcending the Prisoner's Dilemma." (Elster 1976, p.249) In this sense Prisoner's Dilemma became a label for the very essence of political science.

Another transformative idea came from a different line of research. Central to this research was the work of Kenneth Arrow. Arrow (Arrow 1951, 2nd ed. 1963) set out to formally define the core properties of a democratic choice and proved that they cannot all be satisfied. One way to think about Arrow's result is as a formal theory of any voting rule that we would want to use in elections. Before we start arguing if we want to use a simple majority rule, or a plurality rule or approval voting or, perhaps, the American electoral system, we would want to sit down and specify basic properties of any acceptable democratic voting rule. For example, we would want a rule that treats all voters equally (all votes count the same) and one which when used in a two-candidate election where every voter prefers one candidate over the other, would name the preferred candidate the winner. The way I always explain Arrow's Theorem is by stating such—overtly obvious—properties of a democratic choice and ask students to think of them as articles of a constitution. I have yet to find a single individual who would not support it. As it turns out, democratic rule of choice, as defined by the articles our constitution, is a concept with an empty connotation. Arrow has proved that "articles" of this "constitution" are inconsistent. This result, known as Arrow's Theorem, is considered by many to be the most important finding in the modern political science. This is one of the best lessons Hobbes could have hoped for. If Arrow did not construct an axiomatic (formal) theory of choice, we would have never known that what we may want in democracy is not possible to have.

It is important for me to note that my description of the developments of 1940's and 50's is not indicative of how they were seen, or recognized, in political science. All these results came from mathematics and/or economics and very few political scientists knew about them or had sufficient mathematical training to understand this work.[13] Riker who had considerable training in economics was an exception. Not only was he familiar with this research but he clearly understood its importance for political science. He was also aware that the timing to change political science could not have been better and through his own research he began to advance the new paradigm (Riker, 1963, 1982; Riker and Ordeshook, 1973). The scientific form of this paradigm, widely seen as employing formal theory construction, became central to Riker's brand of political science. When in 1960's the University of Rochester was looking for faculty to build analytically rigorous social science departments, Riker was an obvious choice for political science. Backed by a very generous support of the university Riker built a department that made Rochester a famous brand and began a transformation that created the field of political science commonly known as formal theory.

Riker's students, students of Riker's students, and countless followers populate now major departments of political science offering curricula in formal theory at undergraduate and graduate levels. Political science seems to have undergone a permanent and irreversible change. And yet, I am not sure if the scope and the pace of the change would have been seen as satisfactory by Hobbes, or by Wilder.

The bulk of political science still does not use the proper theory construction. The very use of the "formal theory" label suggests that most "theories" are not constructed that way; to put it bluntly, it suggests that most of political

[13] Private communication with William Riker and Peter Orheshook.

theory is not really scientific. The inertia of the discipline and the resistance to change have proved to be considerable. For this reason, papers like this one will be written for a foreseeable future and arguments will be given for and against the use of formal theory construction in social science. The day when symbolic notation, like the one used in the formulation of preference theory, will cover blackboards of classes in political science the way they do now in physics, is still very distant.

For this reason, anyone teaching formal theory in political science finds it imperative to explain to students why formal political science looks so much different than what they see elsewhere in the discipline. I often envy physicists who don't have to justify the use of mathematics or other properly formalized constructs since everyone in physics knows that this is the proper form of science. In the next two sections I offer two arguments which I often use to convince students of politics about the necessity of the proper theory construction.

What I have consistently found to work best as convincing explanations are arguments that use short and simple examples. Each of the next two section is built around one such example.

4. An example that explains the necessity of formal political science

Sometime during the winter break of 2006/07 Nicholas Grossman, who was then a graduate student at the University of Maryland, stopped by my office to tell me about a peculiar finding he had overheard on the radio when driving from Connecticut to Maryland. The data concerned a study involving voters' preferences over three Democratic front runners, John Edwards, Barack Obama and Hilary Clinton, who were to compete for the Democratic nomination in the US presidential election of 2008. The puzzle behind the peculiar results can be shortly described as follows.

Subjects in the study were asked to consider two candidates, Edwards and Clinton, for instance, and pick the one they would vote for. They were then asked to do the same for Edwards and Obama, and Clinton and Obama. In yet another question they were asked to vote for one from the set of Edwards, Obama and Clinton. Interestingly, Edwards was picked by a substantial majority against Clinton and against Obama and yet when subjects were picking one out of three, Edwards ended up with the lowest percentage of votes. This has struck Grossman as an interesting, and a possibly odd, result. Indeed, the data seems counterintuitive enough to warrant a closer examination.

To make the puzzle more explicit let's conjecture specific distributions of choices. Assume, for instance, that Edwards was chosen by 60% over Clinton and by 60% over Obama; for completeness, let's also assume than 60% chose Clinton over Obama. Suppose, moreover, that when asked to pick one out of three, 38% chose Clinton, 32% Obama, and 30% Edwards. What, if anything, might be wrong with this outcome?

If, indeed, there is something wrong with the assumed distributions, the problem must originate in subjects' preferences over the candidates. If, for instance, a subject does not really know any of the three candidates but feels compelled to answer the questions, he may well make his choices at random. But choosing at random means that a person who chose Edwards over Clinton may choose Clinton over Edwards when asked about his preferences again. His choice in one instance is unrelated to his choice in the other. Is it an effect of such random choices that strikes us as "odd" in the data? Do we, indeed, intuitively expect some "consistency of choices" and this consistency is violated by the distribution of votes cast by the subjects?

If the lack of consistency of choices constitutes the essence of the puzzle, then the solution would require that we define the conditions of consistency and then prove that they are violated by the data.

And so, let's first consider the properties of choices that would agree with our intuition of consistency. If, for instance, a decision maker chooses x over y, basic consistency, as I have already noted, would require that he does not choose y over x when we ask him again. Another similarly straightforward aspect of consistency would have a preference relation behave similar to a "greater than" relation on numbers: if a subject prefers x over y and y over z then he should also prefer x over z. Yet another set of intuitions would apply to cases in which a decision maker is indifferent between alternatives. Here our intuition would suggest that indifference is a form of equivalence. Thus, for instance, a "consistent" indifference relation should not be affected by the order in which alternatives are presented: if a person is indifferent between x and y, he should also be indifferent between y and x. In addition, anyone who is indifferent between x and y and y and z should also be indifferent between x and z. One final condition would assume that subjects know the alternatives and, hence, can choose between them. More specifically, we will assume that when a decision maker compares two alternatives he would either prefer one over the other or be indifferent between them.

If we agree to define our intuitions the way I have suggested, we are all set to construct a "formal theory" of individual decision making. In the formulation I will use the following notation: $D = \{x, y, z, \ldots\}$ denotes a finite set of alternatives; $\prec$ denotes a binary preference relation (strict), $\sim$ denotes a binary indifference relation, $\neg$ denotes "it is not true that," $\forall$ denotes "for all," $\exists$ denotes "there exists," $\in$ denotes "belongs to," $\Rightarrow$ denotes "implies," $\vee$ denotes "or," and & denotes "and." A formal system $\langle D, \prec, \sim \rangle$ which satisfies the following five axioms is called *theory of preferences*.

A1. (COMPARABILITY) $\forall x,y \in D: (x \prec y \vee y \prec x \vee x \sim y)$.
A2. (ASYMMETRY OF $\prec$) $\forall x,y \in D: (x \prec y \Rightarrow \neg (y \prec x))$.
A3. (TRANSITIVITY OF $\prec$) $\forall x,y,z \in D: ((x \prec y \,\&\, y \prec z) \Rightarrow x \prec z)$.
A4. (SYMMETRY OF $\sim$) $\forall x,y \in D: (x \sim y \Rightarrow y \sim x)$.
A5. (TRANSITIVITY OF $\sim$) $\forall x,y,z \in D: ((x \sim y \,\&\, y \sim z) \Rightarrow x \sim z)$.

The theory we have just formulated, known as theory of preferences, constitutes the most general concept of rationality in the social sciences. For this reason alone, it may well be the most important formal theory of all. But there is another important reason for its significance. A properly constructed theory of choice made it possible for the science of choice to grow. The expected utility theory, formulated by von Neumann and Morgenstern (1945), is a generalization of the preference theory to the domain of choices that involve lotteries, i.e., probability distributions over alternatives in the domain. Game theory, finally, is a generalization of theories of individual decision making, i.e., theory of preferences and the expected utility theory, to cases in which the outcome is affected not only by the action of a decision maker but also by the actions of others. These connections are important to note because without a proper formulation of a theory of preferences the path that links it with game theory would not have been possible. For any science to progress, it must be cumulative and this cannot be done outside of properly constructed theories.

Now that we have defined a consistent choice behavior, better known as rational choice, we can return to the data and ask if there is any evidence that any of the axioms has been violated. Given the design of the study the only axiom that can be violated is the transitivity of the preference relation. Does, then, 60% choosing Edwards over Clinton, 60% choosing Edwards over Obama, 60% choosing Clinton over Obama, and 38% choosing Clinton, 32% Obama and 30% Edwards when asked to pick one out of three indicate that some of the subjects must have had intransitive preferences?

The answer is given by the following example. Consider a group with the following (transitive) preferences over Edwards (E), Clinton (C) and Obama (O): 8% prefer E over O over C or EOC, in short, 22% ECO, 30% CEO, 8 COE, 30% OEC and 2% OCE. It is easy to check that this distribution of preferences results in the conjectured data. And so, the intuitive feeling that something must be wrong with Edwards winning by considerable majority over Clinton and over Obama and then coming last when voters were to pick one out of three candidates was plain wrong.

This, of course, would not come as a surprise to anyone working with deductions. After all the point of any deductive science, i.e., a formal theory as a political scientist would say, "(…) is to start with something so simple as not to seem worth stating, and to end with something so paradoxical that no one will believe it." (Russell 2009) Implicit in Bertrand Russell's words is an advice that working with proper science we are better off putting our intuitions aside since what we are heading for may be negatively correlated with them. This seems to be the case with the example of the voting study.

Another lesson to be learned from the example is that formal theory, as the proper form of science, is necessary simply because human brain cannot properly process tasks of certain complexity. Just like we do not attempt to multiply two five digit numbers in our head we should not attempt to make scientific inferences outside of a properly constructed formal theory. As Richard Feynman has notably put it: "Science is a way of trying not to fool yourself.

The first principle is that you must not fool yourself, and you are the easiest person to fool." (Coyne, 2009)[14] While we widely recognize and acknowledge our arithmetical deficiencies and wisely use calculators when multiplying five digit numbers, we are not nearly as humble when it comes to our ability of practicing social science. If we were, the formalization of the social science would have started with Hobbes.

5. A trap for a barefoot empiricist

Barefoot empiricism is a common accusation used against many social scientists. How can we, in a simple way, see that sheer data analysis may not make sense without a theoretical reflection? It would be good to see it using the simplest theory possible. If sheer data analysis without any analytical work turns out to be unreasonable in the simplest instance then, obviously, it can only get worse as we move to more complex cases.

Imagine that John, a student of international relations, has been studying networks of military alliances. A pure empiricist, John gave no thought to theoretical properties of such networks and set out to search for interesting empirical relations in the data. The data spanned many decades and for each point in time and every two countries it could tell John whether the pair was in military alliance or not. For exploratory purposes John picked an arbitrary year and started looking at the data. To aid his exploration he would occasionally draw a graph, like these in Figure 1, with countries represented by vertices and alliance relation by edges; absence of an edge between two vertices means that the two countries were not allied.

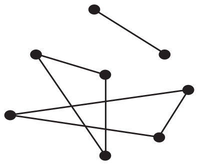


Figure 1

[14] I want to thank Jon Bendor for bringing this quote to my attention.

Imagine now two unrelated scenarios of what could have been the outcome of John's analyses.

Situation 1. After a careful examination of alliances in all groups of countries, and at all points in time, John discovered a remarkable regularity: It was always the case that if *a* was allied with *b* and *b* was allied with *c* then *a* was also allied with *c*. Since this property is called transitivity in the jargon of "relational properties," we can say that John has discovered ubiquitous transitivity of the alliance relation. This regularity has struck him as one of obvious significance and since no one else had noticed it before, John felt that he made an important discovery.

Now, let's reverse the time, assume that Situation 1 has never happened and John did not notice transitivity simply because there was no transitivity in the data. Situation 1 aside, however, the data might have held other interesting regularities for John to observe. The next situation describes one such scenario.

Situation 2. Suppose that John, like many scholars of international relations, was interested in the concept of power. One possible index of country's power is its number of allies. Since more allies makes one more powerful (other factors equal), John started examining patterns in the number of allies. One thought that has crossed John's mind was to look for cases where every country has a different number of allies. In such cases countries would form a strict "power hierarchy" with the country with most allies at the top and the country with the least allies at the bottom. Interestingly, though, John has failed to find them. At all points in time and in all possible subsets of countries there were always at least two countries with the same number of allies. There has never been a set of countries with a strictly hierarchical order that John was looking for—not a single case like that across the recorded history! This observation has struck him as interesting and significant. Since no one else had discovered it before, John felt that he has made an important discovery.

Let's now revisit the two scenarios. Since all that John has done is data analysis, we now need to do some thinking for John. Let's begin with Situation 1 and the case of transitivity. If we consider all theoretically possible graphs of alliances it is obvious that some graphs have a transitive alliance relation, say the graph in Figure 1, but other graphs may not. This means that the transitivity of the alliance relation discovered by John was a genuine empirical finding.

A philosopher would say that a proposition "In all networks of alliances the alliance relation is transitive" is a synthetic proposition—one that may be either true or false depending on what we find in the data. In Situation 1 the data have rendered this proposition true.

Now, let's turn to Situation 2 and, by analogy, ask if the proposition "In all networks of alliances there are at least two countries with the same number of allies" is a synthetic proposition as well. Is it so that in some cases the proposition is true but in some it is false? If this proposition is synthetic then John's discovery, like the discovery of transitivity, would mean a genuine empirical finding. But what if it is not a synthetic proposition? In that case, given the definition of a synthetic proposition this proposition must always be true. But if it is always true, regardless of the data, then it must follow deductively from the structure of the data. Does it?

Clearly, to move ahead with the problem we need to realize the structure of the data—i.e., the underlying formal theory—and try to figure out what follows from its axioms. Fortunately, the theory is trivial, even from a barefoot empiricist's point of view and may well be the simplest theory there is. It consists of a finite set of objects, countries in our interpretation, and a single binary relation, the alliance relation α, on this set. The only assumption we make is the symmetry of this relation: for any *a* and *b*, if *a* is an ally of *b* then *b* is an ally of *a*. Our formal theory is a system $\langle D, \alpha \rangle$ with a single axiom:

A1. (SYMMETRY OF α) $\forall a,b \in D: (a \alpha b \Rightarrow b \alpha a)$.

This, of course, is a well-known theory in mathematics, a subfield of graph theory which is called a theory of undirected graphs. What is important for John, however, is what follows deductively from this simple axiom. And what follows is that his empirical observation was not "empirical" at all. No matter what system of alliances you take there will always be at least two countries with the same number of allies.

To prove this proposition let's suppose, by contradiction, that there is a system with n different countries (n>2) where all countries have different number of allies. The only possible number of allies any country can have is one of the n numbers: 0, 1, …. , n-1. Thus if all countries have different number of allies, then one of them must have 0 allies, another one 1 ally, another one 2 allies, ..., and yet another one n-1 allies. But if a country has n-1 allies, it must be allied with all other countries in the system, and this means that the system cannot contain a country with 0 allies. This, however, contradicts the fact that all countries have different number of allies, and thus proves the proposition.

What John has thought to be an empirical discovery, turns out not to be an empirical discovery at all—it is a deductive consequence of the way he has defined the problem. And so the proposition "In all networks of alliances there are at least two countries with the same number of allies" is not a synthetic proposition; it is not an empirical regularity at all. Philosophers call such propositions analytic. This means that if we properly specify the concepts and the assumptions we use—construct a proper theory, that is—then this proposition follows from the axioms.

Whenever we deal with properly defined concepts, there will always be empirical findings whose status we won't know if we don't check if the observed regularity does not follow from the set of assumptions that define these concepts. In other words, unreflective empiricism is only safe when a theory is trivial: either all deductive consequences of the assumptions we make are obvious or there are no deductive consequences at all. Whenever we venture into the simplest of nontrivial theories, like the theory of alliances, we would not know if our empirical discovery is genuine or not unless we first show that it is not a deductive consequence of theory's assumptions. Atheoretical empiricism only makes sense in a world that has a transparently trivial structure.

Perhaps my example is too simple to fool everyone though I know for a fact that it works on great many people. It shows that our brains are not good with deductions. Even if we realize the form of the theory we are working with we still cannot see some very simple consequences of the assumptions we make. But if our logical skills are dismal then the very idea of pursuing inferences without properly specifying all assumptions needed—i.e., without having a formal theory—makes about as much sense as trying to execute an addition without knowing what numbers we are adding. And so, my last example goes to the core of the problem: How good are our logical powers?

6. The dismal shape of our scientific backbone

Since all valid inferences have to use logic, logic is our scientific backbone. I often come across people who think that logic is an abstract construction of philosophers and mathematicians and as such belongs with other esoteric topics that are of interest to no one but academics. They are wrong. First and foremost, logic is the way we (should) make inferences, not just in science, but in everything we do.

To see logic at work it is enough to recall a visit to a supermarket and reflect on one of the shoppers in front of us whose train of thought must have gone somewhat like this:

> "Oh, good, the cashier has scanned all my products. Why does he keep staring at me? Oh, I know. I have to pay now. OK, let's pay. How can I pay? Well, by cash, credit card, or a check. So, where is my money? It must be in a purse which is in my backpack. Let's get it and pay cash. How much do I have to pay? All right, let's see how much cash is in the purse. I have a ten, and a five, and some singles. No, I don't have eighty-four dollars. All right, let's use a credit card then. But, where are my credit cards? They can be in my purse. Are they? Let's check. No, they are not. Well, they must be in my wallet then. But my wallet can only be in my purse or in the back pocket of my pants. Since it is not in my purse, it must be in the pocket. Oh yes, it is. (…)"

In science people make inferences the same way, albeit faster.

If we can watch logic at work in a supermarket, logic must clearly be the backbone of our actions and thoughts. Since it is important to have a strong backbone we ought to take a closer look at ours.

Consider, for instance, the following dictum of Carl Sagan, a prominent astrophysicist and a popularizer of science. (Davidson, 1999, p. xv.)

> If "experts" could always be trusted to make the right moral decision, then public participation would not be necessary—but they cannot be, and so it is.

If we have read over this argument without any alarms going off, my point that our backbone is not in good shape would have been made. This is because Sagan's inference is wrong; his dictum contains a logical error. Note that the saying has the following structure: "If *p* then *q*. But we observe *not p*, thus we conclude *not q*." For Sagan's dictum to hold we would need "if *p* then *q*" implying "if *not p* then *not q*." But, it does not.

The dictum's logic is not all that more challenging than what we saw in the supermarket story and yet, if we don't clearly see what is wrong with it we would have to concede that our backbone has failed us. Well, a slip, you may say, is not a big deal. Don't we all make mistakes? Yes, we do. This is one point of the story. Another point of the story is that we would expect science to be free of such mistakes. If an engineer designs a bridge, we expect him to be careful about the validity of inferences he used in the construction. This is clearly something we would wish of anyone whose actions affect us. Even the most methodologically flamboyant social scientist will place his trust in logic when heading for a surgery. Whether we want to admit it or not, we all assume that people whose actions affect us do not make logical mistakes.

Picking on Carl Sagan is important for another reason. It is the smart people we should watch because it is their arguments that we trust most. The problem, of course, is that their brains, just like everyone else's, are not perfect. The more important problem with logical mistakes is that while we all make them; it is very seldom that we have an opportunity to recognize that we do. Here is my very own blunder. I must have had many of them, but, unlike the case below, there was no one around to point them out to me.[15] The blunder uses Wason's selection task experiment which I have later replicated dozens of times in my classes.

Suppose we have four cards and each has a letter written on one side and a number on the other. The four cards lie on a table and show A, B, 4 and 7. Which cards do you need to turn over to determine if the proposition "if A is on one side then 4 is on the other" is true? Should we open cards A and 4, or should we do something else? I am most gratified to find that a good majority of people are guessing A and 4. So did I. Some people guess A and 7 though, and they are correct. My guess of A and 4 was wrong and I would not be able to talk myself out of the trouble. The virtue of science is that whenever I am wrong anyone can prove me wrong and make me see my mistake and correct my reasoning.

Let's then go through the reasoning here. Why is opening A and 7 the right answer? For simplicity let's use *p* to denote our proposition "if A is on one

[15] Many years ago professor Andrzej Nowak of the University of Warsaw has shown me this puzzle and was a witness to my blunder. The puzzle is well known among psychologists as Wason selection task (Wason 1968).

side then 4 is on the other." Two things are clear. First, we don't need to open B since no matter what number we find on the other side, it has no bearing on *p* being true or false, *p* remains true regardless. Second, we have to open A to check if 4 is indeed on the other side; if it not then *p* is, clearly, false. Now, how about 4 and 7? Let's consider 4 first. If we open 4 and the letter on the other side is A then *p* is true for this card. But if the letter is not A then *p* is also true—*p* only claims that "if A is on one side then…" No matter what is on the other side of the 4 card *p* remains true. If we want to know if *p* is true or false opening this card is useless. The same, however, is not true about the card with 7: If we find A on the other side it renders *p* false but if the letter is different it renders *p* true. To conclude, if we want to establish if *p* is true or false we have to open cards A and 7 only.

The most compelling aspect of Wason's example is that it uses the simplest of logical structures—a single implication. If we can show that most people get it wrong at this level, I don't think we can get a more persuasive argument for constructing science in a proper form.

7. Conclusion

If our brain's software which we use for reasoning contains glitches, as Wason's selection task experiment would strongly suggest, then we face double peril when practicing science. First, we don't clearly see deductive consequences of the assumptions we tacitly or overtly impose on a problem. And second, and more disturbing perhaps, the framework which we use to make inferences is itself ridden with errors which make us see some true propositions as false or false propositions as true. When practicing science that involves any nontrivial inferences we don't really have a choice—we have to do it in properly constructed theories. In political science they are called formal theories.

The day when political scientists will stop using the qualifier "formal" will be the day when the vision of Hobbes would prevail. As long as there is a need to identify "formal theory" as a specific subfield of political science neither Hobbes, nor Wilder, nor Riker would find the status quo satisfactory.

In conclusion I should note that it was not my objective to come up with an overview of the formal theory field or its comprehensive history. In my judgment, a different vision of the paper was more useful and this is the vision I have implemented. My reasons and objectives are explained in the introduction.

One reason I did not to pursue certain topics is that others have done that before and I found no good reason to either repeat their points or try to improve on their work. In closing it may be useful to mention a few of these sources. Morris Fiorina (1975) wrote an early overview aimed at explaining to a general audience the role and the reasons for using formal theories in political science. It is a very good paper and I would recommend it to anyone interested in the subject. Paper by Amadae and Bueno de Mesquita (1999) provides an interesting and a comprehensive historical account of the, so-called, Rochester School in political science. Lalman, Oppenheimer and Swistak (1993) compiled a comprehensive overview of the field of formal theory in political science. Shepsle's "Analytical Politics" (2010) is the best undergraduate textbook in the discipline. And, finally, anyone interested in the most recent advances in the field should consult the book of Humphreys (2017.) It provides a wealth of specific applications.

References

Amadae, Sonja M., and Bruce Bueno de Mesquita. 1999. "The Rochester School: The origins of positive political theory." *Annual Review of Political Science* 269-295.

Arrow, Kenneth J. 1951, 2nd ed. 1963. *Social Choice and Individual Values.* Yale University Press.

Black, Duncan. 1948. "On the Rationale of Group Decision Making." *Journal of Political Economy* 23-34.

Cohen, Mitchell, and Nicole Fermon, editors. 1996. *Princeton Readings in Political Thought.* Princeton, New Jersey: Princeton University Press

Coyne, Jerry. 2009. "Seeing and Believing." *New Republic*, February 4: 32-41.

Davidson, Keay. 1999. *Carl Sagan: A Life.* New York: John Wiley & Sons.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Elster, Jon. 1976. "Some Conceptual Problems in Political Theory." In *Power and Political Theory: Some European Perspectives*, by Brian Barry (Ed.). London: Wiley.

Fiorina, Morris. 1975. "Formal Models in Political Science." *American Journal of Political Science* 133-159.

Hobbes, Thomas. 1651 (reprinted 1981). *Leviathan.* London: Penguin Classics.

Holt, Charles A., and Alvin E. Roth. 2004. "The Nash Equilibrium: A Perspective." *Proceedings of the National Academy of Sciences* 3999-4002.

Humphreys, Macartan. 2017. *Political Games.* New York: Norton.

Lalman, David, Oppenheimer, Joe and Piotr Swistak. 1993. "Formal Rational Choice Theory: A Cumulative Science of Politics." In *Political Science: The State of the Discipline II*, by Ada Finifter (Ed.) , 77-104. Washington DC: American Political Science Association.

Macpherson, Crawford. 1985. "Introduction." In *Leviathan*, by Thomas Hobbes, 9-63. Penguin.

Oaksford, Mike and Mike Chater. 2009. "Précis of Bayesian rationality: The probabilistic approach to human reasoning." *Behavioral and Brain Sciences* 69-84.

Riker, William and Peter Ordeshook. 1973. *An Introduction to Positive Political Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Riker, William. 1982. *Liberalism Against Populism.* San Francisco: Freeman.

Riker, William. 1963. *The Theory of Political Coalitions.* New York: Yale University Press.

Russell, Bertrand. 2009. *The Philosophy of Logical Atomism.* Routledge.

Samuelson, Paul A. 1969. "Presidential Address The Way of an Economist." In *International Economic Relations*, by Paul A. Samuelson (Ed.) , 1-11. London: Palgrave Macmillan.

Shapley, Lloyd and Martin Shubik. 1954. "A method for evaluating the distribution of power in a committee system." *American Political Science Review* 787-792.

Shepsle, Kenneth. 2010. *Analyzing Politics.* New York: Norton.

Von Neumann, John and Oskar Morgenstern. 1945. *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Wason, Peter. 1968. "Reasoning about a rule." *Quarterly Journal of Experimental Psychology* 273-281.

Piotr Swistak,
Department of Government and Politics,
and Applied Mathematics, Statistics and Scientific Computation Program,
University of Maryland, College Park, Maryland, 20742, USA
e-mail: pswistak@umd.edu

STUDIA METODOLOGICZNE
NR 39 • 2019, 123–152
DOI: 10.14746/sm.2019.39.5

Andrzej Jarynowski, Michał B. Paradowski, Andrzej Buda

Modelling communities and populations: An introduction to computational social science

Abstract. In sociology, interest in modelling has not yet become widespread. However, the methodology has been gaining increased attention in parallel with its growing popularity in economics and other social sciences, notably psychology and political science, and the growing volume of social data being measured and collected. In this paper, we present representative computational methodologies from both data-driven (such as "black box") and rule-based (such as "per analogy") approaches. We show how to build simple models, and discuss both the greatest successes and the major limitations of modelling societies. We claim that the end goal of computational tools in sociology is providing meaningful analyses and calculations in order to allow making causal statements in sociological explanation and support decisions of great importance for society.

Keywords: computational social science, mathematical modelling, sociophysics, quantitative sociology, computer simulations, agent-based models, social network analysis, natural language processing, linguistics.

1. One model of society, but many definitions

Social reality has been a fascinating area for mathematical description for a long time. Initially, many mathematical models of social phenomena resulted in simplifications of low application value. However, with their growing validity (in part due to constantly increasing computing power allowing for increased multidimensionality), they have slowly begun to be applied in prediction and forecasting (social engineering), given that the most interesting feature of social science research subjects – people, organisations, or societies – is their complexity, usually based on non-linear interactions. Consequently, the application of modelling in applied social science has expanded in the 21st century, with a gradual shift towards data science with the growing availability of Big Data. Both abroad and in Poland one can now observe the spread of modelling and analytical knowledge and competences among young quantitative sociologists, so a shift of priorities in sociology towards computational social science is to be expected.

1.1. Defining models and methods

As in sociology the term 'model' may refer to a range of concepts, in this paper we focus on the narrower, strict understanding of a 'hard' model as a projection of society and its properties which allows computational operations. In this limited sense, a 'hard' model should be described by a mathematical formula or formalism, while its 'soft' counterpart may be describableed with words or infographics attempting to systematise the data or knowledge [Pabjan, 2004]. Natural science is not monolithic [Jarynowski, 2017], thus a branch of science dealing with the example described above will be called system science or system dynamics by computer scientists, dynamical systems or non-linear dynamics by mathematicians, and complex systems or complexity science by physicists.

1.2. Aims of this paper

In this paper, we discuss the characteristics of different types of models, notably genesis, limitations and advantages, and their main impact in the field of modelling communities. The main dividing line distinguishes rule-based and data-driven approaches. In the historical section, we present the roots of modelling and the pairing of concepts from philosophy and hard science in order to help understand the development of modern sociology and its subfield of computational social science [Kułakowski, 2012]. We then describe in detail subcategories of models as well as a few more lines of division corresponding to the specificity of the research questions asked and the methods applied.

1.3. Culture of modelling

Sociologists have always been trying to measure properties of society (such as social sentiments) and how to predict and understand social changes. Natural scientists and mathematicians would then incorporate the results of these measurements and experiments into models, usually based on an analogy with a known natural phenomenon (*approach via analogy*). On the other hand, computer scientists and statisticians have developed a set of "black box" tools to predict states of society. Those techniques (*approach via black boxes*) are currently used in a "Big Data" world, and even if the underlying models are not understandable to a layperson, they usually perform more reliably than rule-based techniques.

2. Modelling by Analogy (rule-based phenomenology)

The interface of sociology (social science) with mainly physics and biology (natural science) is supported by mathematics and computer science. Following the positivist tradition in social sciences of imitating natural sciences [Pabjan, 2004], which will be discussed more extensively in the section on the history of the fields, is usually based on the belief that there exists an analogy between particles, atoms, or molecules on the one hand and living organisms, humans, and even entire ecosystems and societies on the other. This analogy has been made in order to answer questions about the relationships, evolution and functions of society. These postulates and research directive seem to be different from the mainstream (especially Polish) sociological perspective, relying on the role of the experiment [Ossowski, 1962]. For a very long time natural science methodologies in social sciences were reduced to statistical hypotheses and the running of regressions [Sztompka, 1973]. Nature obviously provides many interesting examples of phenomena that can occur in humans [Lesniewska, Lesniewski, 2016], for instance coordinated movement of animals such as flocks of birds or schools of fish. The animals that move in such patterns usually lack a leader, and their knowledge is local, but in most cases the herd is moving in the right direction. Their models can explain the gregarious properties of human communities if the variable trait represents the position of the individuals. However, a breakthrough in the practical application of new techniques for social sciences has only become possible at the turn of the 21st century, when the volumes and complexity of digitalised information on social activities have forced a paradigm change in social studies. The need for understanding social phenomena within a broader, interdisciplinary perspective emerged as a natural step in detailed studies. Rule-based approaches allow building more comprehensive, multidisciplinary planes of social knowledge.

3. Black Box modelling (data-driven)

A different modelling approach is data exploration, where prior knowledge about interaction between people is unnecessary. This may reveal complex relations between agents in various systems without the need to understand the ongoing processes. Statistical and computational visualisation methods and quantitative techniques are currently being fully exploited in research devoted to social systems. Basic quantitative methods such as statistical regressions and data-mining procedures are extended to a range of various innovations that improve data on the evolution of social systems and, most importantly, their objective prediction [Komosiński, 2011]. This area has changed the quantitative sociology of the 21st century. Analysis and modelling of human behaviour has been widely applied in academia, and even more so in commerce. Content targeting such as that used in algorithms behind Google, Amazon or Facebook adverts is a nearly trillion-dollar business [Tyler, 2018], and the accuracy of these tools relies on solid data-driven social models. Moreover, understanding of quantitative methods is taken for granted, so current and future generations of social scientists must be increasingly more familiar with data-driven modelling.

4. History of modelling societies

The modelling of societies has difficulty being recognised as part of social science. Still, the application of computer simulations and mathematical methods to the modelling of social behaviour should not come as a surprise, after they proved their value in the understanding of physical and biological systems [Komosiński, 2011]. The current era of the modelling of societies began with the works of Schelling [Schelling, 1971] and Galam [Galam et al., 1982]. The philosophical background, however, is much older.

4.1. The era of positivism and the establishment of modelling tools and theory

In 1766 Daniel Bernoulli presented probably the first mathematical model[1] of society on a population scale [Jarynowski, 2010] and stated, *I simply wish that, in a matter which so closely concerns the well-being of mankind, no decision shall be made without all the knowledge which a little analysis and calculation can provide.*[2] In the 21st century his wish came true in many countries.

In 1784 Immanuel Kant spoke of 'universal laws' which, *however obscure their causes, permit us to hope that if we attend to the play of freedom of human will in the large, we may be able to discern a regular movement in it. Moreover, that what seems complex and chaotic in the single individual may be seen from the standpoint of the human race as a whole to be a steady and progressive though slow evolution of its original endowment* [Kant, 1784].[3]

Moreover, Kant claimed that elementary mathematics could be synthetic *a priori* because its statements may provide new knowledge not derived from experience [Kant, 1781].

[1] On a side note, Bernoulli used census data of the 17th-century city of Wrocław [Bernoulli, 1766], the home town of the first author.

[2] « Je souhaite simplement que, dans une affaire qui concerne de si près au bien-être de la race humaine, aucune décision ne sera faite sans toutes les connaissances dont une petite analyse et de calcul peuvent fournir. » [Bernoulli, 1760]

[3] „[…] so tief auch deren Ursachen verborgen sein mögen, läßt dennoch von sich hoffen: daß, wenn sie das Spiel der Freiheit des menschlichen Willens im Großen betrachtet, sie einen regelmäßigen Gang derselben entdecken könne; und daß auf die Art, was an einzelnen Subjecten verwickelt und regellos in die Augen fällt, an der ganzen Gattung doch als eine stetig fortgehende, obgleich langsame Entwickelung der ursprünglichen Anlagen derselben werde erkannnt werden können." [Kant, 1781]

In 1851-54 August Comte in his book "*Système de politique positive*" introduced term 'social physics' as 'mécanique sociale', a mechanical social science based solidly on statistics and reflecting positivist approaches of world scientification. He wrote that whatever concerns the human species considered *en masse* belongs to the domain of physical facts; the greater the number of individuals, the more the individual will be submerged beneath the series of social facts which depend on the general causes according to which society exists and is conserved.

In 1904 Max Weber presented the concept of the "ideal type" which permits drawing up a schema where some features are very strongly emphasised in order to analyse the complex and amorphous social world. This fundamental axiom (methodological reductionism) is the quintessence of positivism and reductionism in sociology and lies at the foundation of similar concepts of rationality in economics. In the same year George Simmel formulated the theory of social interaction, where for example fashion is a form of social relationship. As a consequence, at the beginning of the 20th century the first mathematical models of societies appeared within this concept, such as that of evolutionary language adaptation by Otto von Wiener, or Pareto's distribution of wealth.

In 1918, when Marian Smoluchowski concluded that probability is the central problem of modern science, he introduced the methodology of social science into physics. He said: *The probability calculus, which since the beginnings of its development found the most successful application—besides mathematical approaches—mainly in the less anticipated areas of social and biological processes, has recently earned itself a (new) very important field of application: physics.* [Smoluchowski, 1918; transl. MBP].[4]

If physics had failed to learn from sociology at that time, we do not know for how long the former field would have remained in the utopia of determinism and reductionism.

[4] "Die Wahrscheinlichkeitsrechnung, welche seit Beginn ihrer Entwicklung mit größtem Erfolg hauptsächlich in dem sonst der mathematischen Behandlung wenig zugänglichen Bereich sozialer und biologischer Vorgänge angewendet wurde, hat sich in den letzten Zeiten ein überaus wichtiges Anwendungsgebiet erobert: die Physik." [Smoluchowski, 1918]

4.2. Anti-positivism/Post-modernism era. Great recession of modelling abilities

Up to end of World War I the main easily translatable sociological theories were grounded in structural functionalism, interactionism, utilitarianism, exchange and conflict theories. However, the development of modelling in social sciences was later suppressed by the boom of the humanist and anti-positivist philosophies of interwar science in the 1920s and the 1930s and weakened by the subsequent discovery of quantum mechanics, which partially altered classical logic. Individualistic mainstream sociologists interested in post-modernism, with the attitudes of "anything goes" of Feyerabend and Kuhn, led to a dispute over the methods of social sciences which, following [Ossowski, 1962] and [Sztompka, 1973], resulting in a "big divorce" of Polish sociology and the techniques of natural sciences. Perhaps one consequence of this misunderstanding was a crisis of Polish sociology, which was even unable to predict the determinant role of the solidarity movement "Solidarność" in the political transformation in Poland. However, owing to the development of computer calculation techniques, modelling was proven in application in other, sometimes fairly remote, areas such as biology.

4.3. Neo-positivism. Modelling comes back on stage

In the same post-war time, with the development of statistics and computations (where computers started to solve equations) one can observe the emergence of independent fields of social science: analytical sociology [Merton, 1968] and social cybernetics [Wiener, 1950]. Although both areas use mathematical descriptions of human behaviour, in most cases they lack analogies to the laws of nature, because the models only rely on engineering (cybernetics) or sociological (analytical sociology [Hedström, Bearman, 2009]) descriptions of processes. For example:

- Merton [1968] was able to explain the rich club phenomenon (most resources are owned by a very small number of people);
- Coleman [et al., 1957] was able to explain the spread of adaptation;

- Axelrod and Hamilton [1981] were able to explain the rationality of cooperation and conflict;
- Schelling [1971] was able to explain segregation processes.

In Poland, these approaches never got enough attention from social science representatives [Szaniawski, 1971], even though cybernetics [Mazur, 1966] and sociotechnics [Kossecki, 1996] were extensively studied in the USSR.

In the 1950s, the frontiers of science propelled the modelling of societies thanks to the works of ([Bertalanffy, 1950]; "open systems") and later ([Anderson, 1972]; "more is different"). Bertalanffy criticised attempts to describe living systems on the basis of closed thermodynamic systems and called for a completely new approach, taking into account the systems' openness and exchange of energy and information with the environment. Anderson's seminal [1972] paper argued that *the behavior of large and complex aggregates of elementary particles […] is not to be understood in terms of a simple extrapolation of the properties of a few particles. Instead, at each level of complexity entirely new properties appear*

Sociologists such as Durkheim argued that changes in the evolution of social systems occur not only incrementally, but also rapidly, in a manner corresponding to non-equilibrium statistical physics. Thus, the fundamental question of social theory is the concept of human action [Granovetter, 1985], which takes into account not only internalised norms and values (oversocialised — collective behaviour), but also the self-interest of the actors (undersocialised — linear changes) in the individual decision-making process (linear changes). The approaches to describing interactions offered in this framework correspond to Newtonian mechanics (undersocialised systems) and statistical physics (oversocialised systems).

4.4. The digital era – Greatest expansion of modelling

A real boom in the field of modelling societies occurred around the turn of the centuries, when a plethora of new models were developed (mainly around Social Network Analysis), aiming to recreate the entire spectrum of social phenomena [Barabasi et al., 2002]. It is worth mentioning the visible representation of Polish scientists in the process, such as Katarzyna Sznajd-Weron (Wrocław), Andrzej Nowak (Warsaw) or Jacek Szmatka (Kraków) [Sznajd-Weron, 2001].

5. Sociology as a hard science

As we have shown above, modelling may be applied to social science because all the predicted results can be verified. However, this condition is not sufficient to consider sociology a 'hard science' in the same way as physics, chemistry or biology, because of lack of falsifiability. The pioneers of logical positivism such as D'Alambert or The Vienna Circle grounded science in observation, while non-science was deemed non-observational and hence meaningless. Falsifiability is the central property of science and every genuinely scientific claim has the potential of being proven false, at least in principle [Popper, 1959]. According to the principles of logical empiricism, physics may fulfil these conditions more than any other science. (Thus, sociology has been partially considered as a science in the narrower domains of sociobiology [Carnap, 1928] and, later on, sociophysics.) Therefore, the concepts of modelling in sociology must be consistent with physicalism defined by the general principles of the positivist program and its conceptual bases: 1) the construction of a universal system which would encompass all of the knowledge furnished by the various sciences, 2) the absolute rejection of metaphysics, in the sense of any propositions not translatable into verifiable scientific sentences [Neurath, 1931].

On the other hand, modelling in hard science may also explain emergent and critical phenomena. However, the existence of a correlation does not always mean causation [Paradowski, 2011]. In this case, we recommend to use the criteria of logical empiricism. For example, whenever a correlation between processes could be explained by interdisciplinary research based on culture and biology, it is more reasonable to consider biological reasons as a cause. "Metaphysical" approaches have been massively criticised and can only be considered as complementary methods, for example in hypothesis generation [Sedlak, 1980] in model explorations.

6. Some techniques of modelling societies

Mathematical methods and computer simulations are becoming increasingly popular and successfully applied in explanations of phenomena observed in real-world social, economic and biological systems. Here we explore two main computational methodologies, agent-based modelling and system dynamics, and one main analytical tool – social network analysis, all of which allow to represent complicated and complex non-linear social systems [Pabjan, 2004], as well as several other techniques. We also mention various other approaches, such as machine learning, decision trees, and neural networks.

7. Data exploration and data analysis

Collecting data on human activity can rely on linear and non-linear techniques of data analysis. Stochastic processes and data-mining allow investigating the properties of the system. One can analyse the data sets (survey, register-based, time series, spatial, panel, longitudinal data, etc.) and (re) construct (simulate processes) with similar characteristics (e.g. distributions), to predict future states [Jarynowski, 2017]. Series may be analysed from a hierarchical [Mantegna, 1998] or fractal [Kwapień, Drożdż, 2012] perspective to explore the features of the processes. A structured database allows running quantitative tree decision algorithms, though in the case of machine learning and neural network approaches this is not even required.

7.1. Regressions and multivariate analyses – artificial intelligence "in people's homes"

Logistic regression is a mathematical model which allows describing the influence of many variables $X = (X_1, \ldots, X_i, \ldots)$ on dependent variable $Y(X)$. A multivariate analysis consists of those statistical techniques that consider two or more related variables as an entity and attempt to produce an overall result taking into account the relationship among the variables. Regression analyses are widely used for risk/likelihood prediction and forecasting future statuses [Duch et al., 2000].

7.2. Structural equation modelling (SEM)

The idea behind SEM is the possibility of finding out causal relationships by systematic analysis [Konarski, 2009]. Given a set of questions (manifest variables) corresponding to theory-related statements (latent variables), one can provide [Jarynowski, 2017]:

- causal modelling, or path analysis, which validates relationships among the variables and tests them with manifest and latent variables;
- confirmatory factor analysis, an extension of factor analysis in which specific hypotheses about the structure of the factor loadings and intercorrelations are tested against the data.

7.3. Natural Language Processing (NLP)

In the field of computational linguistics, (statistical) modelling has usually aimed at the development of probabilistic models able to accurately predict the most likely next word in a provided string [Goodman, 2001]. As sociology has processed human written or spoken signals since the beginning of this field [Thomas, Znaniecki, 1918], NLP techniques have been widely applied in qualitative analyses. NLP is thus central to a range of real-world language modelling, machine-learning/deep learning and AI (artificial intelligence) applications such as automatic speech recognition, speech-to-text software, handwriting recognition (think smartphone or tablet input), information retrieval, text summarization, or machine translation. A lay reader will have encountered algorithms relying on NLP while dealing with Google Suggest or using mobile phone interfaces while typing.

Recent advances in NLP have shifted to the use of recurrent neural networks [Jozefowicz et al., 2016] and networks with long-term memory [Sundermeyer, Ney, Schlüter, 2015]. The main advantage of connectionist models [Bengio et al., 2003] over traditional statistical NLP such as non-parametric *n*-gram models that had been developing since the 1980s [Jelinek, Mercer, 1980], or subsequent shallower feed-forward neural network models, is their improved classification accuracy alongside the increased ability to generalise and scale [Józefowicz et al., 2016]. This line of enquiry is often termed Neural Language Modelling. Usually, words in the training dataset are parametrized as vectors, whereby lexemes sharing similar functional feature values (e.g. expressions both grammatically and semantically close) are analogously proximal in the induced vector space [Yoon et al., 2016].

7.4. Computational language identification (classifiers)

One noteworthy recent development has been computational identification of the native tongues of second-language users [Paradowski, 2017: 69–71]. This area has grown out of the field of stylometry (concerned with authorship attribution based on statistical calculation of textual features [Barr, 2003]) and automated text classification (which frequently applies machine learning algorithms to sort texts by their type or author attributes [Stemle, Onysko, 2015]).

Resting on the assumption that language users from different native tongue backgrounds exhibit distinguishable profiles in their language production, native language detection likewise uses computer classification techniques and machine learning algorithms trained on large databases of learner texts in an attempt to discover the constellations of words, multiword sequences, error types and other textual features most predictive of authors' mother tongue [Paradowski, 2017: 69–71]. The first application of automated text classification by the author's native languages was a study aiming at distinguishing between Chinese and Japanese learners of English [Mayfield, Tomokiyo, Jones, 2001]. The next two decades have seen studies of learners from various language constellations [Paradowski, 2017]. The accuracies of the best classifiers currently range between 80 and 84 per cent [Tetreault, Blanchard, Cahill, 2013]. For instance, the employment of support vector machines managed to differentiate between original and translated Italian texts in 86.7% of the samples in one corpus [Baroni, Bernardini, 2006].

Models can also be used to uncover insights for second-language user pedagogy. For instance [Rezaee, Golparvar, 2016] used random forest modelling to detect the most salient predictors of clause ordering in academic English.

8. Rule-based methodology

Differential equations (used by dynamical systems, system dynamics and other approaches) were the first to be applied to describe and predict phenomena, but recently even more frequent have been agent-based models (ABM) [Jarynowski, 2017]. Sometimes one problem could be solved in multiple ways. Despite the parallel development of numerical methods for differential equations, agent-based models usually give more accurate predictions and hints for decision-makers. On the other hand, differential equations allow us to understand the core process, something that could be missing in an agent-based approach. As a result, both perspectives are common among specialists and depending on the theoretical or applied aspects, their respective prevalence differs.

8.1. System dynamics (SD)

The main idea in SD is to draw up a set of differential equations representing social phenomena. The model (equations or diagrams) can be solved by numerical or approximative procedures, easily available with several types of computer software used for SD, such as Vensim, Dynamo, iThink or Stella. Their graphical notation allows non-mathematicians such as sociologists to build and solve sets of differential equations [Brouwers, 2009]. The dynamical variables are represented as stocks and rates of change as flows.

8.2. Agent-based models (ABM)

ABM is a computational technique used for experiments with artificial systems populated by agents which interact in non-trivial ways. This is probably the most common approach to modelling used by sociologists. The available toolkits include Netlogo, Swarm, RePast or MASON. In NetLogo, an agent (an autonomous, interacting entity), is represented by a turtle, while a patch is the elementary spatial unit in the grid. The goal is to imitate real patterns by running (often computerised) ABMs under different treatments and conditions. This approach is used much more often by economists than by sociologists [Kaminski, 2012].

8.3. Dynamical systems and chaos

The evolution of dynamical systems is ruled by an implicit relation of input and output. If the behaviour of a dynamical system is highly sensitive to the initial conditions, it can be described in terms of chaos.

8.4. Control theory and cybernetics

The general goal of control theory and cybernetics is discovering patterns and finding principles that can be applied to prediction or control [Poczybut, 2006]. Apart from feedback loops—the central point of system theory-many other techniques are used, such as neural networks, artificial intelligence, artificial life, swarming and floating algorithms. A very important issue is emergence – a process whereby interactions among basic entities exhibits properties that cannot be derived from a simple sum of the entities. After Anderson [1972], this has often been often referred to as the "more is different" principle.

8.5. Critical phenomenon and self-organisation.

The critical phenomena theory [Jarynowski, 2007] has been applied to solving problems in behavioural, social and political sciences, although its roots can be traced to the mathematics and physics of the early 20th century. Catastrophes, self-organising processes and chaos have been broadly overlapping in social and natural science within the complex system paradigm. Many socio-economic systems, where the equivalent of the Reynold number is high enough, can be described with the turbulent flow theory of fluids and gases.

8.6. Evolutionary models

Evolution may explain the formation of different population structures, speciation through the rivalry of two forces of nature – natural selection and genetic drift [Leśniewska, Leśniewski, 2016]. This approach can be derived from the theory of thermodynamics. In reference to sociology, different analogies are used depending on the model, but genetic drift may correspond to the transmission of cultural information, and natural selection can be observed in societies favouring the fittest individuals. There is still a lot of variability due to various factors, like fashion, wars or external. Natural selection results in the differentiation of adjacent societies, because each social group adapts to its local conditions. The cultural transmission of civilisation tends to blur the differences between populations [Dybiec, et al., 2012]. In addition to vertical transmission from parents to descendants, there is horizontal transmission (the media, schools, etc.), whose mission is to maintain the social structure [Bourdieu, 1977]. According to the knowledge of biological systems, the average value of the characteristics of a metapopulation depends on external conditions. Sometimes small changes in external conditions lead to drastic changes in the metapopulation. The analogy to genetic drift in a human population can be inserting through social mechanisms in models. Some of the most popular examples of theories based on evolutionary models in the 20th century were memetics, followed by sociobiology, and subsequently human behavioural ecology. Mathematical models and computer simulations had already been applied in memetics, however 21st-century research prioritizes other approaches.

8.7. Algorithms and heuristics

Models can use deterministic algorithms to find stable states and assume rational behaviour of the agents. This is a well-known problem in mathematics called the stable marriage problem, which also frequently surfaces in other fields, primarily in economics and sociology[5] (Memorial Nobel Prize in 2012).

[5] Actually one of only two memorial Nobel prizes linked to sociology (the other being [Brian, 1994] on risk perception).

In the problem, each agent tries to maximise its own satisfaction (find the best partner) without respecting the rest. However, the algorithms stop at a point where the agents are more or less satisfied with their partners and cannot change them any more; these are called stable states or Nash equilibria. Usually there are several possible stable states with one set of initial conditions.

8.8. Game theory (e.g. the prisoner's dilemma)

Game theory problems are well-studied in economics and can lead to an understanding of individual human decisions [Axelrod, Hamilton, 1981]. The question of whether to cooperate or defect can be answered according to mathematical rules. There are many types of games from single to repetitive (which can illustrate an adaptive strategy based on history). An interesting case of the problem initially considered is a minority game, when the payoff decreases dramatically when too many players choose the same strategy.

8.9. Language simulations

A separate chapter in trying to capture language phenomena in a formal manner while incorporating a social component is simulations devoted to language dynamics, which over the past two decades have been the topic of both dedicated workshops and articles posted on arXiv and published across scientific journals [Paradowski, Jonak, 2012b]. The presentations and papers have mainly been penned by physicists, and cover such phenomena as language acquisition [e.g. Nicolaidis, Kosmidis, Argyrakis, 2009], language evolution and change [de Oliveira, 2013, and papers from the thematic issue of the *Mind & Society Symposium* on "A multi-methodological approach to language evolution"], language spread [Atkinson, 2011], linguistic typology and diversity [Baronchelli et al., 2010, or papers from the theme issue of *Philosophical Transactions of The Royal Society B* 'Cultural and linguistic diversity: Evolutionary approaches'], the emergence of creoles [Tria et al., 2015], diglossia/bilingualism [Castelló et al., 2007], language competition [Abrams, Strogatz, 2003], naming consensus [Lipowski, Lipowska, 2009], and semiotic dynamics. While most of the early works focused primarily on formal representations and regular-lattice *in silico* experiments that were not infrequently grossly inadequate to the scenery of the 21st century [Paradowski, 2012], over time the field has tried to move from coarser-grained game-theoretic [Nolfi, Miroli, 2010] and agent-based models [Nowak, Komarova, Niyogi, 2001] towards increasingly accurate and sophisticated work based on the results of rigorous data-driven research and empirical studies that recreate the necessary conditions and parameters as faithfully as possible.

9. Networks models

9.1. Complex networks

The most important aspect of complex networks is their topology: who is connected to whom. Each item is a node, connected to others by links (edges). The degree is the number of links attached to a given node. The shortest path length is the minimum number of connections to go through to get from one node to another. The clustering coefficient is a measure of whether the neighbours of a node are connected to each other (at the level of the network it tells us how tightly clustered the individuals are in general), while centrality tells us which nodes (or links) are the most important (e.g. act as 'brokers' between the most individuals or have the highest degree). The 'small-world' concept comes from the fact that most of us are linked by small chains of acquaintances. Community detection algorithms in turn identify the intermediate scale structure by breaking the network up into separate social groups. The preferential attachment property means that a node is linked with a higher probability to a node that already has a large number of links.

Network theory is useful when it comes to the study of nature from a systems perspective, and there are already several examples where it has helped understand the behaviour of complex systems. Genetic regulatory networks, Internet transfer protocols, financial market dynamics and social interactions [Fronczak, Fronczak, 2008], such as those involved in the social diffusion of linguistic innovation [Paradowski, Jonak, 2012a] or second language acquisition [Paradowski et al., 2020; Paradowski et al., subm.]. The most exciting property of these systems is the existence of emergent phenomena which cannot be simply derived or predicted solely from the knowledge of the system's structure and the interactions between their individual elements. A modelling methodology proves helpful in analysing many issues of complex systems properties including collective effects and their coexistence with noise, long range interactions, the interplay between determinism and flexibility in evolution, scale invariance, criticality, multifractality and hierarchical structure. Thereby, complex networks are mostly an artificial concept developed by physicists and mathematicians and (at least in theory) they obey universal rules. Complex network analysis helps better understand social behaviour and determine the degree to which individual agents build a functioning and working system.

9.2. Social Network Analysis

One can currently observe an exponentially growing interest in—and importance of—multi-layered and multifaceted interactions. A method which has been more and more widely applied in approaching complex networks from a societal perspective is social network analysis (SNA), made all the more powerful with the growing arsenal of versatile tools [Borgatti et al., 2009, Christakis, Fowler, 2007]. Conceptually, the social network was introduced in the 19th century by Durkheim (1893), who compared the structure and functioning of societies to biological systems consisting of interconnected components. He concluded that social phenomena are not an effect of the actions of particular individuals, but of the interactions between them. In a similar vein within sociological concepts of structural functionalism and interactionism, Malinowski (1924-44) skilfully combined anthropological study with knowledge from the borders of psychology, mathematics and economics in an attempt to get a better grasp of how societies work. In the 1990s, Jacek Szmatka and his team at Jagiellonian University in Kraków actively participated in the development of Social Network Analysis in one of the first labs of this kind in Europe. Currently, SNA serves as a methodology and set of tools enabling a multifaceted in-depth exploration of interacting systems. The topological properties of networks have been proven to determine dynamic processes above the network level, such as cascades of information adoption or default contagion in culture networks [Dybiec et al., 2012]. Dynamical network models are crucial for dealing with adaptive systems, such as those investigating the relationship between interactions and change of behaviour. In the complexity science paradigm, models have been proposed assuming interactions at the network level [Holme et al., 2012], but an integrative framework is often missing that would combine both theoretical and empirical approaches.

9.3. Homophily, social contagion and external field

There are correlations between the properties of ego and its alters in the network of social ties. The famous Framingham Heart Study Network [Christakis, Fowler, 2007] showed that the phenomenon of obesity is linked to social networks of relations between peers. This means that people who have many friends with similar characteristics (e.g. overweight friends) were more likely to share this feature or faced an increased likelihood of getting it in the future. Surprisingly, the greatest effect is seen among close friends and not among people sharing the same household or sex. This observational study is a very good example helping understand the core processes of network formation (topology) and the processes taking place in the network (spread of norm). Thus, homophily can drive topological network dynamics, but social contagion and external field influence the process on the top of the network.

10. How to model societies?

There are a huge variety of possible approaches to modelling [Pabjan, 2004]. The main classification based on fundamental methodological and technical differences was discussed in the previous sections. Focusing on research questions may lead to yet other divisions [Gilbert, Troitzsch, 2005]. Modelling a target system requires the researcher to choose the model's ingredients, including its form, structure, content, and properties [Hardt, 2016]. If we are interested in correspondence to reality, models can be precise, but very complicated. If we agree on lower precision, models can be simple and easy to analyse. A model can be deterministic or stochastic; interactions can be implied by forces, energy, or rules; variables can be discrete or continuous. Models may be further divided into two major, substantially different types, macroscopic and microscopic [Jarynowski, Nyczka, Buda, 2014].

10.1. Macroscopic models

Macroscopic models attempt to answer the questions "how?" and "how much?" They do not care what happens at the micro level of individual units of analysis, only how the respective average values behave. Here, we are dealing mainly with all kinds of structural equations. This description is similar to the macroscopic description of complex systems, such as in the case of thermodynamics, which includes temperature, pressure, volume, etc. This approach can answer many quantitative questions; it can also generate more or less accurate predictions. One example is population growth. Until the mid-20th century, population growth on Earth was observed to be exponential. In the Malthus model, the world's population keeps increasing in size exponentially (J-curve), while in the Verhulst model, it slows down in a logistic (S-)curve. This also points to the very important issue of the reliability of models. In 1972, economists gathered around the Club of Rome predicted that human population would by now (2018) have exceeded 10 billion in the slowest growth scenario, which was not the case (overestimation of more than 30%).

10.2. Microscopic models

The main disadvantage of macroscopic models is the lack of answers to the question about the causes of the occurrence of the phenomena ("why?"). Enter microscopic models. In the case of social and economic sciences, microscopic models are further divided into:

- microsimulations, where the objects change their state due to deterministic or stochastic rules [Jarynowski, 2010];
- Agent Based Models, wherein the system is a collection of "agents" interacting according to some dependent model rules [Jarynowski, Nyczka, 2013].

An agent, as the basic element of the system, has some characteristics (described numerically) and usually interacts with other agents or external factors. The features of a single agent, as well as the rules, are affected depending on the specific model. One can say that agents are a generalisation of the concept of particles, many-body systems, etc. known from natural sciences.

10.3. Goals of modelling

The application of adequate theoretical methods and the empirical approaches of rigorous sciences such as mathematics and physics to economic and social issues has many faces. From a historical perspective, any science starts by collecting and systematising empirical observations, then moves on to the search for regularities and patterns, which finally results in theoretical formulation which captures the observed behaviours and mechanisms. Although the current scientific community is trying to make progress on all three stages, there are still methodological and conceptual issues that we believe should be addressed in that context. Here, we present the most important open or re-opened methodological issues [Berman, Jarynowski et al., 2016]:

- How to use statistical mechanics to approach social issues? This should be addressed at all levels of social systems, which sometimes lack micro-, meso- or macro-foundations.
- How can conclusions from mathematical and physical models be translated into practical policies that would affect society? Theoretical predictions and conclusions drawn from mathematical models should eventually be put into simple, applicable ideas.
- How should empirical data be collected and shared? Data sources are the key to good science and should be trustworthy and broadly shared.
- Universality versus specificity. The tendency in the hard sciences is towards the generalisation of phenomena and properties, while in so-

cial sciences there is a tendency for particularity and specificity. These natures should synergise in order to make progress and on the one hand come up with new ideas, on the other make them applicable.

11. Conclusions

Researchers devoted to the modelling paradigm postulate that in order to capture the complex and dynamic nature of social phenomena, there is a necessity to unify non-structural factors into universal laws. However, simplistic models are not satisfactory in predicting or even reproducing known sociological observation [Biccheri, 2006; Paradowski, 2012]. For example, many natural scientists claim the universality of the power law degree distribution of social networks. Based on accurate empirical studies [Jarynowski, Buda, Nyczka, 2014], this is considered not true any more. On the other hand, modellers increase our general understanding of the mechanisms of social dynamics and the impact of social interaction. The authors' own exploratory research aims at investigating issues which have so far remained virtually unexplored, by applying computational approaches and a methodological complex systems apparatus in combinations never before successfully carried out *in natura*.

11.1. Main applications of modelling

There are many applications of modelling societies and computational social science. Modelling introduces new concepts to sociology to understand changes, new questions that can be asked, and offers new explanations for phenomena. Rule-based models such as system dynamics are useful in explaining processes and identifying the main factors. Data-driven models such as machine learning can be applied in classification tasks and predict future states. Social Networks Analysis has been found to play an increasing role in connecting human behaviour with the attainment of individuals and has already been part of standard organisation management [Żak, Zbieg, Możdżyński, 2014]. On the other hand, the application of stochastic models based on physicalism may confirm in a scientific manner hypotheses proposed by sociology or aesthetics [Buda, 2012].

11.2. Evolution of modelling in sociology

Modelling is an interdisciplinary venture. The development of sophisticated tools can be observed in every scientific discipline. A growing body of 21st-century research has sought to integrate methods and techniques from different fields of science in order to further understanding of societies and the governing principles therein. Sociology has gained more quantitative, mainly statistical tools in order to process the increasing amounts of data. In consequence of the ability to comprehend the calculus, modelling has begun to play a leading role in investigating social structures through the prism of mathematical formalism. Its growth has been facilitated by the interactions among traditionally defined scientific disciplines. Natural scientists discovered (and social scientists recently agreed) that alongside many differences in research objects and methodologies, there are also similar principles that underlie seemingly unrelated phenomena [Guastello, 2017]. Recent progress in this field can be also attributed to developments in communication technologies. Powerful computer facilities and ubiquitous GPS devices have made possible the collection, processing and storage of data on individual behaviour [Belik, 2008]. Traffic jams and evacuation process models have not only been successfully applied, but have become standard procedure for urbanism and building planning [Helbing, 2016]. In conclusion, modelling already played an important, but still not a dominant role in sociology, and even lived to see dedicated periodicals (e.g. *Journal of Mathematical Sociology* or *Journal of Artificial Societies and Social Simulation*, devoted to sociological enquiries carried out with the application of statistical methods and numerical evaluations). However, it seems that the 'European fortress' of humanistic sociology is less susceptible to mathematical physicalism than the 'Anglospheric' scientific style of thought. The fault also lay with those representatives of European natural sciences (especially in post-communist countries) who preferred to stay in their 'sphere of comfort' rather than try and solve real social problems. During deep Polish social-realism Hugo Stainhaus, one the best Polish applied mathematicians, said to his students at Wrocław University [in a recollection by Roman Duda, p.c.]: *The USA is not richer than Poland, because Poland can afford raising and educating good theorists without having any profits from their work. The USA cannot afford that.*[6] Even today, most natural scientists working in the field of sociophysics have not shown much interest in solving social issues (beyond mentioning them in grant applications) and are only focussing on the 'physical' properties of their models. Due to their disinclination to take off the blinders, collaboration with sociologists may still not be appropriate, at least in the Polish context.

11.3. The future of modelling in sociology

In the authors' opinion, modern computational sociology is mainly about basic research on understanding social systems. The role of sociologists consists in education, influencing public debate and actively enforcing social change. Sociology, and computational social sciences in particular, is a public service for developing better societies [Sitek, 2007]. For example, any possible social intervention can be verified *in silico* before application [Bernoulli, 1766]. Sociotechniques and knowledge form other kinds of modelling outcomes have been already applied and is being increasingly applied by governments and companies. Even some effects of social paradoxes[7] of modelled intervention have been noticed. However, some paradoxes could be mitigated in certain scenarios. The most frequent argument used by some opponents – that the social world is too complicated to be described and explained by models – is untenable. While in the past the modelling of multi-agent systems was limited by the lack of reliable data and computer power, one can easily combine many perspectives in quantitative sociology. Thus, some modelling techniques and complementary tools such as regressions, structural equation modelling, social network analysis, and agent-based modelling, have become necessary part of curricula of sociology programmes at world's leading universities (e.g. Oxford, Columbia, UNSW). There is no other choice for sociology than to adopt the

[6] „Stany Zjednoczone nie są bogatsze od Polski, bo Polskę stać na wykształcenie oraz utrzymanie teoretyków, z których gospodarka nie ma korzyści. USA na to nie stać."

[7] For example the self-fulfilling prophecy or the self-defeating prophecy.

computational social science paradigm to a larger extent, with the possibilities coming from computer power. Deep Blue (the computer chess-playing system) won its first game using heuristic algorithms against a human world champion in 1996, but in 2017 a standard personal computer needs only 72 hours of unsupervised learning to beat any human player. The problem of quantitative and computational deficit [Pabjan, 2004] will lead to the 'End of Sociology' [McKie & Ryan, 2015] as it was traditionally defined. Therefore, computational sociology theories based on modelling or physicalism may be synthetic *a priori* with additional properties such as falsifiability. Such a narrowing down gives sociology the opportunity to get closer to hard science. 21st-century societies are changing at an increasing speed. Much of that change is being driven by developments in Information and Communications Technology (ICT) [Helbing, 2016]; thus, the same ICT techniques as modelling must be applied to investigate societies. The suitability and feasibility of the conceptually novel approaches and applicability of the methodology to still new domains, at least in Poland in sociology [Winkowska-Nowak et al., 2004], [Winkowska-Nowak et al., 200], [Nowak et al. 2009] is highly promising and some results are very encouraging.

Acknowledgments

The authors acknowledge support from COST Action CA 15109 "European Cooperation for Statistics of Network Data Science (COSTNET)". MBP and AJ's research is sponsored by SONATA-BIS grant № 2016/22/E/ HS2/00034 from the National Science Centre of Poland. MBP is also supported by COST Action 15130 "Study Abroad Research in European Perspective (SAREP)". The authors are grateful to the anonymous reviewers for their helpful commentary.

Literature

Abrams, D.M. & S.H. Strogatz (2003). Modelling the dynamics of language death. Nature, 424, 900. doi: 10.1038/424900a

Atkinson, Q.D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. Science, 332(6027), 346-349. doi: 10.1126/science.1199295

Axelrod, R, W.D. Hamilton (1981). The evolution of cooperation. Science, 211(4489), 4489, 1390-1396. doi: 10.1126/science.7466396

Baker, A. (2008). Computational approaches to the study of language change. Language and Linguistics Compass, 2(2), 289–307. doi: 10.1111/j.1749-818X.2008.00054.x

Albert, R., A.L. Barabási (2002). Statistical mechanics of complex networks. Rev. Mod. Phys. 74: 47–97. doi: 10.1103/RevModPhys.74.47

Baronchelli, A., T. Gong, A. Puglisi, V. Loreto (2010). Modeling the emergence of universality in color naming patterns. PNAS, 107(6), 2403-2407. doi: 10.1073/pnas.0908533107

Baroni, M. & S. Bernardini (2006). A new approach to the study of translationese: Machine-learning the difference between original and translated text. Literary and Linguistic Computing 21(3), 259–74. doi: 10.1093/llc/fqi039

Barr, G.K. (2003). Two styles in the New Testament epistles. Literary and Linguistic Computing 18(3), 235–48. doi: 10.1093/llc/18.3.235

Belik, V. (2008). Transport, disorder and reaction in spreading phenomena. Niedersächsische Staats-und Universitätsbibliothek Göttingen.

Berman, Y., Jarynowski, A., Borysov, S., Balatsky, A., et al. (2016). Manifest on Predictive Data-Driven Research of Social and Economic Systems: From Data Science to Statistical Physics, Nordita.

Bernoulli D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour le prévenir. Histoire et Mémoires de l'Académie des Sciences, Paris. [English translation by Sally Blower]

Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C. (2003) A neural probabilistic language model. Journal of Machine Learning Research 3, 1137-1155. http://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf

Bicchieri, C. (2006). The Grammar of Society. Cambridge UP.

Borgatti S.P., Foster P.C. (2003) The Network Paradigm in Organizational Research: A Review and Typology, Journal of Management 29(6) 991–1013

Bourdieu, P., & Passeron, J.C. (1977). Reproduction in education, society and culture. Sage studies in social and educational change.

Brian, A. (1994). Inductive Reasoning and Bounded Rationality. American Economic Review (Papers and Proceedings) 84 .

Brouwers, L. (2009). MicroSim: Modeling the Swedish Population.

Buda A.(2012) Does Pop Music Exist? Hierarchical Structure in Phonographic Market, Physica A: Statistical Mechanics and its Applications 391 (21), 5153-5159

Carnap, R. (1928) Der logische Aufbau der Welt, Leipzig: Felix Meiner Verlag

Castelló, X., Loureiro-Porto, L., Eguíluz, V.M., San Miguel, M. (2007). The fate of bilingualism in a model of language competition. In: Takahashi, S., Sallach, D., Rouchier, J. (eds.) Advancing Social Simulation: The First World Congress. Berlin: Springer, pp. 83-94.

Christakis, NA., Fowler, JH. (2007). The spread of obesity in a large social network over 32 years. New England journal of medicine, 357(4), 370-379.

Coleman, JS, Katz E. and Menzel, H. (1957). "The Diffusion of a New Drug Among Physicians". Sociometry 20, 253-270.

Comte, A. (1851-1854). Système de politique positive ou Traité de sociologie instituant la religion de l'Humanité, édition originale en quatre tomes. Paris, Carilian-Gœury et V. Dalmont.

de Oliveira, P.M.C. (2013). Dynamic Ising model: reconstruction of evolutionary trees. Journal of Physics A, 46, 365102. doi: 10.1088/1751-8113/46/36/365102

Duch, W., Korbicz, J., Rutkowski, L., & Tadeusiewicz, R. (2000). Biocybernetyka i inżynieria biomedyczna 2000. Sieci neuronowe, 6.

Dybiec, B., Mitarai, N., & Sneppen, K. (2012). Information spreading and development of cultural centers. Physical Review E, 85(5), 056116.

Fronczak, A., Fronczak P. (2009). Świat sieci złożonych. Od fizyki do Internetu. PWN

Galam, S, Gefen, Y. Shapir J., (1982) Sociophysics: A Mean Behavior Model for the Process of Strike, Journal of Math. Sociology 9, 1-13

Gilbert, N., & Troitzsch, K. (2005). Simulation for the social scientist. McGraw-Hill International

Goodman, J.T. (2001) A bit of progress in language modeling. Computer Speech & Language, 15(4), 403-434. doi: 10.1006/csla.2001.0174

Guastello, S.J. (2017). Nonlinear dynamical systems for theory and research in ergonomics. Ergonomics, 60(2), 167-193.

Hardt, L. (2016) On Similarities in Modelling and Metaphorizing Economic Phenomena. Studia Metodologiczne 36, 147-174

Hedström, P, P. Bearman (2009). The Oxford Handbook of Analytical Sociology. Oxford: Oxford University Press.

Helbing, D. (2016). The Automation of society is next. Zurich, ETH.

Holme, P. (October 2012). Temporal networks. Physics Reports , 519 (3).

Jarynowski, A. (2007). Zjawiska krytyczne w przyrodzie. [online]. Wrocław, Uniwersytet Wrocławski. Dostęp: http://th.if.uj.edu.pl/~gulakov/kracylin.pdf [06.01.2018].

Jarynowski, A. (2010). Human-Human interaction: epidemiology. In Life time of correlation. Wrocław: WN.

Jarynowski, A, Nyczka, P. (2013). Dynamic network approach to marriage/divorces problem, ENIC, DOI 10.1109/ENIC.2014.24

Jarynowski, A, Nyczka, P., Buda, A. (2014). Obliczeniowe Nauki Spoleczne w Praktyce, WN: Wroclaw

Jarynowski, A. (2017), Exploring the dynamics and the complexity of human behavior using nonlinear physics methods, PhD thesis (submitted and received positive review), Jagiellonian University in Cracow

Jelinek, F. & Mercer, R.L. (1980). Interpolated estimation of Markov source parameters from sparse data. In Gelsema E.S. & Kanal L.N. (Eds.) Proceedings of the Workshop on Pattern Recognition in Practice. Amsterdam: North-Holland, pp. 381-397.

Jozefowicz, R., O. Vinyals, M. Schuster, N. Shazeer, Y. Wu (2016). Exploring the limits of language modeling. arXiv:1602.02410v2 [cs.CL]

Kamiński, B. (2012). Podejście wieloagentowe do modelowania rynków: metody i zastosowania. Szkoła Główna Handlowa. Oficyna Wydawnicza.

Kant, I. (1781). Kritik der reinen Vernunft

Kant, I. (1784). Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht. Berlinische Monatsschrift, 385-411.

Komosiński, M. (2011). Życie w komputerze: symulacja czy rzeczywistość?. Nauka 2, PAN

Konarski, R. (2009). Modele równań strukturalnych: teoria i praktyka. Wydawnictwo Naukowe PWN.

Kossecki J., (1996) Cybernetyczna analiza systemów i procesów społecznych. Wyd. WZiA WSP, Kielce.

Kułakowski, K. (2012). Od fizyki do socjologii i z powrotem. W: A. Chmielewski, M. Dudzikowa i A. Grobler (red.), Interdyscyplinarnie o interdyscyplinarnosci, Kraków: Impuls, s: 99-111

Kwapień, J., & Drożdż, S. (2012). Physical approach to complex systems. Physics Reports, 515 (3), pp. 115-226.

Leśniewska M, Leśniewski P (2016) Analogy-Making in Biology. An Essay on the Comparative Spirit. Studia Metodologiczne 37, 155-173

Lipowski, A. & D. Lipowska (2009). Language structure in the n-object naming game. Physical Review E, 80, 056107. doi: 10.1103/PhysRevE.80.056107

Mantegna, R. (1999). Hierarchical Structure In Financial Markets. Eur. Phys. J. B .

Mayfield Tomokiyo, L. & R. Jones (2001). You're not from 'round here, are you? Naïve Bayes detection of non-native utterance text. In NAACL '01 Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies (pp. 1–8). Stroudsburg, PA: Association for Computational Linguistics. doi: 10.3115/1073336.1073367

McKie, L., & Ryan, L. (Eds.). (2015). An end to the crisis of empirical sociology?: trends and challenges in social research. Routledge.

Mazur M., (1966). Cybernetyczna teoria układów samodzielnych. PWN, Warszawa

Merton, R. (1968). The Matthew effect in science. Science, Jan 5; 159 (3810), s: 56-63.

Neurath, O. (1931). Physikalismus, Scientia : rivista internazionale di sintesi scientifica, 50, 1931, pp. 297–303

Nicolaidis, A., K. Kosmidis, P. Argyrakis (2009). A random matrix approach to language acquisition. Journal of Statistical Mechanics, P12008. doi: 10.1088/1742-5468/2009/12/P12008

Nolfi, S., Miroli, M. (eds.) (2010). Evolution of Communication and Language in Embodied Agents. Berlin: Springer.

Nowak, A, W. Borkowski, K. Winkowska-Nowak (Eds.) (2009). Układy złożone w naukach społecznych. Wybrane zagadnienia. Warszawa: Scholar.

Nowak, M.A., Komarova, N.L., Niyogi, P. (2001). Evolution of universal grammar. Science 291, 114-118.

Ossowski, S. (2005). Wzory nauk przyrodniczych wobec osobliwości zjawisk społecznych. W: Marek Kucia (red.), Socjologia. Lektury, Kraków: Znak, s: 28-34.

Pabjan, B. (2004). The use of models in sociology. Physica A: Statistical Mechanics and its Applications, Volume 336, Issues 1–2, s: 146–152.

Paradowski, M.B. (2011). Dekalog analityka danych i infografika – *quid, cur, quomodo*. In: Kluza, M. (Ed.) *Wizualizacja wiedzy: Od Biblia Pauperum do hipertekstu*. Lublin: Wiedza i Edukacja, 338–346.

Paradowski, M.B. (2012). The role of data-driven sociolinguistic research in developing human-machine and machine-machine interaction. In: Obrębska, A. (Ed.) *Practical Applications of Linguistic Research*. Łódź: Primum Verbum, 22–8. https://depot.ceon.pl/bitstream/handle/123456789/3034/palr.pdf

Paradowski, M.B. (2017). Computational native language identification (classifiers). In same, *M/Other Tongues in Language Acquisition, Instruction, and Use* (pp. 69–71). Institute of Applied Linguistics, University of Warsaw.

Paradowski, M.B., A. Jarynowski, K. Czopek & M. Jelińska (2020, subm.). Peer interactions and second language learning: The contributions of Social Network Analysis in Immersion/Study Abroad and Stay-at-Home environments. In: Mitchell, R. & H. Tyne (Eds.) *Language and Mobility: Study Abroad in the Contemporary European Context*. Oxon: Routledge.

Paradowski, M.B. & Ł. Jonak (2012a) Diffusion of linguistic innovation as social coordination. *Psychology of Language and Communication* 16(2) [special issue *Language as a Tool for Interaction*], 53–64. DOI:10.2478/v10057-012-0010-z

Paradowski, M.B. & Ł. Jonak (2012b). Understanding the social cascading of geekspeak and the upshots for social cognitive systems. In: Galton, A. & Z. Wood (Eds.) *Understanding and Modelling Collective Phenomena*. University of Birmingham, 27–32.

Paradowski, M.B., J. Ochab, A. Cierpich & C.-C. Chen (subm.) Accounting for variable L2 success in Study Abroad.

Popper, K. (1959). The logic of scientific discovery, Hutchison & Co.

Rezaee, A.A. & S.E. Golparvar (2016). The sequencing of adverbial clauses of time in academic English: Random forest modelling of conditional inference trees. Journal of Language Modelling, 4(2), 225–244. http://jlm.ipipan.waw.pl/index.php/JLM/article/view/131/143

Schelling, T. (1971). Models of Segregation. The American Economic Review, Vol. 59, No. 2

Sedlak, W. (1980). Homo electronicus. Państwowy Instytut Wydawniczy.

Sitek, W (2007) Paradoksy prognoz socjologicznych. W: Socjologia jako służba społeczna, Kraków: WUJ.

Smoluchowski, M.V. (1918). Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik. Naturwissenschaften, 6(17), 253-263.

Stemle, E. & A. Onysko (2015). Automated L1 identification in English learner essays and its implications for language transfer. In H. Peukert (Ed.) Transfer Effects in Multilingual Language Development (pp. 297–321). Amsterdam/Philadelphia: John Benjamins.

Sundermeyer, M., H. Ney, R. Schlüter (2015). From feedforward to recurrent LSTM neural networks for language modeling. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23(3), 517-529. doi: 10.1109/TASLP.2015.2400218

Szaniawski, K. (1971). Metody matematyczne w socjologii, Ossolineum.

Szmatka, J, J. Skvoretz, J. Berger (Eds.) (1997). Status, Network, and Structure: Theory Development in Group Processes, Nowy Jork: Stanford University Press.

Sztompka, P. (1973). O osobliwościach nauk społecznych raz jeszcze. Studia filozoficzne, nr 8 (105).

Sznajd-Weron, K. (2001), grudzień. Opowieści o fizyce egzotycznej. Wiedza i Życie.

Tetreault, J., D. Blanchard & A. Cahill (2013). A report on the first Native Language Identification shared task. In J. Tetreault, J. Burstein & C. Leacock (Eds.) Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications (pp. 48–57). Atlanta, GA: Association for Computational Linguistics.

Thomas, W.I., & Znaniecki, F. (1918). Chlop polski w Europie i Ameryce [The Polish Peasant in Europe and America]. Warszawa: LWS.

Tria, F., V.D.P. Servedio, S.S. Mufwene, V. Loreto (2015). Modeling the emergence of contact languages. PLoS ONE, 10(4), e0120771. doi: 10.1371/journal.pone.0120771

Tyler, J. (2018) The 10 most valuable retail brands in the world. http://www.businessinsider.com/most-valuable-brands-in-the-world-for-2018-brand-finance-2018-2 .

Watts, D., & Strogatz, S. (1998). Collective dynamics of small worlds networks. Nature, 393 (440).

Wiener, N. (1950). The Human Use of Human Beings, London: The Riverside Press.

Winkowska-Nowak, K., Batorski, D., & Peitgen, H.O. (2003). Wprowadzenie do dynamiki społecznej. Wydawnictwo Szkoły Wyższej Psychologii Społecznej" Academica".

Winkowska-Nowak, K., Nowak, A., & Rychwalska, A. (Eds.). (2007). Modelowanie matematyczne i symulacje komputerowe w naukach spolecznych. Academica.

Yoon, K., Y. Jernite, D. Sontag, A.M. Rush (2016) Character-aware neural language models. Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), 2741-2749. https://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/viewFile/12489/12017

Żak, B, Zbieg, A, Możdżyński D. (2014). Mapaorganizacji.pl – partycypacyjna platforma badań sieci organizacyjnych, W: Nauki o Zarządzaniu, 1 (18), Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław, s. 100-110.

Andrzej Jarynowski
Interdisciplinary Research Institute
ul. Oriona 15/9, 67-200 Głogów
e-mail: ajarynowski[-AT-]gmail.com

Michał B. Paradowski
Institute of Applied Linguistics
University of Warsaw
ul. Dobra 55, 00-312 Warsaw
e-mail: m.b.paradowski[-AT-]uw.edu.pl

Andrzej Buda
Interdisciplinary Research Institute
ul. Oriona 15/9, 67-200 Głogów
e-mail: adrzejbudda[-AT-]gmail.com

STUDIA METODOLOGICZNE
NR 39 • 2019, 153–170
DOI: 10.14746/sm.2019.39.6

Łukasz Hardt

Models and Beliefs they Produce

Abstract. This paper does not focus neither on models nor on modelling procedures but rather on the nature of knowledge about the world models give us. It puts forward the thesis that models are producers of beliefs about their targets. These beliefs may differ both in degree and scope. They are offered by various kinds of models, including models understood in terms of isolations as well as minimal models. This paper puts emphasis on what kind of entities beliefs produced by economic models are.

Keywords: philosophy of economics, models in economics, nature of laws in economics, beliefs.

1. Introduction

Economics is a modelling science. It uses models in order to understand the ways economies function. However, these model worlds are always not perfect pictures of the targets they refer to. But still they can explain. This fact alone warrants the curiosity of philosophers interested in analysing the interplay between models as well as their relationship to laws, theories, and empirical phenomena. First, many of them investigate the ways models can be unrealistic. Here enters U. Mäki's ideas of models' realisticness and unrealisticness [see, e.g., Mäki, 1998]. There are many kinds of unrealisticness. For instance, if a given model lacks its target, then we deal with the issue of referentiality [e.g., Mäki, 2017]. Second, as far as the connection between models and theories is concerned one may subscribe to the so-called semantic view and hence theories can be treated as families of models [e.g., van Fraassen, 1980; Giere, 1988]. However, such an approach poses many problems, e.g., it treats models and theories as interrelated entities but what is quite obvious from observing the ways researchers do science is that they often use models as autonomous agents [Morgan, Morrison, 1999]. And third, what many philosophers are interested in is the issue of models' targets – do we start building models by referring to empirical phenomena or maybe we construct models of artificial worlds and only later we try to refer models' insights to some possible facts [Grüne-Yanoff, 2013]? We may easily multiply such questions, however, what seems to be crucial here is to investigate the ways we learn from models.

This paper refers to the above-mentioned issue of learning from models in a very special way. It asks about what kind of insights models offer us. So, I treat models as producers of our knowledge about the world. What preoccupies me is the character of this very knowledge. Here I definitely do not subscribe to the view that the only kind of knowledge models can give us is a *theoretical* knowledge, namely that models are producers of theories. Consequently, the very traditional and dichotomical treatment of the interplay between models and theories offered by the syntactic and the semantic view of theories is definitely inadequate. If not theories, then what is produced by models? This is the issue I put emphasis on here.

In particular, what I try to show here is that models give us beliefs about the ways the world works. Consequently, section 2 starts with the presentation of two kinds of models, namely models of phenomena and minimal models, precisely models of possible and not actual processes, mechanisms, or events. Section 3 introduces the notions of possibilities and beliefs as the right descriptions of the kind of knowledge models give us. Next, in section 4, a further treatment of beliefs is offered. Conclusions follow.

2. Two kinds of models

If one takes the definition of models, for instance, from *Encyclopaedia Britannica*, then it is clear that models are various kinds of entities but all of them refer to real phenomena. It means that models are models *of something*. They are simplified pictures of their targets. Not only scientists use models but we all do it – every map is a model of a given terrain. Therefore, models are idealizations, namely they take on board elements that are crucial from the researcher's perspective and they pass over the ones of secondary importance. Here we should refer to philosophical analysis of models developed by U. Mäki, N. Cartwright, and D. Hausman. They differ in many respects but in one they are compatible – they all refer to J.S. Mill's observation that causes of economic phenomena interfere mechanically and thus one can analyse them in isolation[1]. For Mill it was also clear that what is true inside models is not true vis-à-vis real phenomena. In his own words:

[…] the conclusions correctly deduced from these assumptions, would be as true in the abstract as those of mathematics; and would be as near an approximation as abstract truth can ever be, to truth in the concrete [Mill, 1836/2008, p. 45].

Here we see Mill's emphasis on deduction and thus inference from the model to the world is of deductive nature. Whether this is feasible is a separate issue to which I am to come back soon. But before that let me present Mäki's modern treatment of models as isolations and precisely his following reasoning:

> Agent *A*
> uses object *M* (the model) as
> a representative of target system *R*
> for purpose *P*,
> addressing audience *E*,
> prompting genuine issues of resemblance between *M* and *R* to arise,
> and applies commentary *C* to identify the above elements and to coordinate their relationships" [Mäki, 2009, p. 75][2].

What is however important here is that the issue of resemblance is understood in a very specific way. It is not a perfect isomorphism between *M* and

[1] Mill was of course conscious that some elements of in the socio-economic may interact chemically and hence hinder any isolation.

[2] Cf. Giere [2006, p. 60].

R [cf. Suppes, 2002]; it is not a partial isomorphism [da Costa, French, 2003]; also, it is not a similarity between the two [Giere, 1998]. Rather it is a possibility that the issue of resemblance between *M* and *R* may arise. What is thus required is a researcher's attempt at making a model (potentially) similar to its target. But judging about the existence of this similarity comes only in the second step. If such a judgment is positive then we are equipped with surrogate systems (models) that "are treated as mediating vehicles in attempts to gain indirect epistemic access to the real world" [Mäki, 2009, p. 77]. Resemblance thus matters and importantly one can ask also if we can learn something about the world if a modeller did not make any attempt at making his model similar to some existing phenomena. I am to come back to this issue while analysing minimal models.

Now, what a proper isolation means? In Lawson's words "[abstraction] must be concerned with the essential rather than merely the most general" [Lawson, 1989, p. 69], and in Mäki's account "[…] an isolating theory or statement is true if it correctly represents the isolated essence of the object" [Mäki, 1992, p. 344]. However, even a proper isolation is not a perfect one. So, if assumptions of isolated models are always not perfect then intra-model deductions are not perfectly transferable between models and their targets. But still, if models are similar to experiments then we may claim that in ideal conditions (i.e., in models) we can investigate the way a given causal mechanism is to work. Thus, if intra-model deductions cannot be deductively transferred to our empirical domains then what kind of knowledge they are? Are we hence in need of referring to some blurred techniques based on induction and intuition?

Mill's sagacity can help us in answering the above questions. Our intra-model deductions can be treated as Mill's "abstract truth(s)" and these are only approximation to "truth(s) in the concrete". Since disturbing causes are always present in the concrete then one should not predict the actual result but rather a tendency to the result, precisely "a power acting with certain intensity in that direction" [Mill, 1836/2008, p. 56]. Thus models produce laws but not laws understood as Humean universal regularities but rather "laws of causation […] stated in words affirmative of tendencies only, and not of actual results" [Mill, 1843, p. 523]. So, for instance, instead of saying that lower interest rates stimulate investments one should claim that a decline in the cost of money produces a tendency of investments to rise. So, in a sense, a statement affirmative of tendencies only can be treated as an approximation of Mill's "truth in the concrete".

Now, what is interesting is that Mill's tendencies can be interpreted in metaphysically rich manner. As Cartwright puts it "Substituting the word 'capacity' for Mill's word 'tendency', his claim is exactly what I aim to establish in this book […]. I suggest that the reader take my 'capacity' and Mill's 'tendency' to be synonymous" [Cartwright, 1989, p. 170][3]. Therefore, referring again to our example of price of money-investment interplay one can state that lower interest rates carry capacity to cause higher investments, or that in the very nature of lower interest rates is to cause a rise in investments. So, models understood in terms of blueprints of nomological machines create favourable conditions for the emergence of natures of given causes. Therefore, the knowledge they offer is again not about universal uniformities but rather nature's capacities. Please note here also that in both Mill's treatment of models' results and in Cartwright's one models inform us about possibilities in the world. In section 3 I am to present more insights into the ways we can understand possibilities in the context of models used to raise them.

After the above presentation of models treated as isolations I would like now to refer to a different kind of models, namely models conceptualized as constructions. In Sugden's own words:

It is essential […] that the model world is a construction of the modeller, with no claim to be anything other than this. Its specification is just whatever the modeller has chosen it to be. In particular, there is no claim that it has been constructed by stripping out some features of the real world and describing what remains [Sugden, 2009, p. 17].

Models come first and only later we give them a kind of ex-post interpretation. In the earlier presented approach where models are treated as isolations the world comes first and models only later. The most well-known case of constructivism in model building is Schelling's checkerboard model. There are many research papers dealing with his model with Sugden's ones [e.g.,

[3] There is an interesting debate concerning the synonymous character of Mill's tendencies and Cartwright's capacities (see, e.g., Schmidt-Petri 2008), however, in the context of this paper its extended presentation is not necessary.

2000; 2009] among others. According to Sugden what Schelling's model offers its readers is not a set of isolations of some empirical phenomena but rather a construction of a very specific and abstract *parallel* world. In other words, it *parallels* the real world rather than isolates it [Grüne-Yanoff, 2009, p. 89]. But now, how one learns from such a model? Here inductive inference plays a crucial role and those using models "[infer from] a particular hypothesis, which has been shown to be true in the model world, to a general hypothesis, which can be expected to be true in the real world too" [Sugden, 2000, p. 19]. This very inference is of inductive kind. Although it can be employed both in explanations and predictions, Sugden claims that the most productive kind of inference from model worlds to real phenomena is the one of abduction. It goes as follows: in the model world, *R* is caused by *F*; *R* occurs in the real world, and by a kind of inductive leap we may state the following: there is reason to believe that *F* operates in the real world. So, we do not have any kind of deduction here but rather something more informal. Some even claim, including Sugden himself, that credibility in models is similar to 'credibility' in realistic novels. As Frigg [2010, p. 251] puts it clear "models share important aspects in common with literary fiction" and even Cartwright states that "a model is a work of fiction" [1983, p. 153] and an "intellectual construction" [ibid., p. 144].

But now let us go even further in the constructivist approach to model building, namely to minimal models. In Grüne-Yanoff's [2009, p. 84] words "[…] the representational structure [of such models] as a surrogate is merely declare, and no further claims are made about the truth of its assumptions, the epistemic status of the principles used in its construction, or the similarity of its economic interpretation (or parts of it) with the real word"; and earlier he notes the following: "[minimal models] lack any similarity, isomorphism or resemblance relation to the world, to be unconstrained by natural laws or structural identity, and not to isolate any real factor" [ibid., p. 83]. All isolationists, including Cartwright, Mäki, and most probably also Mill, obviously would claim that we do not learn from such models since they are referentially and representationally unrealistic. However, Grüne-Yanoff takes a different stance. His way of reasoning starts from a claim that minimal models can be used in assessing propositions about impossibility as well as necessity of hypotheses forming our knowledge. As for necessity, people often maintain certain hypotheses about necessary connections in nature (e.g., between unemployment and GDP growth). Similarly for impossibility. But to learn from such a model, (1) it should represent something at least potentially possible and (2) playing with such a model should modify, e.g., my impossibility hypotheses. Thus such models play important role in learning and they do not only play heuristic role in science as, for instance, Hausman [1992] in his account of "conceptual exploration" would claim.

Now, I would like to stress one important aspect of these models, namely that what they produce can be termed as beliefs about the world. Grüne-Yanoff puts it as follows: "[the minimal model] conceptualizes existing beliefs in a new way, and it draws inferences from thus established possibility, given all relevant beliefs" [Grüne-Yanoff, 2013, p. 853]. And most importantly, since this very model can be treated as a work of fiction then it offers us beliefs about possible entities, processes, proprieties, capacities, and so on. Thus appraising such models representationally is a dead end. And importantly, such possibility claims are about possibilities in our world and not in some *possible* worlds. So, we are close here to the ontology of possibilities and not of actual facts. But now, aren't we close again to Mill-Cartwright view of models' claims as statements of tendencies/capacitates only? Answering such a question requires me to introduce notions of 'how actually' and 'how possibly' explanations. It is done in the following section and in section 4 we are to come back to Cartwright's view.

3. Possibilities and beliefs

In his illuminative chapter on three conceptions of explaining using how possibly clauses Persson [2012] refers to an insightful Dray's opinion which is definitely worth citing here:

The demand for explanation is, in some contexts, satisfactorily met if what happened is merely shown to have been possible; there is no need to go on to show that it was necessary as well. To put the point another way, I shall argue that although, as Professor Toulmin puts it, to explain a thing is often to 'show that it might have been expected', the appropriate criterion for [how-possibly explanations] is broader than this; for to explain a thing is sometimes merely to show that it need not have caused surprise [Dray, 1957, p. 157].

We are in need of how possibly explanations if a given element of explanandum, say *X*, is ruled out by our existing belief system – it is simply impossible to have *X* (cf. impossibility hypotheses referred to earlier). Now, coming back to Grüne-Yanoff's insights, to learn about the world (e.g., the one *possibly* consisting of *X* among others) means to change somebody's confidence in impossibility hypotheses. More precisely, and taking into account the above Dray's opinion, how possibly explanations can account how events that are considered impossible (for instance, our *X* above) could have happened. Such a view was not unknown to Hempel who stated the following:

[…] some of the beliefs we hold concerning relevant matters of fact seem to us to make it impossible or at least highly improbable that X should have occurred [Hempel, 1970, p. 428].

Having the above in mind it is clear that on the other hand how actually explanations account how actual events have come into existence. But in what way how possibly explanations (for simplicity let us call them HPE henceforth) can be defined more precisely? Here Persson [2012] writes about three kinds of such explanations. First, they can show that *X* is not epistemologically impossible. For instance, Schelling [1971] just shows that it is not impossible that patterns of segregations may be present even without a strong preference for segregation. Second, HPE may offer us a kind of possibly how explanations of *X*. Here a nice example can be taken from biology where Darwin himself analysed various possible explanations of how birds wings have evolved. He offered some possible how explanations of this very fact without presenting convincing empirical evidence for some of them [Resnik, 1991]. Third, we have partial how explanations. Here one searches for possible mechanisms giving rise to phenomena in question. In economics we (possibly) have many mechanistic explanations [Hardt, 2017][4].

In the context of the above it is important that in science that are models that produce HPE, namely they offer us possibility claims. As Persson [2012]

[4] It is definitely beyond the scope of this paper to investigate whether these mechanical explanations are of ontic or epistemic kind [see, e.g., Illari, Williamson, 2012].

rightly stresses these claims are of epistemic nature, precisely they refer a claim under consideration to the current and known state of our world, e.g., *X* can be epistemically possible if it may be true for all that we know. However, and interestingly, in the context of models producing HPE one may refer also to alethic possibilities [Grüne-Yanoff, 2017]. For instance, one may claim the following: "It is possible that the author of this paper was born in London". Is this possibility claim true? Yes, it could be, however, since I was born somewhere in Poland it contradicts current knowledge. Now, coming back to Schelling's model and assuming that in actual cities racial segregation is caused by strong preference for self-segregating, then his model's claim that segregation may possibly arise event without such a preference degenerates into a kind of alethic possibility. However, and to sum up, both epistemic and alethic possibilities influence our beliefs regarding the world. And this is how we learn, since to learn means to change one's beliefs' structure. Let us give here again the floor to Hempel:

[…] questions of the form 'why is not the case that p?' […] might well be rephrased as 'how-possibly' questions: 'How could it possibly be the case that not-p?' […]. A pragmatically adequate answer again will have to clear up the empirical or logical misapprehensions underlying this belief" [Hempel, 1970, p. 429].

So, and again, we have beliefs at the very centre of the process and the way we learn about the world. But let me now draw my reader's attention to another philosopher who put beliefs at the centre of his theory of how we apprehend the world. Here we have to refer to Charles Peirce, the author of *The Fixation of Beliefs* (1877), who famously states the following:

Doubt is an uneasy and dissatisfied state from which we struggle to free ourselves and pass into the state of belief; while the latter is a calm and satisfactory state which we do not wish to avoid, or to change to a belief in anything else. […] the irritation of doubt causes a struggle to attain a state of belief. I shall term this struggle *inquiry* [Peirce, 1877, pp. 66-67].

So, the very essence of learning is a transition from the state of doubt to the one of belief. But still, a given belief which is formed after abandoning some doubts may again be put into question. Hence degrees of beliefs [ibid.]. Peirce analyses four methods of fixing beliefs, i.e., tenacity, authority, the a priori method, and 'the method of science'. Of these four the last one is privileged. This is so since for Peirce scientists observe mind-independent reality: "There are real things, whose characters are entirely independent of our opinion about them; those realities affect our senses according to regular laws" [ibid., p. 74]. Thus the crucial role of empirical evidence in forming beliefs. Peirce is hence optimistic about the possibility of having *true* beliefs in science. However, if we do not have laws of nature in economics, and if we as economists influence the world we investigate, then 'the method of science' is not to offer us high degree beliefs [cf. Hardt, 2017].

Interestingly, the celebrated 1877 work by Peirce was discussed in the context of economics by R. Backhouse in his 1994 paper which appeared in the inaugural issue of *The Journal of Economic Methodology.* In this very paper Backhouse put forward a thesis that in economics disagreements are impossible to overcome. This is so since (1) empirical evidence provides weak constraints on theorizing and economics is centred on theorising; (2) economic theories change the world economists try to understand; (3) "[…] and great problems are imposed by the changing nature of the economic world" [Backhouse, 1994, p. 41]. While I agree with Backhouse's thesis, I do not accept some of his arguments backing it. First, economics is centred on models rather than on theories. Second, markets definitely evolve and thus are constantly changing but it does not imply that natures of elements they consist of are also in a permanent change. If one, as myself, subscribes to the world of powers and capacities, then natures of given entities do not evolve [Cartwright, 1999, p. 49]. For instance, let me claim the following: in the very nature of lower cost of money is to stimulate investments. This nature is not to change while we modify circumstances but in some contexts it is to manifest itself and in some it will be dormant. Why not to interpret in such a *realist* spirit the following sentence from Peirce's 1877 paper? So, it states the following:

Belief does not make us act at once, but puts us into such a condition that we shall behave in some certain way, when the occasion arises [Peirce, 1877, p. 67].

I am to come back to the above issue later in the next section. Now, what about changing focus of inquiry from theories to models? In such a case beliefs cease to be elements of theories but they are rather products of models. Therefore, Backhouse is not right in claiming that the vagueness of many statements in economic theory is due to its detachment from empirical domains. Here he gives examples taken from Bloor and Bloor [1993], e.g., 'It seems likely that money causes inflation', or 'I wish to suggest that money causes inflation'. In my perspective where models are first and theories only later such elusive statements are nothing wrong. As it was shown in section 2, models' claims about the world are often about tendencies, probabilities, or capacities, and rarely if ever Humean regularities.

Now, referring again to Peirce's idea of the fixation of beliefs one may ask what does this very fixation mean if beliefs are agents' *opinions* about the world that are offered by models. First, let me take the case of *models of something*, namely models understood as isolations or as imperfect pictures of some targets. In such a situation one can only take fixation as something local and not general. In other words, beliefs produced by a given model can be fixed in relation to this very model. So, the more a given situation similar to the one described by a model in question is, the higher degree of beliefs about this very situation we have. As Guala puts it: "The fact that a model turns out not to work under certain circumstances does not count as a refutation of the model but only as a failed test of its applicability in a given domain" [2005, p. 220]. I put forward the idea of models as believable words elsewhere and I do not want to go into details here [Hardt, 2017, pp. 33-168].

But what about minimal models where the issue of referentiality is not in place? As I have mentioned earlier these models help us to learn about the world by revising our beliefs. So, now the following question is worth asking: are beliefs in Grüne-Yanoff's sense [e.g., 2013] similar to the ones appearing in Peirce's seminal 1877 work? I would give here a rather positive answer, however, some reservations are definitely worth making. First, Peirce generally claims that while beliefs nearly always are imperfect descriptions of the world they can nevertheless attain (however rarely) the status of true beliefs [cf. Almeder, 2014, p. 105]. In my perspective beliefs can only be true inside models used to produce them. Second, pragmatists, including Peirce himself, generally put emphasis on prediction and control rather than on explanation. In Grüne-Yanoff's take on models they are more treated as entities aiming at explaining than simply on predicting. Third, pragmatists are generally treated as fallibilists. Therefore, given beliefs may become wrong. But we should remember that these are beliefs that turn out to be wrong in a given context and not necessarily a particular model as such. Once you change a context your belief may turn out to be correct. Despite the above mentioned differences there are striking similarities between Peirce's beliefs and beliefs understood as claims about the world produced by models, including economic ones.

But what is now necessary to complete the analysis of beliefs is to focus more on models' users, namely the ones internalizing beliefs about the world given by models. It is done in the next section.

4. Beliefs and people possessing them

If beliefs are propositional attitudes, as analytic philosophers would generally claim, and if these attitudes are mental models of having some opinions, then it is worth analysing beliefs abstracting from the ways they are produced and taking more into consideration the ones possessing them. So, beliefs are responsible for the productionof behaviour. If I believe that doing *X* is to harm me, then I am to abstain from *X*. However, *X* does not harm in every circumstances. For instance, if *X* stands for lighting a match then I am definitely not to do it if I believe that I am near some flammable materials. So, I will be characterised by a disposition of trying not to do it. But this disposition may be dormant if I am not surrounded by such materials. As Schwitzgebel [2015] puts it "[…] the demand for an absolutely precise specification of the conditions under which a disposition will be manifested, without exception, may be excessive". As Cartwright [1983] has noted, even perfectly respectable claims in the physical sciences often hold only ceteris paribus or 'all else being equal'. So, we are close to the so-called dispositionalism about beliefs, precisely if somebody believes *Y* then this person possesses a particular disposition pertaining to *Y*.

So, in the context of the above it is now worth making a distinction between occurrent and dispositional beliefs . Take, for instance, the following statements: (1) 'The author of this paper runs a ten mile race', and the second one, (2) 'The author of this paper is running a ten mile race'. The former one is a dispositional statement and the latter a statement about a particular occurrence. Here (1) can be true even if the author is asleep. So, (1) can be true even if (2) is false. But (1) cannot be true unless there are circumstances under which (2) would be true. Schwitzgebel [2015] explains it as follows while taking into account an agent having these two kinds of beliefs: "A subject dispositionally believes *P* if a representation with the content *P* is stored in her memory or 'belief box'. When that representation is retrieved from memory for active deployment in reasoning or planning, the subject concurrently believes *P*. As soon as she moves to the next topic, the occurrent belief ceases". Such a dispositional view of beliefs is shared by many philosophers, including D. Armstrong and D. Hume, among others. So, in the two forthcoming paragraphs I am to comment on their ideas concerning beliefs.

In his 1973 book on *Belief, Truth, and Knowledge* Armstrong analyses three possible and interrelated answers to the very question on the nature of beliefs, precisely they can be treated as conscious occurrences in the believer's mind, they may be understood as dispositions of the believer, and they can just be states of the believer's mind. Let us focus here on the second understanding of beliefs. Here Armstrong clearly distinguishes between thing's disposition and the manifestation of that disposition. For instance, between the brittleness of a piece of glass and its actually breaking. Therefore, our knowledge consists of two parts, i.e., capacities (or dispositions to use Armstrong's word) and their manifestations. We are so very close to Cartwright's approach to human knowledge. She states the following: "Our most wide-ranging scientific knowledge is not knowledge of laws but knowledge of the *natures* of things" [Cartwright, 1999, p. 4], and in her earlier book: "The generic causal claims of science are not reports of regularities but rather aspirations of capacities, capacities to make things happen, case by case" [Cartwright, 1989, pp. 2-3]. If beliefs are produced by models, and if they are understood as dispositions, then a powerful conclusion follows: models inform us about entities' dispositions or, in other words, about nature's of things/states/processes/etc. Thus Armstrong's ideas concerning beliefs can be situated close to the Cartwright's ones.

Now, let me refer to Hume, and especially his observations about the nature of beliefs. In the *Treatise* he says that the goal of scientific inquiry is "a degree of belief, which is sufficient for our purpose" [Hume, 1740/2000,

p. 122]; however, a belief which would be free from any doubts is almost not possible: “Belief, being a lively conception, can never be entire, where it is not founded on something natural and easy” [ibid.]. As Hardt [2017, p. 26] explains “[…] the idea of belief was very important for Hume, since it was used as a tool for overcoming the conflict between knowledge and probability”. Please thus take into account the following Hume’s opinion:

Knowledge and probability are of such contrary and disagreeing natures, that they cannot well run insensibly into each other, and that because they will not divide, but must be either entirely present, or entirely absent [Hume, 1740/2000, p. 119].

The above is very interesting: beliefs should not be understood in terms of probabilities. This is so mainly because beliefs refer to prototypical proprieties of things and not to statistical *normality*. For instance, I can still claim that in nature of birds is to fly even if due to some natural disaster all birds cease to fly. What can be claimed here also is that the above dispositional account (including Armstrong’s ideas) and here presented Hume’s understanding of beliefs can be employed to the economic realm. As Cartwright puts it: “[…] our typical methodologies and our typical applications, both in the natural and in the social sciences, belong to a world governed by capacities, and indeed cannot be made sense without it” [Cartwright, 1999, pp. 1-2]. Here we find a clear influence of Aristotelian ideas on her thinking. Take, for instance, the following sentence from *The Nicomachean Ethics*: “Now fine and just actions, which political science investigates, admit of much variety and fluctuation of opinion, so that they may be thought to exist only by convention, and not by nature” [Aristotle, 1995, p. 1730]. However, the above does not mean that we should treat Hume as the one subscribing to such a metaphysically rich treatment of beliefs. On standard reading he is the one subscribing to the deflationary metaphysical claim that there is nothing more to causation than constant conjunction. And thus his beliefs being in-between knowledge and probability do not refer to some specific *ontology of beliefs*[5].

[5] It is however worth noticing that some Hume’s interpreters treat him even as a „quasi realist”, namely the one analysing operations of nature [e.g., Blackburn, 2000]. But such voices still are rare [cf. Hardt, 2017, pp. 102-107].

Now, let me come back to the topic of beliefs and people possessing them. Here one issue is very important, especially in the context of the previously discussed minimal models, namely of whether one may have beliefs without explicit representation. We should thus focus on the possibility (*sic!*) of implicit beliefs, namely such beliefs which while being possessed by an individual do not possess representations. Dennett [1978] claims in this context that we all have such beliefs and that they are derivable from explicit beliefs, namely the ones representing some existing phenomena. Now, Grüne-Yanoff's statement that "[minimal model] conceptualizes existing beliefs in a new way"[2013, p. 853] can be treated as the one related to the formation of implicit beliefs. However, one cannot also exclude the situation in which the previously implicit belief shifts into the explicit one once a context changes. Here enters Grüne-Yanoff's idea that in order to learn from minimal models they should represent entities at least potentially possible. Therefore, it seems to be useful to link literature on minimal models and beliefs they produce or modify with contributions studying the very nature of explicit and implicit beliefs in the human cognitive apparatus.

5. Conclusions

The existing literature in the philosophy of economics which deals with models focuses on both modelling and models but put less emphasis on the nature of knowledge models produce. It was shown in this paper that this very knowledge is best understood as beliefs about the world. Moreover, these beliefs should be treated as statements about dispositions, capacities, and natures. Therefore, they should not be interpreted in terms of probabilities but rather as claims about prototypical characteristics of entities being under investigation. However, beliefs are true within models producing them. Interestingly, products of very divers types of models can be analysed in here proposed framework. However, models treated as isolations (in Mill's tradition) produce *new* beliefs and models-constructions, and especially minimal models, rather modify existing beliefs. Nevertheless, the difference is rather a difference in degree rather than in kind.

Acknowledgments

The research here presented was financed by a grant from the National Programme for the Development of Humanities (grant no. 0266/NPRH4/H2b/83/2016).

References

Almeder R., (2014), “Pragmatism and Science”, in: S. Psillos, M. Curd (eds.), The Routledge Companion to Philosophy of Science, Oxon, Routledge, pp. 103-111.

Aristotle (1995), “The Nicomachean Ethics”, in: J. Barnes (ed.), The Complete Works of Aristotle, Princeton, Princeton University Press, pp. 1729-1867.

Armstrong D.,(1973), Belief, Truth, and Knowledge, Cambridge, Cambridge University Press.

Blackburn, S., (2000), “Postscript to ‘Hume and Thick Connexions’”, in: R. Read, K.A. Richman (eds.), The New Hume Debate, London, Routledge, pp. 100-112.

Bloor M., Bloor T., (1993), “How Economists Modify Propositions”, in: W. Henderson, T. Dudley-Evans, R. Backhouse (eds.), Economics and Language, London, Routledge, pp. 153-172.

Cartwright N., (1983), How the Laws of Physics Lie, New York/Oxford, Oxford University Press.

Cartwright N., (1989), Nature’s Capacities and their Measurement, Oxford, Clarendon Press.

Cartwright N., (1999), The Dappled World: A Study of the Boundaries of Science, Cambridge, Cambridge University Press.

Costa N.C.A., French S., (2003), Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning, Oxford, Oxford University Press.

Dennett D., (1978), Brainstorms, Cambridge, MIT Press.

Dray W., (1957), Laws and Explanations in History, Oxford, Oxford University Press.

Frigg R., (2010), “Models and Fiction”, Synthese, 172(2), pp. 251-268.

Giere R., (1988), Explaining Science: A Cognitive Approach, Chicago, University of Chicago Press.

Giere R., (1999), Science without Laws, Chicago, University of Chicago Press.

Giere R., (2006), Scientific Perspectivism, Chicago–London, University of Chicago Press.

Grüne-Yanoff T., (2009), “Learning from Minimal Economic Models”, Erkenntnis, 70(1), pp. 81-99.

Grüne-Yanoff T., (2013), “Appraising Non-Representational Models”, Philosophy of Science, 80(5), pp. 850-861.

Grüne-Yanoff T., (2017), “Learning from Models about Possibilities”, unpublished manuscript.

Guala F., (2005), The Methodology of Experimental Economics, Cambridge, Cambridge University Press.

Hardt Ł. (2017), Economics without Laws. Towards a New Philosophy of Economics, Cham, Palgrave Macmillan.

Hausman D., (1992), The Inexact and Separate Science of Economics, Cambridge, Cambridge University Press.

Hempel C.G., (1970), Aspects of Scientific Explanation, New York, The Free Press Paperback.

Hume D., (1740/2000), A Treatise of Human Nature, Oxford, Oxford University Press.

Illari P., Williamson J., (2012), "What is a Mechanism? Thinking about Mechanisms Across the Sciences", European Journal for Philosophy of Science, 2(1), pp. 119-135.

Lawson T., (1989), "Abstraction, Tendencies, and Stylised Facts: a Realist Approach to Economic Analysis", Cambridge Journal of Economics, 13(1), pp. 59-78.

Mäki U., (1992), "On the Method of Isolation in Economics", in: C. Dilworth (ed.), Idealization IV: Intelligibility in Science, Special Issue of Poznan Studies in the Philosophy of the Sciences and the Humanities, 26, pp. 317-351.

Mäki U., (1998), "Realisticness", in: J.B. Davis, D.W. Hands, and U. Mäki (eds.), The Handbook of Economic Methodology, Northampton, Edward Elgar, pp. 409-413.

Mäki U., (2009), "MISSing the World. Models as Isolations and Credible Surrogate Systems", Erkenntnis, 70(1), pp. 29-43.

Mäki U., (2017), "Highly Unrealistic Models: Their (Un)realisticness and their Functions", unpublished manuscript.

Mill J.S., (1836/2008), "On the Definition and Method of Political Economy", in: D. Hausman (ed.), The Philosophy of Economics, Cambridge, Cambridge University Press, pp. 41-58.

Mill J.S., (1843), A System of Logic. Ratiocinative and Inductive, London, John W. Parker.

Peirce C.S., (1877/1982), "The Fixation of Belief", Popular Science Monthly, reprinted in: H.S. Thayer (ed.) (1982), Pragmatism: The Classic Writings, Indianapolis, Hackett, pp. 61-78.

Persson J., (2012), "Three Conceptions of Explaining How Possibly and One Reductive Account", in: H. Regt, S. Hartmann, S. Okasha (eds), The European Philosophy of Science Association Proceedings, 1, Dordrecht, Springer, pp. 275-286.

Resnik D.B., (1991), "How-possibly Explanations in Biology", Acta Biotheoretica, 39, pp. 141-149.

Schelling T., (1971), "Dynamic Models of Segregation", Journal of Mathematical Sociology, 1(1), pp. 143-186.

Schwitzgebel E., (2015), "Belief", The Stanford Encyclopedia of Philosophy (Summer 2015 Edition), E.N. Zalta (ed.), https://plato.stanford.edu/archives/sum2015/entries/belief/, accessed 20/2/2017.

Schmidt-Petri C., (2008), "Cartwright and Mill on Tendencies and Capacities", in: L. Bovens, C. Hoefer, and S. Hartmann (eds.), Nancy Cartwright's Philosophy of Science, New York, Routledge, pp. 291-302.

Sugden R., (2000), "Credible Worlds: The Status of Theoretical Models in Economics", Journal of Economic Methodology, 7(1), pp. 1-31.

Sugden R., (2009), "Credible Worlds, Capacities, and Mechanisms", Erkenntnis, 70(1), pp. 3-27.

Suppes P., (2002), Representation and Invariance of Scientific Structures, Stanford, CSLI Publications.

van Fraassen B., (1980), The Scientific Image, Oxford, Oxford University Press.

Łukasz Hardt
Chair of Political Economy,
Faculty of Economic Sciences,
University of Warsaw,
Dluga St. 44/50,
00-241 Warsaw, Poland,
e-mail: lhardt@wne.uw.edu.pl

STUDIA METODOLOGICZNE
NR 39 • 2019, 171–193
DOI: 10.14746/sm.2019.39.7

Marek M. Kaminski

Erosion of belief in "social rationality": How game theory and social choice theory changed the understanding and modeling of social rationality*

Abstract. This article discusses how the developments in game theory and social choice theory profoundly transformed our understanding and modeling of social rationality in the social sciences due to the erosion of the concept of social optimum. I discuss the Prisoner's Dilemma and relevant examples of social situations, analyze the difficulties that arise when games are repeated, and finally, check how the main results of social choice theory influenced our understanding of the "best" social outcome.

Keywords: rationality, Prisoner's Dilemma, Arrow's Theorem, folk theorems, social optimum.

1. Introduction

One of the long-lasting byproducts of the Age of Enlightenment was the confidence in social rationality and linear progress. Bernard Mandeville (1705) firmly believed that individual vices necessarily produce social good. In Adam Smith's (1776) more refined analysis, the concept of an "invisible hand" represented the universal mechanism of aggregating multitudes of individual activities. Unconstrained markets were smart and good. Selfishly motivated actions were automatically and miraculously converted by markets into "optimal," or at least "near-optimal," social outcomes – this was an implicit

* The author is grateful for comments to Barbara Kataneksza. The support of the Center for the Study of Democracy at the University of California, Irvine is gratefully acknowledged.

meaning of "social rationality." The forces of competition and the interplay of demand and supply were inevitably pushing selfishly motivated bakers, butchers and brewers into delivering products for everybody's benefit.

The first half of 20th century brought a better understanding of the concepts of social rationality and social optimum. Pareto (1906) substituted a vague Bentham's (1780) idea of the "greatest happiness of the greatest number" with a precise definition that in a socially optimal outcome, nobody can be made better off without making somebody else worse off. These were minimal conditions since, in a typical economic setting, there are many such outcomes (called *Pareto-optimal*). Bergson (1938) introduced what later became known as Bergson-Samuelson social welfare function (swf). Any well-defined ethical or economic evaluation of social outcomes would produce its own Bergson-Samuelson swf that would assign higher numerical values to more preferred outcomes. If there were consensus in a society on its ethical and economic principles, the relevant swf would provide a perfect implementation of those principles and a measuring rod for the corresponding concept of social rationality.

The optimism of Scottish Enlightenment was attacked over the 19th and 20th centuries both by ideologically motivated socialists and communists, and by the rightist supporters of corporate state. However, the decisive blows to Smith's elegant intellectual construction were slowly hammered out by ideologically neutral developments in microeconomics that were soon represented in more formal ways in two subfields of applied mathematics, game theory and social choice theory.

An implicit – and problematic – assumption behind Smith's and similar approaches was the lack of interdependence among individual decisions, perhaps with the exception of competition among producers. To some extent, such an assumption reflected the low complexity of Enlightenment economies. Uncomplicated markets of bakers, butchers and brewers consisted of many small players who were providing simple private goods. If your neighbor bought a loaf of bread at a local bakery then, assuming no shortage, her consumption wouldn't affect your consumption in any way. In other words, in Smith's markets there were no "externalities," i.e., side costs or benefits for players other than those directly involved in an economic transaction. At that time, there was no concept of "public goods" as introduced by Samuelson

(1954) whose main characteristic was non-rivalrous access to resources such as public broadcasting or national defense, and which was later complimented by related concepts of club goods (Buchanan 1965) and common-pool resources (Ostrom 1990). For such non-private goods, consumer's utility could be strongly dependent on the actions of other consumers. For instance, if many of your neighbors used a community pool at the same time, then the crowd would make swimming utterly uncomfortable. If your neighbor switched to a public channel on her TV, you could still watch it on your TV. However, if nobody paid fees or taxes to support the public broadcaster, your public channels would stay dark. This was a more complex economic environment than Smith's markets. In such an environment, it was easily possible to obtain Pareto inferior outcomes that would be also non-optimal according to any reasonable Bergson-Samuelson swf.

Noncooperative game theory introduced especially useful tools for modeling interdependent decisions that enabled considering more complicated situations than Smith's simple markets. It taught us that humans may be locked in a variety of traps that convert their individually rational decisions into social disasters. This insight also applies to, but is not limited to, certain markets. Sometimes, unconstrained markets may lead to overexploitation or undersupply. Further developments in game theory revealed even more disturbing difficulties with the concept of social rationality. When interactions are repeated with sufficient intensity, in many games stable cooperation is in equilibrium. However, the range of equilibrium behavior may vary from total cooperation to total defection. In other words, if equilibrium is a proxy for social rationality, too many various types of behavior can be labeled as "socially rational" to make such a label useful. Finally, social choice theorists established that even if we assume the existence of a benevolent and wise decision maker willing to take into account the preferences of everybody in order to reach social optimum, the conversion of individual preferences into smart social solutions faces unsolvable difficulties. According to social choice, the very concept of "social optimum" is problematic.

The present chapter discusses how the developments in game theory and social choice theory deeply transformed our understanding and modeling of social rationality in the social sciences due to the erosion of the concept of social optimum. I discuss the Prisoner's Dilemma and relevant examples of social situations, analyze the difficulties that appear when games are repeated, and finally examine how the main results of social choice theory affected our understanding of the "best" social outcome.

2. Game theory and tensions between individual and social rationality

The milestones in the formal development of game theory were von Neumann's (1928) proof of the existence of minimax solution in two-player zero-sum games and von Neumann and Morgenstern's (1944) comprehensive blueprint for the future development of the discipline. In the most important type of game, strategic game, two or more players independently make decisions (choose their "strategies") and the results of their individual choices result in certain payoffs for all of them. Players are interested in their own payoffs only and prefer higher payoffs to lower ones.

One of the first games that generated a lot of interest was the Prisoner's Dilemma (PD), first described by RAND mathematicians Flood and Dresher, and popularized with an eye-catching interpretation by another mathematician, Albert Tucker (Flood, 1950; Tucker 1952). This simple game describes the quintessential problem that may arise when individual decisions lead to an important social outcome.

The Tucker's story unfolds as follows. Don and Tom, two robbers, were caught by police. They are held incommunicado and expect to be charged. They care only about their own sentences; they can stay mute (cooperate with each other) or testify against their partner (defect). The prosecutor describes to Don the consequences of his decisions: if they both stay mute, there is not enough evidence for a serious charge and they are charged with a minor crime. Both inmates get one year in prison. If one of them testifies, the other one is convicted entirely on the strength of such testimony with a three-year sentence. The testifying inmate gets free. If they both testify, such testimonies are relatively less valuable and both sentences are two years. Then the prosecutor presents a symmetric scenario to Tom.

Games like the PD are most conveniently represented in a matrix form. The rows correspond to Don's (Player 1's) strategies and the columns correspond to Tom's (Player 2's) strategies. The numbers at the intersection of some row and column represent the payoffs that Don and Tom respectively obtain from playing the corresponding strategies. Since both players prefer a shorter sentence to a longer one, the numbers representing their payoffs have negative values.

		Tom	
		mute	testify
Don	Mute	−1, −1	−3, 0
	Testify	0, −3	−2, −2

Figure 1: The Prisoner's Dilemma

The dilemma is constituted by two facts.

1. First, testifying is a dominant strategy, i.e., it is always better to testify than to stay mute. This is the concept of individual rationality employed in the PD. To see this, let's consider Don's decisions assuming that he cannot influence in any way Tom's decisions. If Tom stays mute, then Don breaks free for testifying versus getting one year for staying mute; if Tom testifies, then Don gets two years for testifying and three years for staying mute. Tom's situation is analogous.

2. Second, when both inmates testify, they both get higher payoffs than when they stay mute: two years versus only one year.

As an effect, in the PD individual rationality leads to Pareto-inferior outcomes since both players could get higher payoffs under a different strategy profile. Both Don and Tom always want to testify; when they do, they are worse off than when they both stay mute.

In the generic version of PD, the strategy 'testify' is usually called 'defect' (D) and 'mute' is called 'cooperate' (C). The dilemma is between individual rationality that recommends defection and looking for social benefits from cooperation. Since individual rationality leads to inefficient outcomes, the PD provides an example of situation when an invisible hand doesn't work.

While the prison interpretation of the PD made the game attractive to present to a lay audience, it obscured its enormous importance and universal applicability. If the PD were a curiosity, its place wouldn't be at the center of political science and economics. And the applications of the PD are massive. PD was essentially the game played by the USA and the Soviet Union in the arms race. From applying PD we can learn that we do not need to assume that the opponents must be "evil": the structure of interactions forces them into defection. When we extend the basic model to many players, PD-like situations appear in global warming and other environmental problems, the formation of interest groups or the provision of all public goods and services. Even the most fundamental problem of humans forming a society versus total anarchy has arguably the structure of the PD!

The empirical problems resulting from the PD-like games motivated high-profile work in political science, economics, sociology and psychology that focused on modeling the tension between individual and social rationality. The situations involving such tensions were called collective action problems, the tragedy of the commons, social traps or social dilemmas (Olson 1965, Hardin 1968, Platt 1973, Dawes 1980). Especially important were extensions of the PD to multi-player games that retain its fundamental properties. PD can be extended in a variety of ways. In our next example, we will follow one possible path for one specific problem. Let's face the following problem of pollution in a big city:

> **Automobile pollution:** In Los Devils, where automobile pollution is a big problem, annual benefits from clean air are estimated to be equal to $1000 per person. A pollution-control filter costs $100. We know that:
> (a) the city's population is 1 mln and everybody has one car;
> (b) everybody pollutes equally and the benefits are proportional to the number of filters;
> (c) status quo payoff (no filters) is zero.

Every strategic game includes, as its components, players, strategies and payoffs. Let's reconstruct those components:

1. Players: 1 mln car owners;
2. Strategies: 1 (cooperate, buy filter), 0 (defect, do not buy filter);
3. Payoff of Player *i* (measured in dollars): $P_i(\mathbf{s}) = r(\mathbf{s})/1000 - 100\times s_i$, where $r(\mathbf{s})$ is the total number of cooperators when the strategy profile is **s** and s_i is Player *i*'s strategy.

Let's consider our player's payoffs from the two available strategies when the number of initial cooperators is *r*. When Player *i* defects, his payoff is equal to $r/1000$ since he doesn't have to buy the filter. When he decides to cooperate, his payoff is equal to $(r+1)/1000 - 100$ since he must pay 100 for the filter and he becomes one more cooperator. The difference between both payoffs is 99.999. The extra benefits to a player from his own cooperation (buying the filter) are practically negligible while the cost of cooperation is high. Thus, the incentive to defect is very strong. Defection is a dominant strategy, i.e., you are always better off if you defect.

Similarly to the two-player PD, not only both players have dominant strategies but also the equilibrium in dominant strategies is inefficient. To check the second fact let's denote the strategy profile of all cooperators ALL C and all defectors ALL D:

$$\mathrm{P_i(ALL\ C)} = 1000000/1000 - 100 = 900$$
$$\mathrm{P_i(ALL\ D)} = 0$$

When everybody plays the individually rational strategy 'defect', everybody gets the payoff of 0. When everybody cooperates, everybody's payoff is 900. The substantial difference represents a potential gain from cooperation that is forfeited in the inefficient Nash equilibrium (see Figure Automobile Pollution). The lower line in the figure represents payoffs from cooperation; the upper line represents payoffs from defection. The fact that we do not have the continuum of players as well as the continuum of values is neglected. The dots at the ends of lines mean that we cannot have 1 mln cooperators if Player *i* defects (in such a case 999999 is the maximum) or no cooperators if Player *i* cooperates (we must have at least one cooperator).

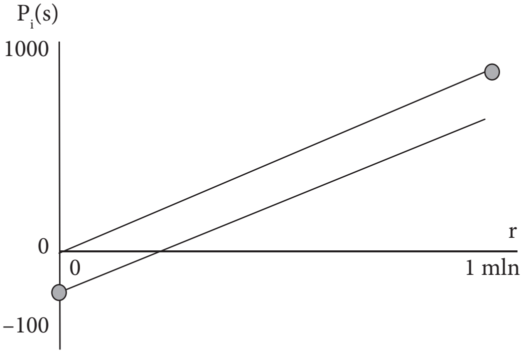


Figure 2: Automobile Pollution.

Automobile Pollution is another frightening version of the tragedy of the commons or a collective action problem. With a large number of players, achieving cooperation seems hopeless. Every player must pay the cost of cooperation but the benefits are spread thinly over a million.

From the PD perspective, it is surprising that humans so often cooperate. Cooperation may be facilitated by a variety of factors. One of them is establishing new rules of the game that essentially force players into cooperation. The rules may be imposed externally by a state or another central authority, or they may be self-imposed by interested players on themselves. Such new rules are called in political science and economics 'institutions.' The theoretical recognition of PD and similar games provided the decisive impulse for the development of modeling of institutions. One of the most profound contributions of political science was the discovery and description of the incredibly imaginative institutions that players themselves invent in order to get out of PD-like situations. In economics, increasingly larger parts of economic theory are being converted into the subfield of Industrial Organization that studies markets and other institutions that coordinate economic exchange.

In the case of Automobile Pollution, examples of game-changing institutions are easy to imagine. Sanctions may be imposed on car owners or car manufacturers that make buying a filter a dominant strategy. Products that are not environmentally safe may be banned from the market (so the strategy of defection may be effectively removed). Informal social pressure may be imposed on carmakers to manufacture only environmentally safe products. In general, we model new institutions as modified versions of the original ones that have some strategies banned, payoffs changed by sanctions, or that were subject to more complex transformations. Cooperation is possible, but the road to cooperation is far from straight and easy.

3. Repeated games and theoretical predictive impotence

In addition to institutional change, a process facilitating cooperation is the repetition of the game. It turns out that players may routinely cooperate in PD and similar games when the game is repeated. Repeated games are typically used to model evolutionary behavior, including the evolution of social norms or repeated economic exchanges. The developments in the theory of repeated games opened its own Pandora Box of surprising effects that deeply affected game-theoretic modeling.

In a finitely repeated PD, players play PD a fixed numbers of times and the final payoff is a sum of payoffs received in all rounds. In such a game, there is no dominant strategy, i.e., against certain strategies of the opponent there may exist better strategies than ALL D (always defect). Nevertheless, the strategy profile when both players play ALL D constitutes the unique Nash equilibrium of the game, or a situation in which no player can improve his payoffs by unilaterally changing his strategy. Such a strategy profile is the one that a game theorist would be quick to predict to happen. However, surprisingly and disturbingly, players seriously deviate from total defection. The first experiment with repeated PD (in which the PD was also introduced as a game) showed this phenomenon unambiguously (Flood 1952). Out of 100 rounds, the two subjects participating in the experiment cooperated on average in 73 rounds.

The phenomenon of cooperation in finitely repeated games hasn't been explained convincingly so far. One can easily design one's own simple experiment and run it on any audience with the game of Centipede (see Figure 3).

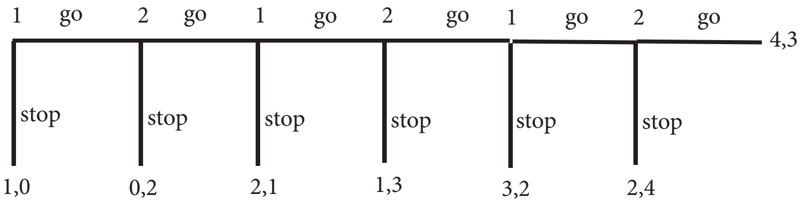


Figure 3: Centipede with three rounds.

Centipede may be considered a simpler version of the repeated PD. It can be described as follows: two players start a game with a small endowment: Player 1 starts with one dollar while Player 2 starts with nothing. They can always continue playing the game or stop it. When a player stops, both players collect the prizes that they have accumulated so far. If a player continues with the game, she must give up one dollar but the other player simultaneously receives two more dollars.

Centipede is an example of a *sequential game*. It can be solved by *backward induction* that finds all *subgame perfect equilibria*, a more restrictive concept than the Nash equilibrium. The backward reasoning is as follows: At the last stage, Player 2 wins four dollars by stopping and three dollars by going. Thus, he chooses stop. Player 1 can predict such action in her last move, and can represent her own choice as the one between receiving 3 for stopping and 2 for going (because of the predicted choice of Player 2). Thus, Player 1 stops in her last move as well. And so on: at every stage both players have incentives to stop. In the unique subgame perfect equilibrium of the game, Player 1 stops immediately, and in all possible decision situations both players would stop as well.[1]

The dilemma is quite similar to the PD. Both players receive in the subgame perfect equilibrium the payoffs of 1 and 0, respectively; if they cooperated, their payoffs would be 4 and 3, respectively. However, there is a difference between the Centipede and the one-shot PD. Experimental subjects, who are properly taught the rules of the game, tend to cooperate for a few rounds and then they stop one or two rounds before the end. Immediate stopping almost never happens!

[1] For a comprehensive evaluation and extensions of backward induction see Kaminski (2017, 2019)..

In many games, whenever interactions are repeated a finite number of times, cooperation miraculously appears. The troublesome problem for this phenomenon is that, while some explanations were offered, none is fully satisfying. Thus, the predictions made by solution concepts that we use in game theory to model rational behavior and the actual behavior in Centipede and similar games are at odds.

In infinitely repeated games, another problem emerges. Equilibria are numerous and, in many games and under reasonable assumptions, any level of cooperation may be achieved in equilibrium. In the 1970s, a number of disturbing theorems was proved that confirmed the existence of multiplicity of equilibria in repeated games. Those results were called "folk theorems" since game theorists long suspected that repeated games have similar properties.

Before we formulate one of the folk theorems, we have to introduce an infinitely repeated game created on a basis of a finite one-shot game. It is defined as follows: the players from {1, 2, …, *n*} are unchanged and non-empty strategy sets S_i (for i=1,…,n) from the original game are actions that are taken at every stage of repetition. The payoff functions are typically defined as a weighted sum of partial payoffs obtained at consecutive stages. Let *s* be a strategy profile such that at stage *k* players choose actions $s^k = (s^k_1 ,\dots , s^k_n)$. Then the partial payoff of player *i* at stage *k* is equal to $\mathrm{P}_{\mathrm{i}}(s^k)$, where P_i is player *i*'s payoff function in the original game. The total payoff in the entire repeated game is equal to an infinite sum of discounted partial payoffs:

$$\mathrm{P}_i(\mathrm{s}) = \Sigma_{\mathrm{i}=1,2,\dots}\ r^{i-1}\,\mathrm{P}_{\mathrm{i}}(s^k)$$

where $0<r<1$ is a discount parameter. Since the discount parameter is between zero and one, and the payoffs in a finite one-shot game are limited, the series always converges to a finite number.

While versions of folk theorems can be formulated for any finite game, repeated PD received most attention from game theorists and the formulation is especially simple in this case. Let *p* denote the general payoff from mutual cooperation (corresponding to -1 in Figure 1) and *q* denote the general payoff from mutual defection (corresponding to -2).

Folk Theorem [Friedman 1971]. In the repeated PD, when the discount parameter is sufficiently high, any vector of payoffs (x,x), where $q \leq x \leq p$, may be reached in a subgame perfect equilibrium.

In other words, by varying the discount parameter, we can find a full range of equilibria: from such that both players only defect to such that both players only cooperate! The equilibrium concept, as a predictor for what will happen in a game, is under such circumstances useless.

Folk theorems provided researchers with incentives to go beyond simple analysis of equilibrium existence and to model outcomes in repeated games with the use of computer simulation. Certain results, both empirical and theoretical, suggested that cooperation may be somehow privileged against defection. Axelrod (1984) published a book that, among others, included the results of his PD 'tournaments.' He asked a number of game theorists to submit computer programs generating strategies in the repeated PD, and then played them pairwisely against each other. The simplest examples of such strategies are ALL D (Always defect) and ALL C (Always cooperate) but a strategy can be much more complex since it can take into account the information about all previous moves of both players. A strategy TFT ("tit-for-tat") submitted by Anatol Rappoport (who was one of two subjects in the first PD experiment ever played) proved especially successful. TFT starts with cooperation and then repeats the previous action of the opponent, i.e., responds with C to C and with D to D. TFT won the first tournament, i.e, it received the highest average payoff, and then, even when its success was revealed to the participants, it won the second tournament. TFT is not a dominant strategy since no such strategy exists in a repeated PD. Moreover, because of its cooperative character it cannot even win against any other strategy in a pairwise contest. Nevertheless, it seems to be very efficient in generating high average payoff across large environments consisting of many various strategies.

The privileged status of cooperative strategies in Axelrod's experiments motivated researchers to include additional variables into modeling repeated interactions. One of such variables was the 'robustness' of a strategy, i.e., informally, the minimal proportion of such strategies in the population that guarantee the existence of equilibrium. Bendor and Swistak (1997) confirmed that equilibria based on cooperative strategies, such as the TFT, are most robust.

The less efficient the strategy (in terms of cooperation), the less robust it is, i.e., frequent defectors require higher proportions to defend against intrusion. This means that frequent defectors can be destabilized by a small change in the population. This is the essence of the evolutionary advantage of cooperation.

Despite all results suggesting a special character of cooperative strategies in repeated PD, the wide range of potential equilibria remains a disturbing phenomenon.

4. Social choice theory and the problems with aggregation of preferences

The motivating questions of social choice theory differ from those of game theory. Instead of asking 'what happens when rational players make independent decisions?,' social choice enquires about the properties of various methods of making social decisions, principles of distributive justice or the existence of methods satisfying certain properties (Arrow 1951, Sen 1969, Lissowski 2013). The seeds of social choice theory were in the work of a French mathematician, social scientist, and, incidentally, one of the fathers of the Enlightenment, marquis Nicolas de Condorcet. Condorcet studied the Estates-General, the French pre-revolutionary parliament and its three main blocs of voters constituted by clergy, nobility and the others: bourgeoisie, wage-laborers and the peasantry. He noticed that voter preferences of the three groups could form a very curious pattern. When paired with majority rule, such a pattern produced paradoxes.

Let's assume that we have three voters (or homogenous groups of voters of roughly equal size) that are numbered simply 1, 2 and 3. There are three alternative policies on crime that can be implemented *x*, *y*, and *z* (e.g., spending more on police force, building more prisons and spending more on prevention). The preferences of voters in the Condorcet profile are listed below from most preferred to least preferred alternative:

1: *xyz*
2: *yzx*
3: *zxy*

Under such preferences, Players 1 and 3 prefer x to y; 1 and 2 prefer y to z; 2 and 3 prefer z to x. Thus, for every alternative, there is a majority that prefers something else to it. We have a *cycle*:

$$x \text{ P } y \text{ P } z \text{ P } x$$

where $x\text{P}y$ means that a majority of voters prefers x to y.

When three alternatives are at stake, voting by majority is impractical since – as demonstrated by the Condorcet Paradox – a possible outcome would be that no alternative receives majority. Thus, voting by majority over three and more alternatives is potentially indecisive. Parliamentary procedures solve this problem by making consecutive alternatives compete with each other pairwisely according to a pre-specified agenda. However, such a solution raises a question: which agenda should be used? The question is of an utmost practical importance since the relation of social preference based on pairwise comparisons using majority is, as we established, intransitive in cases of a Condorcet profile.

If some agenda setter, say, the House Speaker, can specify the order of voting, he has a considerable power in hand. If the Speaker would like to make x the winner, he could easily create an agenda supporting his wishes. Turning x into a winner requires in this case simply that x is added as the last alternative to the agenda. The winning agenda for x looks as follows:

1. Vote between y and z;
2. Vote between the winner of Session 1 and x.

According to the above agenda, denoted formally as yzx, y wins in the first round against z, and loses in the second round to x. In our example, any alternative can win provided that it is the last one on the agenda.

The *Condorcet's Paradox* showed that the naïve interpretation of voting as representing "social will" or uncovering some underlying social interest is deeply suspect. Voting can be often manipulated, and there are fundamental reasons why voting cannot be considered to be as an "objective" method of making social decisions. Over years that followed, many other methods of voting were discovered and all of them were sooner or later found to have various troublesome or paradoxical properties. The next step in the understanding of such paradoxes was made when the contemporary discipline of social choice was created in mid-20th century by Duncan Black (1958) and Kenneth Arrow (1951). Arrow's work is especially relevant for the present article.

Arrow was a student of a great Polish logician and mathematician Alfred Tarski. Arrow's language developed for social choice theory was based on Tarski's terminology used in his popular logic textbook. Arrow claimed that he was not familiar with the Condorcet's paradox when he started his work but his approach could be considered the natural logical extension of Condorcet's problem.

Let's assume that a social decision must be made over some number of at least three social alternatives. We want to rank the alternatives from the best to the worst one, i.e., we want to order them in a *transitive* fashion (we denote this condition by T) using some social decision method F. The only information that can be used are preferences of some finite number (at least two) voters who may be also indifferent between or among some alternatives. An example of such a method would be a dictatorial rule of some voter *i* that simply says "whatever *i* prefers is also preferred by a society." Certainly, such a voting method wouldn't be acceptable for most social decisions. We are looking for such an F that wouldn't be dictatorial and perhaps satisfy a few simple properties as well. Arrow singled out the following properties:

1. Unrestricted domain (U): Our method F is defined for all possible voting profiles (configurations of voter preferences);

2. Pareto property (P): If everybody prefers *x* to *y*, then *x* must be socially preferred to *y*;

3. Non-dictatorship (D): F is non-dictatorial, i.e., there is no dictator among the voters. A voter *i* is a dictator according to F if for any *x* and *y*, whenever *i* prefers *x* to *y*, then F must prefer *x* to *y*;

4. Independence of Irrelevant Alternatives (IIA): F is based on pairwise comparisons only, i.e., if in two voting profiles the individual preferences over *x* and *y* are identical, then F must rank *x* and *y* socially the same way.

Arrow's Theorem [1951] : Conditions U, P, D, IIA listed above and the requirement of transitivity of social ranking T are inconsistent.

At least one of the five properties must be violated. For instance, majority rule applied to all alternatives violates U since for some preference profiles – such as the Condorcet's profile – there are no majority winners, and we cannot even designate the top position in the social ranking. If we modify majority by using certain pre-defined agenda and, possibly in some cases, a tie-breaker between alternatives, we can satisfy U but IIA will be violated. In general, practically all sensible voting methods violate IIA.

Arrow chose conditions that were looking "obvious" and "simple" (his original conditions were slightly different from those defined above that became later standard in the presentation of Arrow's Theorem). His point was that making social decisions is a complicated and troublesome process, far from the naivety of "social physics" and automatic rationality. 'Social rationality' cannot be modeled by a mechanical utilization of metaphors of optimization taken from physics and other natural sciences. No 'social physics' is possible. If we accept Arrow's conditions as fundamental, then every method of social decision-making must have certain fundamental deficiency or deficiencies. This pessimistic statement applies not only to voting but to all other social decisions that are based on individual preferences and result in rankings, including various policy decisions or welfare comparisons of welfare economics.

Arrow's work not only won him a Nobel Prize but also started a new discipline, where scholars uncovered a large number of similar *impossibility theorems*. Probably the most interesting and important of such results is a theorem anticipated by a philosopher Allan Gibbard (1973) and formally proved by a mathematical economist Mark Satterthwaite (1975). The Gibbard-Satterthwaite Theorem states formally what many social choice theorists suspected since Arrow: practically all voting methods are vulnerable to manipulation. Thus, this theorem extends the main point of the Condorcet's Paradox to all voting methods.

Let's assume that a voting method V is based on individual preferences of a finite number (at least two) of voters over at least three alternatives. In this case, V produces not a ranking but a single winner, i.e., a single alternative. We make the following assumptions about V:

1. Unrestricted domain (UD): V is defined for all possible voting profiles;

2. Non-dictatorship (ND): there is no voter *i* such that if *i*'s most preferred alternative is *x*, then *x* must become the winner selected by V;

3. Range constraint (RC): There are at least three alternatives that, under certain preference profiles, are selected by F as winners.

Out of the three properties, the two first ones are straightforward. The third one demands a word of explanation. The requirement of RC is in fact less demanding than the Pareto condition that appeared in Arrow's Theorem. The Pareto condition would demand in the present context that for every alternative *x*, whenever *x* is preferred unanimously to everything else, then *x* would be the winner. This would mean that every alternative must be sometimes a winner and also specify conditions when this must happen. RC demands only that we have at least three different winners, and assumes nothing about the circumstances under which this must happen.

Gibbard-Satterthwaite Theorem [1975]: If the three conditions UD, ND and RC are satisfied by V then V must be *manipulable*.

Manipulability means that there exists a preference profile such that a certain voter or voters have incentives to lie about their preferences since this misrepresentation would result in the choice of their more preferred alternative. In a less normatively loaded terminology, the type of voting resulting from misrepresenting one's preferences is called *sophisticated* or *strategic*.

Let's take a look at our example of Condorcet's Paradox and the agenda *yzx* that we used in order to make *x* the winner. Let's recall the voting profile from our example:

1: *xyz*
2: *yzx*
3: *zxy*

The agenda was: vote between *y* and *z*, then vote between the winner of the first round and *x*.

Voter 1 has certainly no incentive to vote strategically since *x* is this voter's top choice. It is left to the reader to check that also voter 3 has no such incentives. With voter 2, the situation is different. Let's assume that only voter 2 can vote strategically. If 2 votes in the first round for *z*, his second choice, instead of his top choice *y*, such a vote would make *z* the winner of the first round. When *z* makes to the second round, it beats *x* and becomes the overall winner. Thus, by voting strategically, voter 2 was able to make the winner his second-best choice *z* instead of his worst alternative *x*. Gibbard-Satterthwaite's Theorem assures us that for all voting methods that satisfy the truly basic conditions 1-3, we can find voting profiles that are similarly manipulable.

Like Arrow's Theorem, Gibbard-Satterthwaite's Theorem generated a wave of work studying various types of manipulation or paradoxes. It turns out that problems with the agenda (as those exhibited in the Condorcet Paradox), the ubiquity of strategic voting, the manipulation via introducing fake candidates or vote trading, gerrymandering (strategic redistricting) or other electoral engineering are unavoidable aspects of politics. We cannot free ourselves from lying and manipulation simply through electing better candidates for our offices. The incentives for manipulation are present due to the nature of social decisions and there will always be politicians who will not miss the opportunity.

5. How research methodology and policymaking were affected?

The gradual dismantling of Enlightenment optimism about social rationality that happened in the middle of 20th century had substantial consequences for the methodology of research studying optimal social outcomes and the evaluation of policy outcomes. We can classify methodological reactions as *business as usual*, pessimistic *resignation*, *focus on institutions* and *fragmentation*. Below I will give examples of all types of reactions.

An example of the *business as usual* reaction is a cost-benefit analysis (CBA) that essentially employs the Bergson-Samuelson approach to evaluating policy outcomes. CBA used in public sector examines the costs and benefits associated with various policy projects. For instance, the construction of a new highway may be evaluated taking into account direct costs, environmental impact, pollution, disturbances for affected people and various types of economic benefits. In short, CBA's appropriateness stems from using monetary estimates that can be justified as well-measurable and additive across affected parts. While problems discovered by Arrow still apply, the crucial Arrowian axiom of IIA (see Section 4) loses much of its appeal in this context.

Resignation followed the widespread pessimism about dismal prospects for collective action (Olson 1965) and the inevitability of the “tragedy of the commons” (Hardin 1968). A possible example – admittedly, a non-falsifiable hypothesis – are the problems of NATO with enforcing cooperation among its members. Olson and Zeckhauser’s (1966) influential article examined the difficulties of NATO with making its members to raise defense spending to the level of percent of GDP comparable with the United States. The pessimistic conclusion was that a “large” player (United States) is doomed to be cheated by “small” players (other members of NATO). As a consequence, the United States cannot avoid making disproportionately large contributions to the common cause. Motivated by the authors’ desire to show a persuasive example of a collective action problem (Zeckhauser 2015), the article quickly became a required reading for practically all graduate students of political science and public policy in the United States. Its pessimism very likely discouraged American government officials and policy analysts from pressing the allies to increase their defense spending.

Nudging NATO allies to spend more on defense by President Donald Trump provides a good example of the next reaction to the pessimism of Olson and Hardin, i.e., *institutionalism*. Trump, the first American president with no prior governmental or military experience, was probably not aware of the “impossibility” of solving the NATO’s spending dilemma. He threatened with withdrawal of American troops from Europe; moving American bases to higher-spending allies; offered public shaming and cold shoulder to Chancellor Merkel of Germany; threatened with tariffs on consumer goods; made an impression that he was a “mad” decision maker. All those attempts at solving the problem were institutional in nature, i.e., he attempted to change the rules of the game in order to push the outcome in the desired direction.

In general, the explicit focus on institutions (described in more detail at the end of Section 2) has generated several Nobel Memorial Prizes in Economics over the past decades starting with Ronald Coase (1991), Douglass North (1993), and Oliver Williamson and Elinor Ostrom (2009). Especially the work of Ostrom (1990) provided brilliant theoretical and empirical arguments against the pessimism of Olson and Hardin. While a common resource can be overexploited, Ostrom demonstrated how humans in small, local communities are surprisingly successful in changing the initial rules of game when managing fisheries, pastures, oil fields, irrigation systems or forests.

By *fragmentation* I mean the substitution of the concept of a measurable social good with various piecewise analyses that became especially popular in voting theory. Since the objective of finding the "best" voting method was unattainable, research strategies evolved towards a less ambitious goal of partial evaluation of such methods. When Arrow's work gained recognition, a popular research strategy became axiomatic analysis adopted from logic. Among its proponents, Amartya Sen received the 1998 Nobel Memorial Prize in Economics for related work while Michel Balinski and Peyton Young's results were instrumental for applying the axiomatic method to proportional representation algorithms.

William Riker (1982) famously refuted the "populist" concepts of looking for "best" politicians or implementing the "social will," and defended "liberalism" defined as merely rejecting the worst. Another fundamental research strategy in voting theory – motivated to some extent by Riker's minimalist idea of rejecting bad options – focused on the evaluation of the extent of paradoxes and problems by using computer simulation (Dougherty 2011). While no ideal voting method exists, one can try mapping the frequency of a particular problem. For instance, voting theorists confirmed by computer simulation that simple plurality method (a candidate with most votes wins) often generated paradoxical and even dangerous results in presidential elections. Thus, having a second round in such elections or using other methods was well justified.

6. Conclusion

Developments in mathematics of the second half of 20th century, especially game theory and social choice theory, demolished many of the Enlightenment's myths of a rosy society being on an auto-pilot of progress. The discovery of fundamental problems underlying the concept of social rationality gave impulse to, among others, the rise of institutional analysis that became central for political science and economics, the use computer simulation that helps deal with many questions that are not easy to treat analytically, and the ascent of experimental methods investigating real-world decision-making.

Game theory, social choice theory and related mathematical approaches owe their modeling success to the precision of their mathematical tools and minimalistic assumptions. For instance, in strategic games described in Section 2, player identities are unimportant (they may be "bakers", "butchers" but also "states" or "voters") and strategies may come from any nonempty set; any preferences over strategy profiles are allowed. This model applies specifically to players simultaneously making independent choices (or equivalent situations), and having well-defined preferences over different outcomes – and to all such cases. We can ask questions such as "Is there any outcome such that no player wants to unilaterally change his/her strategy?" An answer will apply universally to all relevant empirical cases. Other related questions, such as dealing with coalitional opportunities or repeated play, require making additional assumptions in our model or using a different formalism. This "minimalism and precision" approach led to a development of a number of alternative modeling frameworks such as many types of noncooperative and cooperative games, repeated games, games with incomplete information, etc. This allowed for shifting the attention from providing "one explanation for all social phenomena" towards accurate matching empirical phenomena with relevant models.

The Prisoner's Dilemma and its generalizations, analyzed within the simplest framework of strategic games, destroyed the faith that markets are always effortlessly efficient and optimal. In a variety of human interactions, unconstrained decisions lead to such inefficiencies as overexploitation, undersupply or arms races. Reaching optimality via changing the rules of game, while not impossible, may be a difficult and time-consuming task.

Developments in repeated games and evolutionary game theory highlighted another troublesome aspect of social interactions. In repeated PD and other games, any level of cooperation may be sustainable in equilibrium. A troublesome conclusion may come to mind that our predictive power in such cases is impotent and "anything may happen." Surprising good news is that more cooperative strategies, such as TFT, have some evolutionary edge over less cooperative ones.

Finally, social choice theory revealed that the very existence of socially optimal states couldn't be taken for granted. Under typical conditions, no social decision rule satisfies all basic and reasonable properties that we might believe should be satisfied. Also, practically all voting rules are also vulnerable to manipulation. Whenever we make social decisions, we have to accept – consciously or not, whether we like it or not – tradeoffs between fundamental values and principles.

The pessimism about social rationality inherited with game-theoretic and social-theoretic developments had multifaceted impact on research strategies and methodology of policy making. While methods such as cost-benefit analysis remained largely unaffected, in other settings widespread pessimism labeled certain non-optimal outcomes as inevitable. Nevertheless, institutional analysis indicated that while it may be impossible to switch from an equilibrium in a game to a non-equilibrium outcome that is Pareto-superior, it is possible to change the game itself into one that would generate better outcomes. Finally, the by-product of pessimism associated with Arrow's Theorem was the emergence of the axiomatic method and computer simulation methods that instead of social optimality, investigated specific desirable properties of social decision rules.

References

Arrow, Kenneth J. 1951, 2nd ed. 1963. *Social Choice and Individual Values*. New York: Wiley.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Balinski, Michel L., and H. Peyton Young. 1982. *Fair Representation: Meeting the Ideal of One Man, One Vote*. Brookings Institution Press (2nd ed., 2010).

Bendor, Jonathan, and Piotr Swistak. 1997. The Evolutionary Stability of Cooperation. *American Political Science Review* 91: 290-307.

Bentham, Jeremy. 1780. *An Introduction to the Principles of Morals and Legislation*. London: T. Payne and Son. http://www.econlib.org/library/Bentham/bnthPML.html (online reprint of original edition).

Black, Duncan. 1986 (first ed. 1958). *The Theory of Committees and Elections*: Springer.

Buchanan, James M. 1965. "An Economic Theory of Clubs." *Economica* 32 (125): 1–14.

Dawes, Robyn M. 1980. Social dilemmas. *Annual Review of Psychology* 31 (1):169-93.

Dougherty, Keith L., and Julian Edward. 2011. *The Calculus of Consent and Constitutional Design*. Studies in Public Choice. Springer New York.

Flood, Merrill M. 1952. Some Experimental Games. Research Memorandum RM-789. Santa Monica, CA: RAND Corporation.

Friedman, James W. 1971. A Non-cooperative Equilibrium for Supergames. *Review of Economic Studies* 38: 1-12.

Gibbard, A.S. 1973. Manipulation of Voting Schemes: A General Result. *Econometrica* 41: 587-602.

Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162: 1243–48.

Hardin, Russell. 1982. *Collective Action*. Baltimore: Johns Hopkins University Press.

Kaminski, Marek M. 2017. Backward Induction: Merits and Flaws. *Studies of Logic, Grammar and Rhetoric* 50(63): 9-24.

Kaminski, Marek M. 2019. "Generalized Backward Induction: Justification for a Folk Algorithm." Games 10.3: 34.

Lissowski, Grzegorz. 2013. *Principles of Distributive Justice.* Opladen: Barbara Budrich Publishers.

Mandeville, Bernard. 1989 (first ed. 1705). *The Fable of the Bees: Or Private Vices, Publick Benefits*: Penguin Classics.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Olson, Mancur, and Richard Zeckhauser. 1966. "An Economic Theory of Alliances." *The Review of Economics and Statistics*, 266–279.

Ostrom, Elinor. 1990. *Governing the Commons*. Cambridge university press.

Pareto, Vilfredo. 1906. *Manual of Political Economy*. Augustus M. Kelley (English translation).

Platt, John. 1973. Social Traps. *American Psychologist* 28:641-51.

Riker, William H. 1982. *Liberalism against Populism*. San Francisco: WH Freeman.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *The Review of Economics and Statistics*, 387–389.

Satterthwaite, Mark. 1975. Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory* 10: 187-217.

Sen, Amartya K. 1970. *Collective Choice and Social Welfare.* San Francisco: Holden-Day.

Smith, Adam. 2009 (first ed. 1776. *The Wealth of Nations*. Edited by T. Books. London.

Tucker, Albert W. 1950. A two-person dilemma. Mimeographed paper. : Stanford University.

von Neumann, John. 1928. Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen* 100: 295-320.

von Neumann, John, and Oskar Morgenstern. 1944. *The Theory of Games and Economic Behviour*. Princeton: Princeton University Press.

Zeckhauser, Richard. 2015. "Mancur Olson and the Tragedy of the Unbalanced Commons." *Decisions (special issue ed. by Marek M. Kaminski)*, no. 24: 191–202.

Marek M. Kaminski
Department of Political Science and Institute for Mathematical Behavioral Sciences,
University of California,
3151 Social Science Plaza, Irvine, CA 92697-5100, U.S.A.; email: marek.kaminski@uci.edu.

STUDIA METODOLOGICZNE
NR 39 • 2019, 195–232
DOI: 10.14746/sm.2019.39.8

Elay Shech, Axel Gelfert

The Exploratory Role of Idealizations and Limiting Cases in Models

Abstract. In this article we argue that idealizations and limiting cases in models play an exploratory role in science. Four senses of exploration are presented: exploration of the structure and representational capacities of theory; proof-of-principle demonstrations; potential explanations; and exploring the suitability of target systems. We illustrate our claims through three case studies, including the Aharonov-Bohm effect, the emergence of anyons and fractional quantum statistics, and the Hubbard model of the Mott phase transition. We end by reflecting on how our case studies and claims compare to accounts of idealization in the philosophy of science literature such as Michael Weisberg's three-fold taxonomy.

Keywords: exploration, idealization, models.

1. Introduction

Idealizations and the use of models, which are by their very nature imperfect or highly fictitious representations of reality, are ubiquitous in science.[1] How is one to make sense of the fact that, in attaining empirical adequacy and giving us knowledge about the world, our best scientific theories invoke falsehoods and distortions of reality? A standard, albeit naïve, response to such a worry has been not to allocate any substantive role to idealizations and

[1] Some examples of idealizations include nonviscous fluid flow, a perfect vacuum, perfectly rational agents, and isolated populations, while examples of (idealized) models include the Ising model, the Hardy-Weinberg equilibrium model, and Schelling's segregation model. See Shech ([2018a]) for a related review article.

models. Bluntly put, idealizations and models are used to simplify and abstract away irrelevant details, render computationally tractable various systems of study, or else are taken as auxiliary tools for application of theory. In principle, so the argument goes, idealizations and models can be dispensed with.[2]

In contrast, many philosophers of science have attempted to articulate substantive roles for idealizations and models to play in science, with the emphasis having largely been placed on explanation. Our goal in this paper is to build upon recent work and join this latter camp.[3] However, whereas previous authors have concentrated on the explanatory role, we wish to fill what we take to be a missing gap in the literature and stress the exploratory roles of idealizations and models.[4]

In particular, we will present three case studies that illustrate our claims including the Aharonov-Bohm effect, the emergence of anyons and fractional quantum statistics, and the Hubbard model of the Mott phase transitions (Sections 2-4). Although we do not intend for our list to be exhaustive, we submit that idealizations and models can be exploratory in at least four substantive manners: they may allow for the exploration of the structure and representational capacities of theory; feature in proof-of-principle demonstrations; generate potential explanations of observed (types of) phenomena; and may lead us to assessments of the suitability of target systems.[5] Last, we conclude the paper by comparing our case studies with Michael Weisberg's ([2007], [2013]) recent taxonomy of idealizations and models (Section 5). We argue that his three-fold classificatory scheme is lacking in that it does not make room for the exploratory role of idealizations and models, thereby offering a distorted view of the case studies that we present.

[2] See Norton ([2012]) for a recent defense of the claim that idealizations ought to be dispensed with.

[3] For instance, see Batterman ([2002]) for a discussion of explanatory idealizations, and Batterman and Rice ([2014]) and Bokulich ([2008]) for the explanatory role of imperfect models.

[4] Similar themes have been explored by, among others, Redhead ([1980]), Bailer-Jones ([2002]), Yi ([2002]), Wimsatt ([2007], and Ruetsche (2011, p. 337). See Gelfert ([2016], [2018]) and Massimi ([2018]) for the exploratory uses of scientific models, and see Earman [2017] and Shech ([2015a], [2015b], [2016], [2017], [2018a], [2018b]) for exploratory idealizations.

[5] This list partially follows Gelfert's ([2016], pp. 83-94) fourfold distinction of exploratory functions of models.

A caveat is in order before beginning. We do not endeavor to define what idealizations and models are. The literature on these questions is vast, and ultimately not much will be at stake for our purposes.[6] Instead, we will appeal to a generic understanding of these notions, on the assumption that, whatever one's preferred account of idealizations and models, the proposal developed in the present paper can be adapted accordingly. This means that at times we will allow ourselves to talk about an 'idealization,' or an idealized system or object, and a 'model' interchangeably since both idealizations and models, insofar as they are used to represent physical phenomena, are misrepresentations of sorts.

2. The Aharonov-Bohm Effect

2.1. Case Study: AB Effect

Consider a standard double-slit experiment undertaken with a beam of electrons. Experiments have shown that electrons manifest a behavior consistent with wave interference patterns (see Figure 1). Now add to this configuration an infinitely long and absolutely impenetrable solenoid (in between the double-slit screen and the detector screen) (see Figure 2). If we turn on the solenoid, what type of behavior should we expect to witness? Intuitions may vary on this point, but there is a straightforward sense in which no answer can be given: we cannot ever build an apparatus with an infinitely long and absolutely impenetrable solenoid, so we cannot know what would happen in such a scenario. However, the question can be answered within the context of a theory. For instance, if we take our thought experiment to manifest in a world governed by classical physics, there is no reason to think that anything will happen. Ac-

[6] For more on idealizations see Weisberg ([2007], [2013]), Ladyman ([2008]), Elliott-Graves and Weisberg ([2014]), Shech [2018a], and Fletcher *et al.* [Forthcoming], and for more models see Morgan and Morrison ([1999]), Frigg and Hartmann ([2012]), and Gelfert ([2016]). See Norton ([2012]) for more on the distinction between idealization and approximation, and see Jones ([2005]) for more on the distinction between idealization and abstraction. Psillos ([2011]) differentiates between the process of idealization/abstraction and the idealized/abstracted system/model that is the product of such a process. Also see Shech ([2015b], [2016]) for more on misrepresentation and depiction.

cording to the setup the solenoid is infinitely long so that the magnetic field ***B*** produced is wholly confined to a region S_{in} inside the solenoid. The solenoid is also absolutely impenetrable, so that the beam of electrons is completely confined to a region S_{out} outside the solenoid. Since there is no local (physical or causal) interaction between the electrons and the magnetic field, classical physics makes no novel prediction about this particular idealized system.

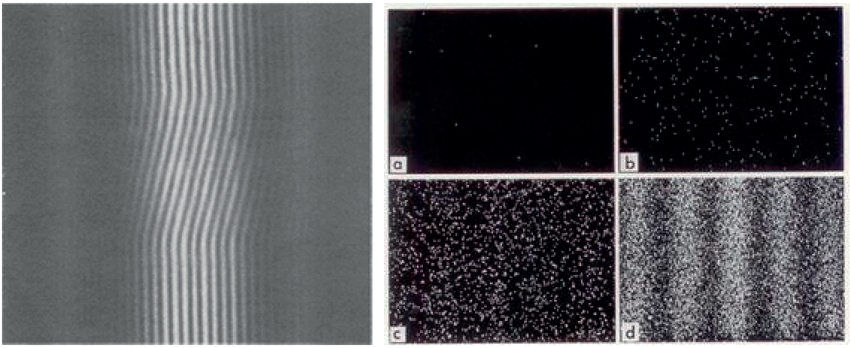


Figure 1. (Left) An example for an interference pattern from a double-slit experiment (from Möllenstedt and Bayh 1962, 304). (Right) Single-electron build-up of (biprism) interference pattern (from Tonomura [1999], p. 15). (a) 8 electrons, (b) 270 electrons, (c) 2000 electrons, and (d) 60,000 electrons.

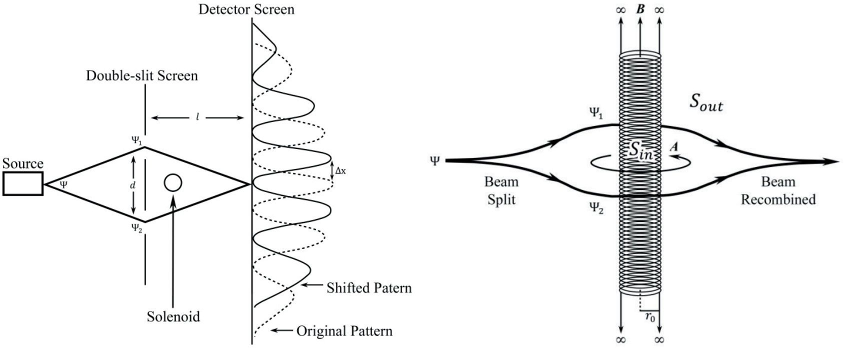


Figure 2. The AB effect. A beam of electrons Ψ is split in a region S_{out}, made to encircle a solenoid (that generates a magnetic field inside the region S_{in}), and then to recombine on a detector screen. The original interference pattern is shifted by an amount Δx.

In stark contrast to these classical intuitions, Yakir Aharonov and David J. Bohm ([1959]) showed that quantum mechanics predicts a shift in interference pattern, which has become known as the (magnetic) Aharonov-Bohm (AB) effect.[7] In modeling the idealized scenario, they began with the standard Hamiltonian used for a charged particle in electromagnetic fields: $H^I_{AB} = (\boldsymbol{P} - q\boldsymbol{A}/c)^2 / 2m$, where m and q are the electron mass and charge, respectively, $\boldsymbol{P} = -i\hbar\nabla$ the momentum operator, $\boldsymbol{A}$ the electromagnetic vector potential operator generating the magnetic field such that $\boldsymbol{B} = \nabla \times \mathbf{A}$, and the electromagnetic scalar potential has been set to zero. Since S_{in} is a region inaccessible to the beam of electrons represented by the quantum state Ψ, H^I_{AB} acts on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^3 \setminus S_{in})$ of square-integrable functions defined on a non-simply connected configuration space $\mathbb{R}^3 \setminus S_{in}$, that is, on three dimensional Euclidean space from which the interior of the solenoid has been excised. This means that H^I_{AB} is not a self-adjoint operator, and so it does not generate the dynamics of the system. In order to remedy the situation, Aharonov and Bohm ([1959]) chose a unique self-adjoint extension of H^I_{AB}, symbolized by $\bar{H}^I_{AB} = (\boldsymbol{P} - q\boldsymbol{A}/c)^2 / 2m$, which is picked out by Dirichlet boundary conditions in which the wavefunction vanishes at the solenoid boundary (i.e., $\Psi = 0$ at the boundary). One can then derive the shift in interference pattern by calculating the relative phase factor $e^{i\theta}$ between the two components of the wave function, Ψ_1 and Ψ_2, as is done in standard textbooks, e.g., Ballentine ([1998], p. 321-325).[8]

2.2. Exploration in the AB Effect

The first sense of exploration that we wish to consider is exploration of the structure of a scientific theory. It is by making use of an idealization, viz., an infinitely long and absolutely impenetrable solenoid, and appealing to the corresponding idealized model H^I_{AB}, that Aharonov and Bohm ([1959]) were

[7] See Peshkin and Tonomura ([1989]) for more on the theory of the AB effect and its experimental confirmation.

[8] But see Shech ([2017]) and Earman ([2017]) for further intricacies regarding the derivation of the AB effect.

able to highlight one of the blatant contrasts between classical and quantum physics: the two theories have a very different structure in that they make vastly different predictions about an idealized model – the model representing the behavior of electrons in the vicinity of a shielded magnetic field. Moreover, exploring the quantum physics of infinite and absolutely impenetrable solenoids is what allowed Aharonov and Bohm ([1959]) to discover a possibly additional manifestation of non-locality in quantum mechanics,[9] since the electrons exhibit a dependency on the magnetic flux while remaining in a region devoid of any such flux. In other words, in this case study, idealizations (in the form of an infinitely long and absolutely impenetrable solenoid) played in indispensable role in exploring the modal structure of non-relativistic quantum mechanics in order to shed light on foundational issues (e.g., locality) and intertheoretic relations.[10]

One may object that it is misleading to talk about the AB effect as an exercise in theoretical exploration via idealization since, in fact, experiments have shown that the AB is a real, physical effect (Tonomura et al. [1986]). In reply, we draw an analogy with Shech's ([2013], pp. 1172-1173) distinction between concrete and abstract phase transitions: concrete phase transitions are the sharp but continuous changes that arise in various thermodynamic

[9] See Healey ([1997], [1999]) and Maudlin ([1998]) for a debate about whether or not the AB effect portrays a type of quantum non-locality comparable with Bell inequalities. In this paper, we shall refrain from making any comment on this issue.

[10] Compare with Massimi's ([2018, p. 339]) discussion of 'perspectival modeling' (original emphasis): But what makes 'perspectival models' stand out in the broader class of exploratory models is a particular way of modeling possibilities … I contend that perspectival models are an exercise in imagining, or, to be more precise, physically conceiving something about the target system so as to deliver modal knowledge about what might be *possible* about the target system. In a way, they perform hypothetical modeling but of a distinctive modal type – they model either epistemic or objective modalities about the target system (within broad experimental and theoretical constraints). And this is also the reason that sets them aside from phenomenological models, in general, which are designed to model data or phenomena known to exist and be actual (indeed phenomenological models are designed to model observed occurrences rather than possibilities, as is the case with perspectival models).

We are clearly sympathetic to such a point of view and only add that the type of modal exploration that 'perspectival models' facilitate—models that are also abstract and/or highly idealized—fits well with our first sense of exploration viz., 'exploring theoretical structure and representational capacities.'

potentials and may be observed in the laboratory. Abstract phase transitions, i.e., phase transitions as they are conventionally and theoretically defined, are discontinuous changes governed by a non-analytic partition function that are used to mathematically represent concrete phase transitions (See Figure 3).

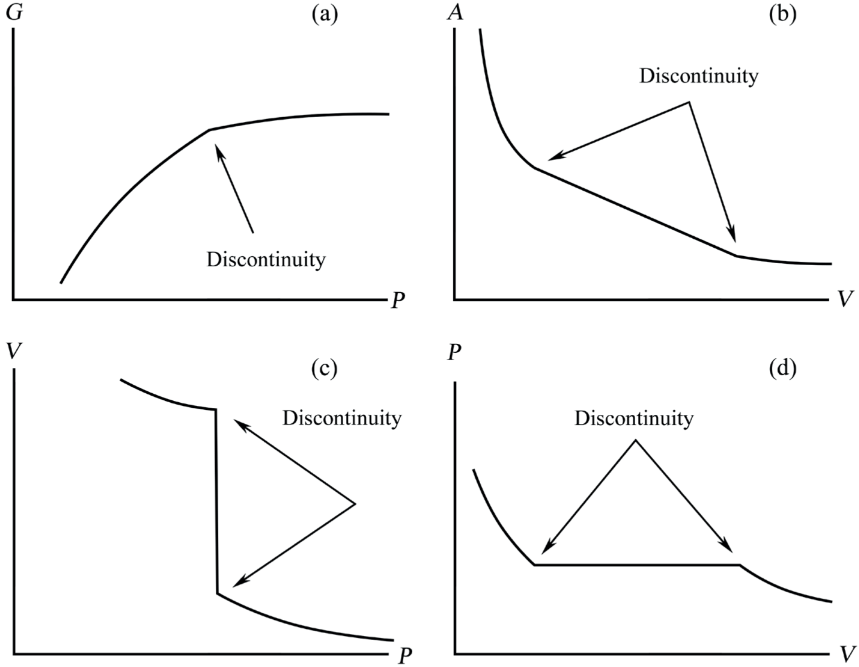


Figure 3. Graphs displaying a first-order phase transition. Graph (a) displays the Gibbs free energy (or Gibbs thermodynamic potential) G as a function of the pressure P, graph (b) displays the Helmholtz free energy (or Helmholtz thermodynamic potential) A as a function of the volume V. Graphs (c) and (d) display functional relations between P and V. Based on Stanley ([1971], p. 31).

Abstract phase transitions are defined in idealized infinite systems through the thermodynamic limit, in which a system's volume and particle number diverge.[11] Similarly, we must make a distinction between two kinds of (magnetic) AB effects. On the one side, the abstract AB effect as it is conventionally defined applies only to idealized systems where there is a strictly null intersection between the regions occupied by the electron wavefunction and the magnetic flux. It cannot, in principle, ever manifest in the laboratory, and yet it plays an exploratory role in the senses discussed in this section. On the other side, there is the concrete AB effect that has been empirically confirmed and shows that a beam of electrons exhibits a type of quantum dependency on magnetic flux that is unaccounted for by classical physics. Only recent rigorous results in mathematical physics have shown that the abstract AB effect is a good approximation of the concrete one.[12]

A second sense of exploration that may be brought about through the consideration of highly idealized models concerns generating potential explanations, for instance, by envisaging scenarios that, if true, would give rise to the kinds of phenomena that constitute the explanandum.[13] Given the odd nature of the AB effect as a possibly non-local effect, and the fact that the idealization of an infinitely long and absolutely impenetrable solenoid cannot be instantiated in reality, it is not surprising that early claims of experimental verification (e.g., Chambers [1960], Tonomura et al. [1982]) were met with skepticism (e.g., Bocchieri and Loinger [1978]). Attempts to understand and explain the effect and its experimental manifestation took various forms, including potential explanations given within the fiber bundle formalism of electromagnetism.[14] In this context, the electromagnetic fields are represented by the curvature of, and the electromagnetic vector potential is represented by a connection on, the principal fiber bundle appropriate for the formulation of classical electromagnetism (see Table 1). That is to say, a principal bundle where the base space is the spacetime manifold and where the structure group is the group of rotations in the complex plane $U(1)$. The relative phase factor

[11] Compare with Kadanoff ([2000], p. 238): 'The existence of a phase transition requires an infinite system. No phase transitions occur in systems with a finite number of degrees of freedom.' See Stanley ([1971]) and Kadanoff ([2000]) for the theoretical treatment of phase transitions.

[12] See Ballesteros and Weder ([2009], [2011]) and de Oliveira and Pereira ([2008], [2010], [2011]) for such results.

[13] See, e.g., Gelfert ([2016], pp. 87-98).

[14] See Healey ([2007], Ch. 1-2) for an introduction.

$e^{i\theta}$ which, according to the theory, gives rise to shifted interference pattern that is the AB effect arises as the non-trivial holonomy of a closed curve encircling the solenoid.

Table 1. Comparison of terminology between non-relativistic quantum mechanics and the fiber bundle formulation of the AB effect.

	Electro-magnetic Vector Potential	Magnetic Field Produced by Solenoid	Shift In Interference Pattern (due to a Relative Phase Factor)	Space or Spacetime
Non-Relativistic Quantum Mechanics Formulation	$\boldsymbol{A}$	$\boldsymbol{B}$	$e^{i\theta} = \exp\left(\frac{iq}{\hbar}\oint_C \boldsymbol{A}\cdot d\boldsymbol{r}\right)$	$\mathbb{R}^3$ Or $\mathbb{R}^4$
Fiber Bundle Formulation	Connection	Curvature	Non-trivial Holonomy	Base Space

It is now possible to generate a potential (although non-actual) explanation of the AB effect. In particular, one may arrive at a non-trivial holonomy by considering a fiber bundle base space that is non-simply connected. The rationale for this explanation is that vanishing electromagnetic fields around the solenoid correspond to a curvature that is zero. Zero curvature means that 'the connection on this bundle is flat everywhere in this region' (Healey [2007], p. 42). Moreover, if 'there is a nontrivial holonomy . . . and if the connection is flat, the base space [representing physical space] must be nonsimply connected' (Batterman [2003], p. 542; original emphasis). In other words, a non-simply connected base space, which represents physical space (as opposed to the electron configuration space), also allows one to derive a non-trivial holonomy. However, while a derivation based solely on such topological considerations may be considered a potential explanation of the non-trivial holonomy, it is not the actual explanation of the non-trivial holonomy that represents the AB effect. After all, the AB effect is a dynamical effect that depends on the interaction between the electron beam and the solenoid (not on holes in physical space or on a particular mathematical formalism). The upshot is that considerations of the highly idealized (abstract AB effect) model, within the fiber bundle formalism, have allowed us to discover a potential explanation of the non-trivial holonomy in terms of a non-simply connected base space.

This concludes our discussion of the AB effect in which we emphasized two senses of exploration: exploring the modal structure of a theory for the purposes of gaining insight into foundational issues and intertheoretic relations, and generating potential explanations.

3. Anyons and Fractional Quantum Statistics

3.1. Case Study: Anyons

Consider a collection of non-interacting, identical particles in thermal equilibrium. What are the possible ways that such a collection may occupy a set of available discrete energy states? Roughly, quantum and statistical mechanics tell us that there are two such ways, and that the expected number of particles in some specific energy state will depend of the type of particle at hand. Bosons manifest a behavior consistent with Bose-Einstein statistics, while fermions distribute themselves according to Fermi-Dirac statistics. This division into particle types, along with corresponding statistics may be captured by what has become known as the symmetrization/anti-symmetrization postulate: 'The states of a system containing *N* identical particles are necessarily either all symmetrical or all antisymmetrical with respect to permutation of *N* particles' (Messiah [1962], p. 595).[15] That is to say, if a collection of *N* identical particles is represented by the quantum state $\Psi_{(1,2,\ldots,N)}$ and the same collection with, say, particles 1 and 2 permuted is represented by $\Psi_{(1,2,\ldots,N)}$, then the symmetrization/anti-symmetrization postulate tells us that state must be related in the following manner:

$$\Psi_{(1,2,\ \ldots,N)} = e^{i\theta}\Psi_{(1,2,\ \ldots,N),}$$

[15] See Earman ([2010]) for a discussion.

where the exchange phase θ can take on a value of $\theta = 0$ for a system of bosons with a corresponding phase factor $e^{i\theta} = +1$ and a symmetric quantum state, or it can take a value $\theta = \pi$ for a system of fermions with a corresponding phase factor of $e^{i\theta} = -1$ and an antisymmetric quantum state.

There are two fundamental frameworks for understanding permutation invariance in quantum mechanics, which ground the symmetrization/anti-symmetrization postulate and its consequences, viz., that there are two basic types of particles and quantum statistics. Following Landsman ([2016]), we will call the first, due to Messiah and Greenberg ([1964]), the operator framework, and the second, due to, among others, Laidlaw and DeWitt ([1971]), Leinaas and Myrheim ([1977]), the configuration space framework. Landsman ([20136]) has argued that, in dimensions greater than two, both frameworks are equivalent and give equivalent verdicts regarding possible particle types and statistics. However, it turns out that in two dimensions, according to the configuration space framework, the exchange phase can take on any value. This allows the framework to represent bosons and fermions, as well as other particles known as 'anyons,' which are said to exhibit 'fractional quantum statistics.'[16]

Recall, the manner by which a collection of identical particles occupies energy states will depend on the kind of quantum statistics that such a collection manifests, which in turn depends on the type of particle considered. Particle type is decided by how such a collection behaves under permutation, and such behavior is captured by the value of the exchange phase θ and the corresponding phase factor $e^{i\theta}$. In short, on the configuration space framework, two central theorems (which may be found in Morandi [1992], pp. 119-120) dictate that the phase factor $e^{i\theta}$ is equivalent to the one-dimensional unitary representation γ of the fundamental group π_1 of the configuration space Q of the collection of identical particles, symbolized by $\gamma = e^{i\theta}$.[17] It has been

[16] The name is due to Nobel laureate Frank Wilczek ([1982]). Note that anyons and fractional statistics have nothing to do with so-called paraparticles and parastatistics (which arise from higher dimensional representations of the permutation group). For more on anyons see Wilczek ([1990]), Khare ([2005]), Shech ([2015a]) and references therein.

[17] See Hatcher (2002) for relevant background in algebraic topology. Roughly, the 'one-dimensional unitary representation' will allow us to represent groups with numbers. The 'fundamental group,' also known as the first homotopy group, is a topological invariant that allows

shown by Artin ([1947]), Fadell and Neuwirth ([1962]), and Fox and Neuwirth ([1962]) that the fundamental group for the two-dimensional ($d = 2$) and three-dimensional ($d = 3$) cases are given by:

$$\pi_1\ (Q) = B_N \text{ for } d = 2$$
$$\pi_1\ (Q) = S_N \text{ for } d = 3$$

where S_N is the permutation group and B_N is the Braid group. In other words, in three dimensions the fundamental group of the configuration space is the (finite and discrete) permutation group S_N which admits of the known one-dimensional unitary representation: $\gamma = \pm 1$ (+1 for bosons and −1 for fermions). In two-dimensions, on the other hand, the fundamental group is the (infinite and discrete) braid group B_N with one-dimensional unitary representations giving rise to phase factors of the form: $\gamma_{(\theta)} = e^{i\theta}$ where $0 \leq \theta \leq 2\pi$ so that the exchange phase can take on a continuous range of factors allowing for bosons, fermions, and anyons.

In the remaining part of this subsection, we will be working through the configuration space framework's treatment of the simple two-particle scenario in order to better convey how the appeal to the $d = 2$ idealization brings about the novel mathematical structure needed to represent anyons. Readers uninterested in such technicalities may skip to the next subsection without loss of philosophical continuity.

The fundamental group of the configuration space of the simplest scenario of two particles $N = 2$ in the $d = 2$ and $d = 3$ cases is as follows:

$$\pi_1\left(\frac{\mathbb{R}^2 \setminus \Delta}{S_2}\right) = \pi_1\left(RP_1\right) = Z \text{ for } d = 2$$
$$\text{for } d = 3,$$
$$\pi_1\left(\frac{\mathbb{R}^3 \setminus \Delta}{S_2}\right) = \pi_1\left(RP_2\right) = Z_2 \text{ for } d = 3,$$

one to classify topological spaces according to whether paths or loops in the space can be continuously deformed into each other. For instance, all paths can be continuously deformed into each other, and all loops can be shrunk to a point, in three-dimensional Euclidean space $\mathbb{R}^3$. Such a space is said to be 'simply connected.' The electron configuration space in the context of the abstract AB effect, $\mathbb{R}^3 \backslash S_{in}$, is not simply connected because loops encircling the region S_{in} cannot be shrunk to a point.

where S_2 is the permutation group for two particles, Δ an excised set of diagonal points that represent points where particles coincide, Z is the cyclic group of order one, i.e., the infinite group of integers under addition. Z_2 is the cyclic group of order two, i.e., it is the multiplicative group of, say, 1 and –1. RP_1 and RP_2 are the real projective one- and two-dimensional spaces, respectively.

Pictorially, for the $d = 3$ case the configuration space reduces to the real projective space in two dimensions RP_2. This can be visualized as the surface of a three-dimensional sphere with diametrically opposite points identified (see Figure 4) or a hemisphere with opposite points on the equator identified (see Figure 5). Consider three scenarios, corresponding to three paths *A*, *B*, and *C* in configuration space including no exchange (Figure 4a), exchange (Figure 4b), and a double exchange (Figure 4c), respectively.

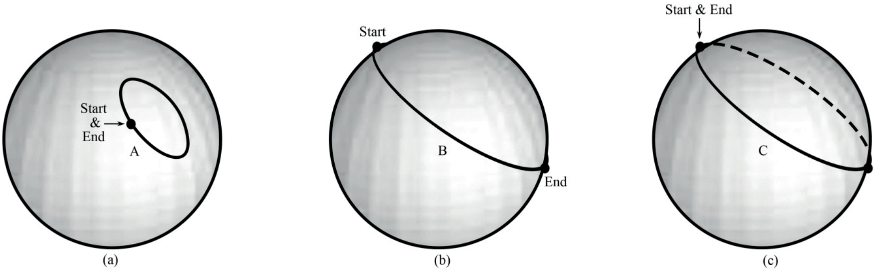


Figure 4. The real projective space in two dimensions RP_2, represented by a sphere with diametrically opposite points identified. Cases (a), (b), and (c), correspond to no exchange, exchange, and double exchange, respectively.

Concentrating on the no-exchange case (Figure 4a). We trace a path *A* in configuration space in which the two particles move and return to their original positions. Path *A* is a loop in configuration space, with the same fixed start and end points, which can be shrunk to a point. This corresponds to a trivial homotopy class in which the phase factor is trivial.

Moving onto the exchange case (Figure 4b), we start at one end of the configuration space and trace a path *B* to its diametrically opposite point. This represents an exchange or permutation between the two particles. Notice that since diametrically opposite points are identified (because the particles are identical), this path is actually a closed loop in configuration space. However, since the start and end points of Figure 4b are fixed, the loop cannot be shrunk to point. This corresponds to a non-trivial homotopy class with a non-trivial phase factor.

The double-exchange (Figure 4c) case includes tracing a path *C* in configuration space similar to that of *B*, but then tracing around the sphere back to the original starting point. Path *C* is a closed loop in configuration space that can be shrunk to a point, and so it is in the same homotopy class of path *A* with a corresponding trivial phase factor. Equivalently, we may visualize the paths *A*, *B*, *C* on a hemisphere with opposite points on the equator identified as in Figure 5, where paths *A* and *C* can be continuously deformed to a point but path B cannot because of the diametrically opposed fixed start and end point on the equator.

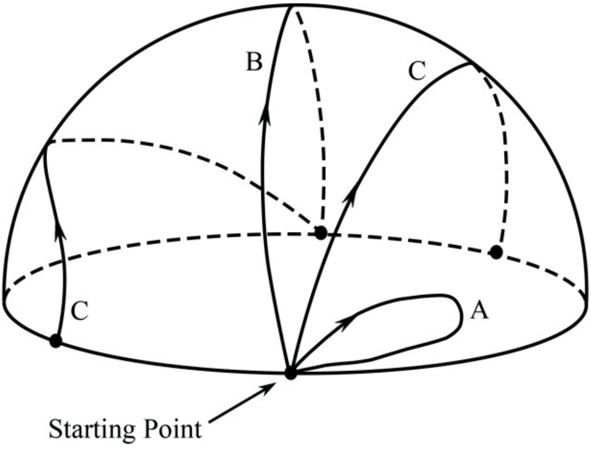


Figure 5. The real projective space in two dimensions RP_2, represented by the northern hemisphere with opposite points on the equator identified.

On the other hand, in the context of the $d = 2$ case, we are dealing with the real projective space in one dimension RP_1. We can visualize this configuration space as a circle with diametrically opposite points identified (see Figure 6).

Again, consider three paths *A*, *B*, and *C* in configuration space that correspond to no exchange (Figure 6a), exchange (Figure 6b), and a double exchange (Figure 6c), respectively. Path *A* traces a closed loop in configuration space (where the particles move but then return to their original positions with no exchange) which can be continuously shrunk to a point and has a corresponding trivial phase factor (as in the $d = 3$ case of figures 4a and 5a). Next, we trace a path *B* across half the circumference of the circle. Since diametrically opposed points are identified, this represents a particle exchange (Figure 6b). Path *B* traces a closed loop in configuration space that cannot be continuously shrunk to a point and has a corresponding non-trivial phase factor (as in the $d = 3$ case of figures 4b and 5b).

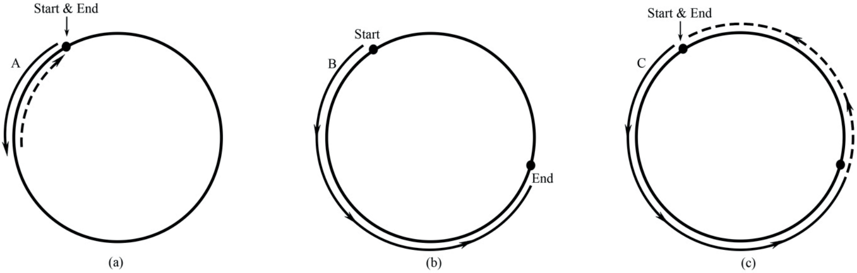


Figure 6. The real projective space in one dimension RP_1, represented by a circle with diametrically opposite points identified. Cases (a), (b), and (c), correspond to no exchange, exchange, and double exchange, respectively.

The main difference between the $d = 3$ and $d = 2$ cases arises when we consider path *C* (Figure 6c), in which the particles are permuted twice, represented by traversing the entire circular configuration space. Path *C* is a closed loop in configuration space but, unlike the $d = 3$ case, it cannot be shrunk to a point because the circle itself (so to speak) acts as an obstructive barrier. Moreover, path *C* cannot even be continuously deformed to overlap with path B. This means that, not only is the phase factor corresponding to the two paths non-trivial, but each path has a different phase factor for each path belongs to a different homotopy class. In fact, for every traversal (in configuration space) of half a circle, we get a closed loop that is in its own homotopy class.[18] In other words, by transitioning from three dimensions to two dimensions, we have transitioned from a doubly connected space to an infinitely connected space, and it is this change in topology that allows for the fractional statistics and the emergence of anyons.

3.2 Exploring Fractional Statistics

The configuration space framework for permutation invariance in quantum mechanics exemplifies, in a very direct way, the manner by which an idealization allows one to explore the modal structure of a theory and its representational capacities. The transition from $d = 3$ to $d = 2$ is an example of an idealization since two-dimensional systems, strictly speaking, do not exist. However, it is exactly by exploring two-dimensional systems that we discover the full spectrum of the exchange phase, as well as the fact that quantum mechanics has the representational capacities to represent more than just bosons and fermions. The novel mathematical structure that emerges in two-dimensions is the braid group B_N, with its corresponding one-dimensional unitary representation $\gamma_{(\theta)} = e^{i\theta}$ where $0 \leq \theta \leq 2\pi$. In three-dimensions, by contrast, the structure of the permutation group S_N and its one-dimensional unitary representation $\gamma = \pm 1$ is too sparse to represent anyons. Moreover, it is by appealing to the $d = 2$ idealization that we can clearly differentiate between the operator and configuration space frameworks, since it is only in $d = 2$ that the latter differs from the former in its verdict regarding possible particle type with corresponding statistics.

An additional sense of exploration, generating potential explanations, also arises in this context. Currently, the empirical evidence confirming the existence of anyons is inconclusive,[19] but physicists believe that anyons or ap-

[18] If we symbolize this by $\pi_1(Path)$ we get that $\pi_1(Path\ A) = 0$ for the trivial homotopy class, but the rest of the paths will be elements of non-trivial homotopy classes: $\pi_1(Path\ B) = 1$, $\pi_1(Path\ C) = 2$, … and so on, so that we generate all of the integers Z. Negative integers corresponding to traversal of the circular configuration space in the opposite direction.

[19] Recent supposed confirmations include Camino et al. ([2005]), but there is no consensus in the physics community regarding the reality of anyons and fractional statistics (instead of, say, composite fermions or composite bosons).

proximate anyons are likely to manifest in what is known as fractional quantum Hall effect (FQHE) systems.[20] Insofar as anyons are found in FQHE systems, which are built to constrain the dynamics of the system to approximately two-dimensions, the configuration space framework allows for one potential explanation of such particles. That is to say, if the idealized two-dimensional systems that allow for anyons and fractional statistics (via the configuration space framework) are good approximations of real FQHE systems, then it is because of the approximate two-dimensional nature of FQHE systems that anyons may emerge in the first place. Ultimately though, since the results derived from the configuration space framework hold for strictly two-dimensional systems and not approximate ones (such as thin layers embedded in three-dimensional space), such an explanation remains only a potential one at this time.

Another example of generating potential explanations can be found in the so-called flux tube model of the anyon introduced in Frank Wilczek's ([1982]) original paper on the subject. The goal is to explain how a composite particle that is neither a boson nor a fermion could come about. Here an anyon is described by spinless particle of charge e in the xy-plane orbiting around a very thin and long solenoid with magnetic flux Φ, set perpendicular to the plane, in the direction of the z-axis (see Figure 7). We are then asked to appeal to further idealizations:

> In the limit where the solenoid becomes extremely narrow and the distance between the solenoid and the charged particle is shrunk to zero, the system may be considered as a single composite object — a charged particle-flux tube composite. Furthermore, for a planar system, there can be no extension in the z-direction. Hence, imagine shrinking the solenoid along the z-directions also to a point. The composite object is now pointlike… (Rao [2001], p. 15)

[20] See Chakraborty et al. ([1995]), Ezawa ([2013]), and references therein for more on both the integer and fractional quantum Hall effects. For philosophical assesments see Bain ([2013], [2016]), Guay and Sartenaer ([2016a], [2016b]), Lancaster and Pexton ([2015]), Lederer ([2015]), and Shech ([2015], [2018b], [2018c]).

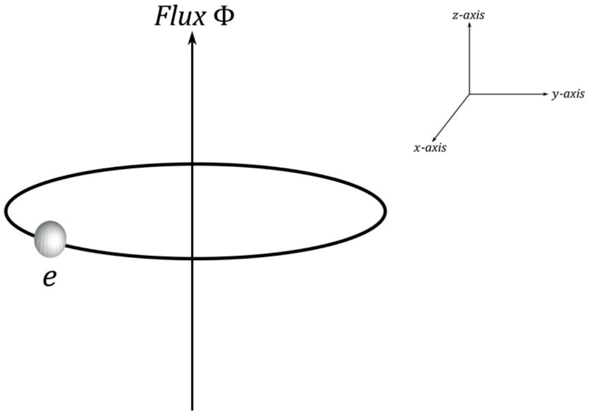


Figure 7. The Flux-tube model of the anyon. A spinless charge *e* particle orbiting around a thin and long solenoid with magnetic flux Φ.

It is difficult to think of this highly idealized model as the actual explanation of how an anyon, thought of as a composite particle, may come about. Nevertheless, the flux tube model of the anyon can be taken as generating a potential explanation – more specifically, a 'how-possibly' explanation, which demonstrates how a particular effect may be brought about in principle, without committing itself to the claim that this is how it was in fact brought about[21] – and, as such, it may point the way towards more realistic models to be developed in the future.

The third sense of exploration that we wish to discuss concerns proof-of-principle demonstrations, specifically of a kind that establishes the viability of a certain type of approach or methodology for the purpose of generating potential representations of target phenomena. In this sense, our case study shows that it is theoretically fruitful and perhaps empirically viable to consider the physics of systems that are not three-dimensional. This methodology is in part vindicated via the theoretical discovery of anyons and has led to the

[21] For some historical background on the notion of 'how-possibly' explanations, see Gelfert ([2016], p. 92); for a more substantive discussion, which links the distinction between 'how-possibly' and 'how-actually' explanations to different explanatory contexts arising from contrasting ways of framing a problem, see Bokulich ([2014]).

exploration of physics of various two- and one-dimensional systems that may perhaps be manifested approximately in the laboratory. In fact, (non-Abelian) anyons, such as bound states of the Mayorana Fermion, are the best candidates from which to build quantum computers (Nayak at al. [2008]), [22] and are regularly discussed in introductory textbooks to the subject (e.g., Pachos [2012]). Also, if the existence of anyons is empirically confirmed, we will know that the configuration space framework is the correct framework for understanding permutation invariance in quantum mechanics (and vice versa). This shows how an idealization (or idealized model) may establish that a certain type of methodology, namely, a framework for permutation invariance in quantum mechanics, is able to generate a potential representation of anyons (whereas the operator approach cannot).

The last sense of exploration that we wish to discover concerns using idealizations and models to assess the suitability of target systems. We will elaborate on this sense in the following section. Here we only wish to introduce the idea with some background history regarding anyons. In particular, anyons were first discussed in the literature solely as theoretical constructs. Before attempts were made to apply such constructs to FQHE systems, there were different target systems that physicists were hoping fractional statistics could shed light on:

> Surely the most dramatic result of the study of anyon statistics […] has been the demonstration [of] a new mechanism of superfluidity (and, for charged anyons, superconductivity). This superfluidity is quite a robust consequence of fractional quantum statistic at appropriate values of the fraction. […] [It] is tempting to speculate that the anyon mechanism of superconductivity will shed light on the copper oxide high temperature superconductors. Whether or not this speculation works out, the mechanism is of considerable theoretical interest and will undoubtedly play an important role in physics in the future. (Wilczek [1990], p. 325)

In other words, it was first speculated (and hoped) that anyons could shed light on high-temperature superconductivity. When later investigations did not bear fruit in this regard, physicists did not reject anyons altogether as useless idealizations. Instead, a new target system was sought: specifically,

[22] See Mourik et al. ([2012]) for recent experimental results.

FQHE systems. Hence, we see how the application of idealization and the investigation of an idealized model such as the flux tube model of the anyon may lead us to reassessments of the suitability of target systems.

4. The Hubbard Model of the Mott Phase Transition

4.1. Case Study: Hubbard Model

Consider a solid, such as a macroscopic piece of iron, made up of atoms at specific sites in a crystal lattice, and electrons which are either bound to specific atoms, shared between them, or moving about in the solid, as happens in electric conductors where an external electric field can easily bring about an electric current. Whereas in the previous two cases discussed in this paper, the quantum systems were constrained only by geometry, dimensionality, or external fields (as in the Aharonov-Bohm effect), in many-body systems such as crystal domains in a metal, a vast number of atoms and electrons interact with each other. Of special significance are substances such as iron, nickel, and cobalt, not only because they are ingredients of various technologically important alloys, but also because they exhibit strongly correlated behavior: the Coulomb interaction between the negatively charged electrons in such materials is so strong that the one-particle picture for calculating the electronic band structure, which governs many important physical characteristics, is no longer sufficient to describe them.[23] Such systems exhibit salient many-body effects, which manifest themselves macroscopically, e.g. in ferromagnetic behavior, when a solid exhibits a permanent magnetization even in the absence of an external field. Due to their large number ($\sim 10^{23}$) of interacting particles, strongly correlated electron systems cannot easily be studied on the basis of theoretical 'first principles', but require the use of many-body models.[24]

As a model of strongly correlated electrons in a crystal lattice, the Hubbard model has become one of the most extensively studied models in condensed

[23] For relevant background material in solid state/condensed matter physics see standard textbooks such as Ashcroft and Mermin ([1976]).

[24] For a review, see Gelfert ([2015]).

matter physics. Commonly, electronic states in a solid may be classified into those that are *localized*, i.e., centered around lattice sites and akin to atomic states, and *delocalized* (itinerant) states, i.e., Bloch states filled by electrons in accordance with Fermi statistics. Originally formulated for the study of collective magnetic order in solids, the Hubbard model is now widely employed for the study of various correlation effects in systems with such itinerant electrons. Its uses have proliferated beyond the question of the origins of spontaneous magnetism, and now include the study of metal-insulator transitions, high-temperature superconductivity, lattice gases, organic molecules and nanoparticles. While, at the descriptive level of scientific practice, this proliferation already points towards the exploratory utility of the Hubbard model, the specific character of exploration involved can be made more precise by looking at how features of the model have enabled a reassessment of its initial intended target. This will be done in the second subsection (4.2); first, let us motivate and summarize the Hubbard model.[25]

The Hubbard model was first developed for systems with narrow energy bands, in which the electrons, though delocalized and mobile, are still likely to be found near the lattice sites, i.e., the atoms (ions) that they are associated with. At a general level, the Hamiltonian of any system consisting of electrons and ions can be expressed as the sum of three components

$$H = H_{kin} + H_{ie} + H_{ee}$$

where the first term indicates the purely kinetic energy of the electrons, the second term the interaction between the electrons and the lattice potential, $V(r)$, due to the ions that make up the crystal lattice, and the third term comprises the electron-electron interaction. Assuming the lattice potential $V(r)$ to be strong and the mobility of the electrons to be small (though non-negligible), the sum of the *atomic* (single-particle) Hamiltonians at site i, $h_{at}^{(i)}$, can still be regarded as an acceptable representation of the system, at least near the lattice sites. (One way to think of this as a starting point for model-building is to regard the atoms as 'too far apart' to yet 'feel' any forces acting between them.)

[25] For the canonical derivation due to Hubbard, see his ([1963]).

The non-relativistic Schrödinger equation then is $h_{at}^{(i)}\varphi_n\left(\boldsymbol{r}-\boldsymbol{R}_i\right)=\varepsilon_n\varphi_n\left(\boldsymbol{r}-\boldsymbol{R}_i\right)$ where φ_n is an atomic wave function and n signifies the relevant set of quantum numbers. In the case at hand, the overlap between atomic wave functions associated with different lattice sites is assumed to be small, and the wave function will be centered strongly around the respective lattice sites.

The atoms, of course, form a crystal lattice, so electrons will not only feel the atomic Hamiltonian, but also the lattice potential. In the non-interacting parts of the Hamiltonian, $H_0=H_{kin}+H_{ie}=\sum_{i=1}^{N}h_0^{(i)}$, one therefore needs to include the lattice potential, $h_0^{(i)}=h_{at}^{(i)}+V\left(\boldsymbol{r}\right)$, which gives rise to the 'non-atomic' single-particle Schrödinger equation:

$$h_0\psi_{n\boldsymbol{k}}=\varepsilon_n\left(\boldsymbol{k}\right)\psi_{n\boldsymbol{k}}\left(\boldsymbol{r}\right)$$

In order to simplify the problem, the wave function $\psi_{n\boldsymbol{k}}$ can be approximated by the atomic wave functions:

$$\psi_{n\boldsymbol{k}}\left(\boldsymbol{r}\right)=\frac{1}{\sqrt{N_i}}\sum_{j=1}^{N}e^{i\boldsymbol{k}\boldsymbol{R}_j}\varphi_n\left(\boldsymbol{r}-\boldsymbol{R}_j\right).$$

While this approximation introduces some degree of error, the assumption is that the total error will be tolerable, since the approximation is nearly exact near the lattice sites (where the ions render the dynamics of the system almost 'atom-like') and becomes substantial only where the value of φ_n is already very small. It is also worth noting that the formula for $\psi_{n\boldsymbol{k}}$ satisfies Bloch's theorem, according to which, in an idealized system with periodic lattice potential, the wave function should be invariant with respect to translation, except for a phase factor: $\psi_{n\boldsymbol{k}}\left(\boldsymbol{r}+\boldsymbol{R}^m\right)=e^{i\boldsymbol{k}\boldsymbol{R}^m}\psi_{n\boldsymbol{k}}\left(\boldsymbol{r}\right)$.

The (general) Schrödinger equation for the Bloch functions $\psi_{n\boldsymbol{k}}$ can now be evaluated using the approximation, given in the previous formula, by the atomic wave functions. This results in a compact formula for the Bloch energies,

$$\varepsilon_n\left(\boldsymbol{k}\right)=\varepsilon_n+\frac{v_n+\frac{1}{\sqrt{N_i}}\sum_{j\neq 0}\gamma_n^{(j)}e^{i\boldsymbol{kR}_j}}{1+\frac{1}{\sqrt{N_i}}\sum_{j\neq 0}\alpha_n^{(j)}e^{i\boldsymbol{kR}_j}}$$

where $v_n=\int d^3rV\left(\boldsymbol{r}\right)\left|\varphi_n\left(\boldsymbol{r}\right)\right|^2$ reflects the influence of the lattice potential on a single electron, and the *overlap integrals*, $\alpha_n^{(j)}=\int d^3r\varphi_n^*\left(\boldsymbol{r}\right)\varphi_n\left(\boldsymbol{r}-\boldsymbol{R}_j\right)$ and $\gamma_n^{(j)}=\int d^3r\varphi_n^*\left(\boldsymbol{r}\right)V\left(\boldsymbol{r}\right)\varphi_n\left(\boldsymbol{r}-\boldsymbol{R}_j\right)$ which are a measure of the mutual influence between electrons at different lattice sites, can be assumed to be very small in value for any $\boldsymbol{R}_j\neq 0$, so that higher-order terms can be neglected. Restricting interactions to those between nearest neighbors (‘*n.n.*’) one thus arrives at

$$\varepsilon_n\left(\boldsymbol{k}\right)=T_0^{(n)}+\gamma_n^{(1)}\sum_{n.n.}e^{i\boldsymbol{kR}_{(n.n.)}}$$

which can be translated back into the formalism of creation and annihilation operators, with the non-interacting part of the Hamiltonian reducing to the simple formula $H_0=\sum_{ij\sigma}T_{ij}a_{i\sigma}^{\dagger}a_{j\sigma}$ where the T_{ij} are the *hopping integrals* associated with those contributions arising from the movement of a particle at site *j* to another site *i*.

The interacting part of the Hamiltonian can similarly be expressed in terms of creation and annihilation operators:

$$H_{ee}=\frac{1}{2}\sum_{ijkl}v\left(ij;kl\right)a_{i\sigma}^{\dagger}a_{j\sigma'}^{\dagger}a_{l\sigma'}a_{k\sigma}$$

where the matrix element *v*(*ij*; *kl*) is constructed from atomic wave functions:

$$v\left(ij;kl\right)=\frac{e^2}{4\pi\varepsilon_0}\ d^3r_1d^3r_2\frac{\varphi^*\left(\boldsymbol{r}_1-\boldsymbol{R}_i\right)\varphi^*\left(\boldsymbol{r}_2-\boldsymbol{R}_j\right)\varphi\left(\boldsymbol{r}_2-\boldsymbol{R}_l\right)\varphi\left(\boldsymbol{r}_1-\boldsymbol{R}_k\right)}{\left|\boldsymbol{r}_1-\boldsymbol{r}_2\right|}.$$

The matrix element bears a close resemblance to the classical Coulomb potential, but it also takes into account the quantum effects between different particles, as indicated by the 'mixed' integral. Because of the small overlap between atomic wave functions centered on different lattice sites, the intra-atomic matrix element $U = v(ii; ii)$ can be expected to strongly dominate the dynamics of interaction; neglecting, as a final approximation, all other matrix elements, simplifying the operator combination using the number operator $n_{i\sigma} = a_{i\sigma}^{\dagger} a_{i\sigma}$, and combining the non-interacting and interacting parts of the Hamiltonian, gives the standard *Hubbard Hamiltonian*:

$$H = \sum_{ij\sigma} T_{ij} \hat{a}_{i\sigma}^{\dagger} \hat{a}_{j\sigma} + \frac{1}{2} U \sum_{i\sigma} \hat{n}_{i\sigma} \hat{n}_{i,-\sigma}.$$

The Hubbard model contains only a small number of parameters, most explicitly, the ratio between the Coulomb repulsion and the kinetic energy of the electrons and, less overtly, the filling of the energy band and the geometry of the crystal lattice (which is implicit in the summation range, e.g. by performing the sum over nearest neighbors in a unit cell).

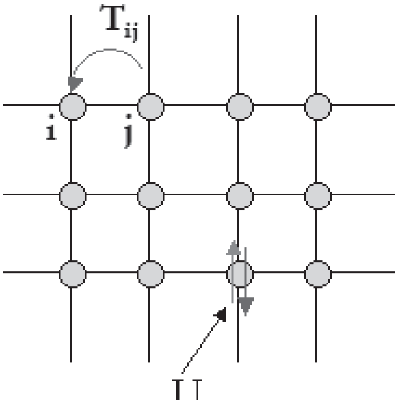


Figure 8. Schematic representation of the processes modeled by the Hubbard Hamiltonian: *U* is the Coulomb repulsion experienced by two electrons (of opposite spin) occupying the same site in a crystal lattice; T_{ij} are the hopping integrals that apply to the movement of an electron from one lattice site, *j*, to another, *i*.

Yet this has proven to be sufficient to render the Hubbard model one of the simplest and most fruitful frameworks for modeling itinerant, strongly correlated electrons in crystal lattices. As a recent *Nature Physics* editorial on 'The Hubbard Model at Half a Century' puts it, while the Hubbard model was initially 'introduced to provide an explanation for the itinerant ferromagnetism of transition metals, such as iron and nickel, […] the past 50 years have seen its relevance go far beyond that original context', and the (anonymous) authors express their confidence that, even after half a century, the model 'should be a stimulus *for further explorations*' ('The Hubbard Model', [2013], p. 523; italics added).

4.2 Exploration in the Hubbard Model

The exploratory utility of the Hubbard model is partly due to its relative simplicity and the ease with which it can be 'customized' to fit single-band and multiband scenarios, and phenomena beyond spontaneous magnetism, such as high-temperature superconductivity and artificial lattices of cold atoms. It has also given rise to other models, such as the *t-J model*, which can be derived from the Hubbard model in the limit of large U via an (operator-level) transformation, in which doubly-occupied electron states are neglected: that is, an electron 'hops' from one lattice site to another (transferring energy t, similar to what happens in the Hubbard model), but only when that destination site is empty. Higher-order processes, in particular, such as two electrons hopping onto the same (previously unoccupied) lattice site, are excluded on this picture.[26] Yet the Hubbard model's simplicity and ease of adaptability alone do not exhaust its exploratory features. In this subsection, we shall discuss a salient example of the fourth type of exploration distinguished earlier, *viz.* exploration via assessing the suitability *of the target system*.

The suggestion that exploration may consist in holding the model and its representational means fixed, while searching for target systems which the model may be able to adequately describe, might at first seem to put the cart before the horse: should we not try to find models that fit our target systems,

[26] See Spałek ([2007]) for more on the t-J model.

rather than 'shopping around' for target systems that fit our preferred models? (On this point, see Gelfert 2006, Section 4.5.4.) In response, two points come to mind. First, in the early days of research, before a good theoretical grasp of a phenomenon has been gained, it may be unclear whether the putative phenomenon is indeed the result of a more stable, reproducible set of circumstances. As noted in our earlier discussion of exploration in the Aharonov-Bohm effect (Section 2.2) and of fractional quantum statistics (Section 3.2), it often takes time for theoretical descriptions and experiments to become sufficiently refined to establish that a phenomenon is indeed what physicists call a recurrent 'effect'. In this early stage of research, commitment to putative target systems and phenomena is tentative, and one's choice of target system or phenomenon will be subject to revision. Second, choices of target systems for a given model are not made arbitrarily, but often are the results of *re*assessing a model's target system in the light of new findings. That is, model equations are not deployed arbitrarily to represent a new target; rather, the new target suggests itself on the basis of findings generated in the course of a sustained exploration of the model's utility for describing the initial target. This (fourth) sense of exploration, we submit, can be illustrated in the Hubbard model's evolution from a model for describing itinerant electrons in transition metals (and other conductors) to a potential model for insulators of a particular kind, the so-called *Mott insulators*.

According to A.H. Wilson's metal/non-metal criterion, an insulator (or semiconductor) is characterized by having either completely filled or completely empty energy bands, separated by an energy gap that prevents electrons from becoming mobile (unless a significant external field is applied).[27] A metal, by contrast, has partially filled bands, such that even the slightest external perturbation is enough to excite an electron into an (infinitesimally higher) empty state in the same energy band. Due to such 'conduction bands', metals are able to conduct electricity, whereas insulators are not. As discussed in the previous section, the Hubbard model, with its emphasis on the ease with which itinerant electrons can 'hop' from one lattice site to the other, was

[27] See Wilson ([1936]) for the original discussion, and Martin ([2004], pp. 40-44) for a more recent textbook presentation.

initially intended to model ferromagnetic conductors such as nickel and iron (rather than insulators). Historically, however, it quickly turned out that the classification into metals and non-metals was not as clear-cut as Wilson's criterion had suggested. Thus, as early as the 1930s, it was shown that certain transition-metal oxides were insulators, even though their crystalline structure suggested that they had partially filled bands and should be conductors.

Theoretical physicists, including Rudolf Peierls, Lev Landau, and Nevill Mott, explored the possibility of the breakdown of Wilson's criterion being the result of correlations associated with the repulsive Coulomb interaction between the electrons. Mott, in particular, clarified the metal-insulator criterion in relation to transition-metal monoxides, and the problem became known as the puzzle of 'Mott insulators' (Mott [1949]). Subsequently, in the 1950s, it was realized that all Mott insulators are antiferromagnets and remain insulators even above the Néel temperature, the critical temperature above which antiferromagnetic order is destroyed and a solid turns into a paramagnet. While it was realized, notably by John Slater, that antiferromagnetism led to a 'splitting up' of electron bands, thereby increasing the chances for the emergence of energy gaps characteristic of insulators, numerical simulations indicated that this alone could not explain the behavior of Mott insulators. Exactly why substances that, on the basis of their crystalline structure and electronic characteristics, should be expected to conduct electricity are nonetheless insulators – and why these Mott insulators are antiferromagnets – remained an open question, and a lively debate ensued within theoretical condensed matter physics in the decades to come. This was compounded by the fact that certain types of substances, such as transition metal oxides, displayed a transition from insulating to metallic behavior – a 'Mott transition', as it became known – as the result of certain factors (e.g. slight modifications of its composition known as 'doping').

With the Hubbard model having been developed as a model of ferromagnetic metals, one might wonder how it could possibly shed light on the phenomenon of Mott insulators – which, after all, are *anti*ferromagnetic *insulators* – and their associated phenomena. Yet, in recent years, it has become one of the most extensively studied models for exploring the Mott metal-insulator transition and related phenomena, attesting to the exploratory power of the

Hubbard model, which in this case manifests itself in a thoroughgoing reassessment of the range of its intended targets. We are, of course, not suggesting that the Hubbard model is *no longer* used for its original purposes – it remains the model of choice for many researchers interested in the behavior of ferromagnetic substances and strongly correlated electron systems in model – but, rather, that the range of its target systems has been expanded to include types of targets that could not have been foreseen by, and indeed would likely have seemed outlandish to, its initial proponents.

Why consider the Hubbard model a potential model of Mott insulators in the first place? This latest turn in the Hubbard model's varied career came as something of a surprise to many condensed matter physicists and is due to an exact mapping that obtains in a limiting case of the standard Hubbard model. In particular, it results from the limit of strong interactions, when *U* is much larger than the other matrix elements (*W*). Evaluating the model using second-order perturbation theory, the multiple degeneracy of the ground state of the model is lifted and, at half-filling ($n = 1$), only the following effective Hamiltonian remains *as the limiting case of the Hubbard model*:

$$H_{eff} = \sum_{\boldsymbol{R}_1 \neq \boldsymbol{R}_2} \frac{2\left|t\left(\boldsymbol{R}_1 - \boldsymbol{R}_2\right)\right|^2}{U} \left(\hat{\mathbf{S}}_{\boldsymbol{R}_1} \cdot \hat{\mathbf{S}}_{\boldsymbol{R}_2} - \frac{1}{4} \right)$$

which is rendered a compact formula thanks to the use of spin operators $\hat{\mathbf{S}}_{\boldsymbol{R}_i} = \left(\sum_{\sigma,\sigma'} \hat{a}^{\dagger}_{i\sigma} \boldsymbol{\sigma}_{\sigma,\sigma'} \hat{a}_{i\sigma'} \right) / 2$ (with $\boldsymbol{\sigma}$ the vector of Pauli matrices). As it turns out, this limiting case is identical to another much-studied quantum model, the antiferromagnetic *Heisenberg model*

$$H_{Heis} = \sum_{\boldsymbol{R}_1 \neq \boldsymbol{R}_2} J\left(\boldsymbol{R}_1 - \boldsymbol{R}_2\right) \left(\hat{\mathbf{S}}_{\boldsymbol{R}_1} \cdot \hat{\mathbf{S}}_{\boldsymbol{R}_2} - \frac{1}{4} \right)$$

with $J\left(\boldsymbol{R}_1-\boldsymbol{R}_2\right)=2\left|t\left(\boldsymbol{R}_1-\boldsymbol{R}_2\right)\right|^2/U$.[28] The Heisenberg model, in turn, has long been known as 'the "standard model" for the description of magnetic insulators' (Gebhard [2000], p. 75).

The situation, then, is this: whereas for $U = 0$, the Hubbard model at half-filling describes a metal, for $U>>W$ it maps onto the standard model of antiferromagnetic insulators. This behavior, which could not easily be 'read off' from the model's original formulation, and certainly played no role in the type of derivation given in the preceding subsection, provides a stunning example – or so we believe – of the exploratory potential that is afforded by some models in virtue of their structure and the relations they stand in with other models.[29] For, having convinced themselves that the Hubbard model at half-filling, and for low interaction strength U, describes a metal, and for very large values of U behaves like an antiferromagnetic insulator, researchers 'are confident that the model is indeed capable of describing a Mott transition at a critical interaction strength U_c, somewhere in between.

Once again, exploration is borne out to be a fruitful strategy in model-based research – this time, this time by being open towards the reconsideration of one's target system. This, together with the construction of limiting cases of highly idealized models, led to a widening of the potential domain of applicability of the Hubbard model to include not only ferromagnetic metals, but also Mott insulators such as the transition-metal oxides, in which the phenomenon of Mott insulation was first observed.

5. Conclusion

We have presented three case studies, which illustrate our suggestion that idealizations, in the construction of models or by considering limiting cases, play exploratory roles in science. To end, we first want to consider a general

[28] For details of the derivation, see various textbooks on quantum magnetism, e.g. Gebhard ([2000]).

[29] Gelfert ([2009]) discusses this case, but relates it primarily to the role of rigorous results in establishing 'cross-model' support, rather than to its exploratory uses.

objection to our account, and then wish to reflect on how Michael Weisberg's ([2007], [2013]) recent taxonomy of idealizations and models compares with our case studies.[30]

One of the main worries associated with taking idealizations and models to play a substantive exploratory role in science, is that it may seem that the concept of exploration is so generic as to border on the trivial. For instance, concerning the exploration of theoretical structure, one may raise the following objection: Is it the case that whenever one considers what quantum mechanics says about *any* idealized system, no matter how unrelated to actual scientific investigations, one is 'exploring' the structure of quantum mechanics? If so, haven't we trivialized the notion of exploration? Should there not be any constraints on what counts as a bona fide exploration? Yet, in reply, it is worth emphasizing that the notion of 'exploration' is not a free-for-all: Whether in relation to experimentation, scientific modeling, or, as in our case, idealizations and limiting cases, exploration is marked by recurring strategies, not by haphazard arbitrariness. Indeed, this is why the case studies we have presented are important: They demonstrate not random moves in the investigation of salient scientific questions, but structured approaches that exhibit continuity and stability.

Perhaps, however, the objection is meant to be more specific, targeting not the various forms that exploration in science may take, but the lack of specific criteria of success. Yet, much the same could be said about other basic concepts such as scientific explanation. For one, we readily accept that there are various notions of scientific explanation, e.g., causal-mechanical, unificationist, deductive-nomological, so that what counts as a successful explanation will greatly depend on which notion is involved and the type of why-question one is looking to answer. Similarly, it is to be expected that there will be various notions of exploration. Second, whether or not an argument proffered merely as a *potential* explanation counts as a good one will not only depend

[30] We wish to give John Earman credit for first bringing to our attention (in conversation and correspondence) the idea that Weisberg's ([2007], [2013]) account may well be wanting. Earman is the first to argue that the three-fold scheme offers a distorted view of the story of the AB effect. Here we extend and elaborate on his insights to cover the anyon and Hubbard model case studies.

on whether it turns out to be an *actual* explanation, but also on its merits in terms of explanatory power, novelty, simplicity, etc., and historical context (e.g., on whether the explanation does justice to the well accepted science of the time). And just as there are better and worse potential explanations, some exploratory moves in science will be more or less fruitful. As we have seen throughout the brief historical sketches included in our case studies, scientists are often keenly aware of which exploratory moves are further removed from the phenomena, and which stand a realistic chance of leading to the genuine discovery of new phenomena, effects, and explanations.

Finally, we wish to reconsider some of the philosophy of science literature on idealizations in light of the case studies presented here. Michael Weisberg ([2007], [2013]) has recently presented a well-received taxonomy of idealizations. His three-fold classification scheme includes *Galilean* idealizations, *minimalist model* idealizations, and *multiple-model* idealization (see Figure 9). We will consider each notion in turn in light of the case studies that we have discussed. A Galilean idealization is a distortion used to simplify, and render computationally tractable, the treatment of a target system. Can the two-dimensional idealization in the context of anyons and fractional statistics be understood along these lines? We submit that it cannot. By moving from three to two dimensions we saw how a novel type of mathematical structure emerged (namely, the fundamental group for the configuration space became that of the braid group instead of the permutation group), allowing for the representation of anyons and fractional statistics. The goal of such an exercise was not specifically to simplify a target system (although we are happy to grant that simplification may be a partial goal or by-product of such an idealization). In contrast with, for instance, the Ising model, the two-dimensional scenario in the anyon case was *not* implemented because it made the model computationally more tractable.

Next, minimalist model idealizations are distortions used to expose key causal or explanatory factors in the behavior of the target system. In the case of anyons, we grant that the two-dimensional idealization may be used to generate a potential explanation of the phenomenon of fractional statistics. Nevertheless, it is difficult to motivate the idea that the two-dimensional idealization generates the actual explanation of fractional statistics since the results obtained from the configuration space framework hold for systems that are, strictly speaking, exactly two-dimensional. Systems that are approximately two-dimensional cannot manifest anyons and fractional statistics according to the configuration-space framework. Moreover, at no point was there a discussion of any causal factors involved; in this sense, the exploration afforded by the two-dimensional idealization runs deeper than the search for causes and effects. So, again, a minimalist model idealization does not seem to be the kind of idealization involved in the case of anyons.

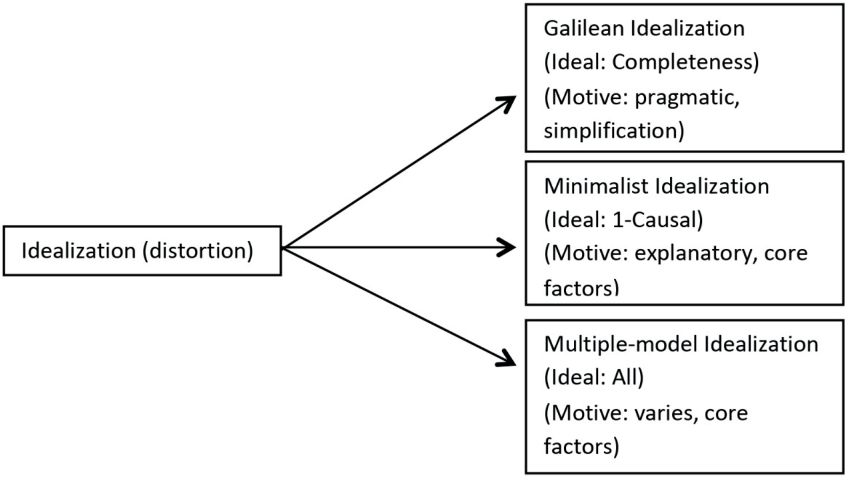

Idealization (distortion)
Galilean Idealization
(Ideal: Completeness)
(Motive: pragmatic, simplification)
Minimalist Idealization
(Ideal: 1-Causal)
(Motive: explanatory, core factors)
Multiple-model Idealization
(Ideal: All)
(Motive: varies, core factors)

Figure 9. Weisberg's taxonomy of idealizations.

Last, multiple-model idealizations are multiple incomplete models, designed to serve different epistemic/pragmatic goals, which typically trade off against each other. While they 'may retain a single, complex target, [they] construct multiple models for the target' (Weisberg [2013], p. 113). Yet, as we saw in both the anyon case study and the connection with the Hubbard model, exploration may arise from using the *same* model and varying – not arbitrarily, but in a way that is clearly motivated by the course of inquiry – its prospective target system or phenomenon. To be sure, there is a sense in which performing a limiting procedure, for example varying the model's parameters so as to realize the scenario $U>>W$, gives rise to a sequence of models, one for each set of parameter values. On this understanding, *any* limiting case in scientific modeling would be associated with 'multiple models'. But it seems to us that this is not the sense intended by Weisberg who, rightly, links multiple-model idealization to competing, qualitatively distinct models, whether due to unavoidable tradeoffs or due to the complexity of the target system. In neither sense, however, is the case of the Mott-Hubbard transition adequately characterized by treating it as a multiple-model idealization (which is not to say that, in other circumstances, it cannot be used in this way; the derivation of the t-J model, hinted at in Section 4.2 may be a case in point). Similarly, in the context of anyons and fractional statistics we did not discuss multiple models but only one idealized two-dimensional model or scenario, and there was one only one goal: to show how anyons and fractional statistics may emerge. Indeed, this holds more generally for models aiming at 'how-possibly' explanations of unexpected phenomena (see footnote 19), where the goal is to demonstrate how a surprising (or perhaps merely speculative) 'effect' *may* possibly arise, while neither requiring that this is how it, in fact, emerges, nor demanding that the model be useful for other purposes.

Thus, we think that Weisberg's ([2007], [2013]) scheme is at the very least incomplete, in that it does not make room for the exploratory role of idealizations and models, thereby offering a distorted view of the types of situations illustrated by our case studies. Moreover, we submit that similar claims can be made about other taxonomies of idealizations and models in the literature, e.g., McMullin ([1985]), Nowak ([1980]), Shaffer ([2012]), but due to space constraints we will leave the details to be worked out in future work. In addition, it seems to us that a deep *understanding* of scientific theory can only be gained by making room for the exploratory role of models and idealizations (cf. Shech ([Manuscript])). If this is the case, we will have found a substantive sense for which idealizations are essential to science, namely, for scientific understanding. Whether or not these further claims can be substantiated given our best accounts of scientific understanding is another issue that we leave for further study. Last, although we have concentrated here on case studies in physics, it would be interesting to see whether the notions of explorations discussed also arise in other sciences as well.

References

Aharonov, Y., D. Bohm. [1959]: 'Significance of electromagnetic potentials in the quantum theory', *Physical Review*, 115, pp. 485-91.

Artin, E. [1947]: 'Theory of Braids', Annals of Mathematics, 48(1), pp. 101-126.

Ashcroft, N.W. and Mermin, D.N. [1976]: S*olid State Physics*, New York: Holt, Rinehart and Winston.

Bailer-Jones, D., [2002]: 'Models, metaphors and analogies', in Machamer, P. and Silberstein, M. (*eds.*), *The Blackwell Guide to the Philosophy of Science*, Oxford: Blackwell, pp. 108–127.

Bain, J. [2013]: 'Emergence in effective field theories', *European Journal for Philosophy of Science*, 3, 257–273.

Bain, J. [2016]: 'Emergence and mechanism in the fractional quantum Hall effect.' *Studies in History and Philosophy of Modern Physics*, 56, 27–38.

Ballentine, L.E. [1998]: *Quantum Mechanics: A Modern Development*, Singapore: World Scientific.

Ballesteros, M. and Weder. R. [2009]: 'The Aharonov–Bohm effect and Tonomura et al. experiments: Rigorous results', *Journal of Mathematical Physics*, 50, 122108.

Ballesteros, M. and Weder. R. [2011]: 'Aharonov–Bohm effect and high-velocity estimates of solutions to the Schrödinger equation', *Communications in Mathematical Physics*, 303(1), pp. 175-211.

Batterman, R. [2002]: *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, London: Oxford University Press.

Batterman, R. [2003]: 'Falling cats, parallel parking, and polarized light', *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 34, pp. 527–557.

Batterman, R. and Rice. C. [2014]: 'Minimal Model Explanations', *Philosophy of Science*, 81(3), pp.349-376.

Bocchieri, P. and Loinger, A. [1978]: 'Nonexistence of the Aharonov-Bohm Effect', *Nuovo Cimento,* 47A(4), pp. 475-482.

Bokulich, A. [2008]: *Re-examining the Quantum-Classical Relation: Beyond Reductionism and Pluralism*, Cambridge: Cambridge University Press.

Bokulich, A. [2014]: 'How the Tiger Bush Got Its Stripes: "How Possibly" vs. "How Actually" Model Explanations', *The Monist*, 97(3), pp. 321-338.

Camino, F.E., Zhou, W., Goldman, V.J. [2005]: 'Realization of a Laughlin Quasiparticle Interferometer: Observation of fractional statistics', *Physical Review B*, 72, pp. 075342.

Chakraborty, T. and Pietilinen, P. [1995]: *The Quanutm Hall Effects*, Berlin: Springer.

Chambers, R.G. [1960]: 'Shift of an Electron Interference Pattern by Enclosed Magnetic Flux', *Physical Review Letter*, 5(1), pp.3-5.

Elliott-Graves, A. and Weisberg, M. [2014]: 'Idealization', *Philosophy Compass,* 9(3), pp. 176–185.

Ezawa Z.F. [2013]: *Quantum Hall Effects: Recent Theoretical and Experimental Developments*, Singapore: World Scientific.

Earman, J. [2010]: 'Understanding permutation invariance in quantum mechanics.', Unpublished manuscript.

Earman, J. [2017]: 'The Role of Idealization in the Aharonov-Bohm Effect', *Synthese* (Online First) 1-29. https://doi.org/10.1007/s11229-017-1522-9

Fadell, E. and Neuwirth, L. [1962]: 'Configuration Spaces', *Mathematica Scandinavica*, 10, pp. 111-118.

Fletcher, S., Palacios, P., Ruetsche, L., and E. Shech. [Forthcoming]: *Special Issue: Infinite Idealizations in Science* in *Synthese*.

Fox, R. and Neuwirth, L. [1962]: 'The Braid Groups', *Mathematica Scandinavica*, 10, pp. 119-126.

Gebhard, F. [2000]: *The Mott Metal-Insulator Transition: Models and Methods*, Berlin: Springer.

Gelfert, A. [2009]: 'Rigorous Results, Cross-Model Justification, and the Transfer of Empirical Warrant', *Synthese*, 169 (3), pp. 497–519.

Gelfert, A. [2015]: 'Between Rigor and Reality: Many-Body Models in Condensed Matter Physics', in Falkenburg, B. and Morrison, M. (eds.), *Why More Is Different: Philosophical Issues in Condensed Matter Physics and Complex Systems*, Heidelberg: Springer, pp. 201–226.

Gelfert, A. [2016]: *How to Do Science with Models: A Philosophical Primer,* Cham: Springer.

Gelfert, A. [2018]: 'Models in Search of Targets: Exploratory Modelling and the Case of Turing Patterns', in Christian, A., Hommen, D., Retzlaff, N., and Schurz, G. (eds.), *Philosophy of Science: Between Natural Sciences, Social Sciences, and Humanities*, Dordrecht: Springer 2018, pp. 245–271.

Guay, A., & Sartenaer, O. [2016a]: 'A new look at emergence. Or when after is different', *European Journal for Philosophy of Science*, 6, 297–322.

Guay, A., & Sartenaer, O. [2016b)]: 'Emergent quasiparticles. Or how to get a rich physics from a sober metaphysics', In: O. Bueno, R. Chen, & M.B. Fagan (Eds.), *Individuation across experimental and theoretical sciences*. Oxford: Oxford University Press. http://hdl.handle.net/2078.1/179059.

Hatcher, A. [2002]: *Algebraic Topology*, Cambridge: Cambridge University Press.

Healey, R. [1997]: 'Nonlocality and the Aharonov–Bohm Effect', *Philosophy of Science*, 64, pp. 18–41.

Healey, R. [1999]: 'Quantum Analogies: a Reply to Maudlin', *Philosophy of Science*, 66, pp. 440-447.

Healey, R.A. [2007]: *Gauging What's Real: The Conceptual Foundations of Contemporary Gauge Theories*, New York: Oxford University Press.

Hubbard, J. [1963]. 'Electron Correlations in Narrow Energy Bands', *Proceedings of the Royal Society of London*, 276 (1365), pp. 238–257.

Jones, M. [2005]: 'Idealization and abstraction: A framework', in Jones, M. and Cartwright, N. (eds.), *Correcting the model: Idealization and abstraction in the sciences*, Amsterdam: Rodopi, pp. 173-218.

Kadanoff, L.P. [2000]: *Statistical Physics: Statics, Dynamics and Renormalization*, Singapore: World Scientific.

Khare, A. [2005]: *Fractional Statistics and Quantum Theory*, New Jersey: World Scientific.

Ladyman, J. [2008]: 'Idealization', in Psillos, S. and Curd, M. (eds.), *The Routledge Companion to Philosophy of Science*, London and New York: Routledge, pp. 358-366.

Laidlaw, M.G. and DeWitt, C.M. [1971]: 'Feyman Functional Integrals for System of Indistinguishable Particles', *Physical Review D*, 3, pp. 1375–1378.

Landsman, N.P. [2016]: 'Quantization and superselection sectors III: Multiply connected spaces and indistinguishable particles', *Reviews in Mathematical Physics* Vol. 28, No. 09 1650019. Available via https://doi.org/10.1142/S0129055X16500197

Lederer, P. [2015]: 'The quantum Hall effects: Philosophical approach', *Studies in History and Philosophy of Modern Physics*, 50, 25–42.

Leinaas, J.M. and Myrheim, J. [1977]: 'On the Theory of Identical Particles', *Nuovo Cimento*, 37B, pp. 1-23.

Martin, R. [2004]: *Electronic Structure. Basic Theory and Practical Methods*, Cambridge: Cambridge University Press.

Massimi, M. [2018]: 'Perspectival Modeling', *Philosophy of Science*, 85(3):335–359.

Maudlin, T. [1998]: 'Discussion: Healey on the Aharonov–Bohm Effect', *Philosophy of Science*, 65, pp. 361–368.

McMullin, E. [1985]: 'Galilean Idealization', *Studies in the History and Philosophy of Science* 16, pp. 247–273.

Messiah, A. M. [1962]: *Quantum Mechanics*, New York: John Wiley & Sons.

Messiah, A.M., Greenberg, O.W. [1964]: 'Symmetrization Postulate and its Experimental Foundation', *Physical Review B*, 136, pp. 248-267.

Möllenstedt, G., W. Bayh. [1962]: 'Kontinuierliche Phasenschiebung von Elektronenwellen im kraftfeldfreien Raum durch das magnetische Vektorpotential eines Solenoids', *Zeitschrift für Physik* 169:263.

Morandi, G. [1992]: *The Role of Topology in Classical and Quantum Mechanics*, Berlin: Springer-Verlag.

Morgan, M. and Morrison, M. (*eds*) [1999]: *Models as Mediators. Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press.

Mott, N. [1949]: 'The Basis of the Electron Theory of Metals, with Special Reference to the Transition Metals', *Proceedings of the Physical Society. Section A*, 62, pp. 416-422.

Mourik, V., Zuo, K., Frolov, S.M., Plissard, S.R., Bakkers, E.P. A.M., Kouwenhoven, L.P. [2012]: Signatures of Majorana Fermions in Hybrid Superconductor-Semiconductor Nanowire Devices', *Science*, 336, pp. 1003-1007.

Nayak, C., Simon, S., Stern, A., Freedman, M., Das Sarma, S. [2008]: 'Non-Abelian anyons and topological quantum computation', *Reviews of Modern Physics*, 80, pp. 1083.

Norton, J.D. [2012]: 'Approximations and Idealizations: Why the Difference Matters', *Philosophy of Science*, 79, pp. 207–32.

Nowak, L. [1980]: *The Structure of Idealization*. Dordrecht: Reidel.

de Oliveira, C.R. and Pereira. M. [2008]: 'Mathematical Justification of the Aharonov-Bohm Hamiltonian', *Journal of Statistical Physics*, 133, pp. 1175-1184.

de Oliveira, C.R. and Pereira. M. [2010]: 'Scattering and Self-adjoint extensions of the Aharonov-Bohm Hamiltonian', *Journal of Physics A: Mathematical and Theoretical* 43, pp. 1-29.

de Oliveira, C.R. and Pereira. M. [2011]: 'Impenetrability of Aharonov-Bohm Solenoids: Proof of Norm Resolvent Convergence', *Letters in Mathematical Physics*, 95, pp. 41-51.

Pachos, J.K. [2012]: *Introduction to Topological Quantum Computation*, New York: Cambridge University Press.

Peshkin, M. and Tonomura. A. [1989]: *The Aharonov-Bohm Effect*. Berlin: Springer Verlag.

Psillos, S. [2011]: 'Living with the Abstract: Realism and Models', *Synthese*, 180, pp. 3-17.

Rao, S. [2001]: 'An Anyon Primer', Available via arXiv:hep-th/9209066

Redhead, M. [1980]: 'Models in Physics', *British Journal of Philosophy of Science*, 31(2), pp. 145-163.

Ruetsche, L. [2011]: In*terpreting Quantum Theories*. Oxford: Oxford University Press.

Shaffer, M.J. [2012]: *Counterfactuals and Scientific Realism*. London: Palgrave Macmillan.

Shech, E. [2013]: 'What is the "Paradox of Phase Transitions?"', *Philosophy of Science*, 80, pp. 1170–1181.

Shech, E. [2015a]: 'Two Approaches to Fractional Statistics in the Quantum Hall Effect: Idealizations and the Curious Case of the Anyon', *Foundations of Physics*, 45(9): 1063-110.

Shech, E. [2015b]: 'Scientific Misrepresentation and Guides to Ontology: The Need for Representational Code and Contents', *Synthese* 192(11): 3463-3485.

Shech, E. [2016]: 'Fiction, Depiction, and the Complementarity Thesis in Art and Science', *The Monist*, 99(3): 311-332.

Shech, Elay. [2017]: 'Idealizations, Essential Self-Adjointness, and Minimal Model Explanation in the Aharonov-Bohm Effect', *Synthese* (Online First,), 1-25. https://doi.org/10.1007/s11229-017-1428-6

Shech, E. [2018a]: 'Infinite idealizations in physics', *Philosophy Compass*. e12514. https://doi.org/10.1111/phc3.12514

Shech, E. [2018b]: 'Philosophical Issues Concerning Phase Transitions and Anyons: Emergence, Reduction, and Explanatory Fictions', *Erkenntnis* (Online First), 1-31. doi: 10.1007/s10670-018-9973-z

Shech, E. [2018c]: 'Infinitesimal Idealization, Easy Road Nominalism, and Fractional Quantum Statistics', *Synthese* (Online First), 1-25. doi: 10.1007/s11229-018-1680-4

Shech, E. [Manuscript]: 'Do Idealizations Afford Understanding? The Case of the AB Effect.' Manuscript.

Spałek, J. [2007]: 't-J model then and now: A personal perspective from the pioneering times', Available via arXiv:0706.4236 [cond-mat.str-el]

Stanley, H.E. [1971]: *Introduction to Phase Transitions and Critical Phenomena*. New York/ Oxford: Oxford University Press.

'The Hubbard Model at Half a Century' [2013], Editorial, *Nature Physics*, 9, p. 523.

Tonomura, A., Osakabe, N. , Matsuda T., Kawasaki T., Endo J.,Yano S., Yamada., H. [1982]: 'Observation of the Aharonov-Bohm Effect by Electron Holography', *Physical Review Letters*, 48, pp. 1443.

Tonomura, A., Osakabe, N. , Matsuda T., Kawasaki T., Endo J.,Yano S., Yamada., H. [1986]: 'Evidence for Aharonov-Bohm Effect with Magnetic Field Completely Shielded from Electron Wave', *Physical Review Letter*, 56, pp. 792-795.

Weisberg, M. [2013]: *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press.

Wimsatt, W.C. [2007]: *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.

Wilczek, F. [1982]: 'Quantum Mechanics of Fractional-Spin Particles', *Physical Review Letters*, 49, pp. 957-959.

Wilczek, F. (*ed.*) [1990]: *Fractional Statistics and Anyon Superconductivity*. Singapore: World Scientific.

Wilson, A.H., [1936]: *The Theory of Metals*. Cambridge: Cambridge University Press.

Yi, S.W. [2002]: 'The Nature of Model-based Understanding in Condensed Matter Physics', *Mind* & *Society*, 3(1), pp. 81-91.

Elay Shech,
Department of Philosophy
Auburn University
eshech@auburn.edu

Axel Gelfert,
Institute of History and Philosophy of Science, Technology, and Literature
Technische Universität Berlin
axel@gelfert.net

STUDIA METODOLOGICZNE
NR 39 • 2019, 233–252
DOI: 10.14746/sm.2019.39.9

Robert Mróz

Value judgements and economic models – a Weberian perspective

Abstract. The paper argues for the need to introduce analysis of value judgements into literature on economic modelling, which does not currently deal with this topic. It starts with a prescription formulated by Max Weber, that because social science is so permeated with value judgements (such as acceptance of certain ethical values and policy ends, or some methodological convictions), social scientists should openly state values and policy ends they accept while doing research. From this, a meta-theoretical prescription is formulated: whenever analysing a piece of research as its user or methodologist, value judgements expressed or assumed by the author need to be taken into account. If this is so, then a meta-theory of how to identify these components will be useful. As economics is a model-based science, it is desirable that this meta-theory be about models, or be part of a broader theory of models or modelling. Uskali Mäki's "model of a model" is an example of such theory of models that is easy to amend and refocus to account for this requirement.

Keywords: Max Weber, economic modelling, value judgements, model of a model.

1. Introduction

A question of objectivity is arguably more pressing in social than in physical sciences. Numerous factors contribute to this, from the fact that social scientists are themselves part of society and not only observe but also participate in social life, to the observation that political entanglement of scientists, most notably economists, creates incentives to substitute value judgements for actual analysis. But this characterisation is already problematic as it reveals that there are a few meanings of "objectivity" one can focus on.

One of these meanings relates to ontology – there is an objective social world, which means that scientific theories are capable of uncovering reality. A related statement would be that social facts are at least partially independent of the minds and thoughts of scientists and other agents acting in society. Another meaning relates to epistemology – theories produced by scientists can be empirically and theoretically adequate, i.e. well-grounded on the basis of some reasons deemed to be "right reasons" (so scientific method rather than, say, political bias). Yet another could be seen as merging objectivity and intersubjectivity – conclusions of scientific inquiry tend to converge towards a consensus among scientists. In this paper, however, I am interested in still different sense of "objectivity", namely the one relating to the fact-value distinction. On this understanding of the term social science would be objective if scientific inquiry were value-free and unconcerned with particular interests.

In fact, even here there are two separate issues at play. First: if it is at all possible for research in social sciences, particularly economics, to be independent of researcher's convictions. Second: even if it is impossible to achieve absolute separation of research and convictions, does that preclude objectivity in economics? This latter question is the core of Max Weber's discussion of objectivity of social science, most notably in his essay "'Objectivity' in Social Science and Social Policy" [1904] (published in English as part of "The Methodology of Social Sciences" [1949]). Since Weber, there was much debate about value-neutrality of economic research, from Robbins [1932] to Sen [1970], but after that the issue seems to have died down among economists. Indeed, it seems that mainstream economists today readily accept that their inquiries, at least as long as explicitly classified as *positive* – as opposed to *normative* – economics, are value-free (even in popular press, as in Mankiw [2011]). As a corollary, it can be observed that while welfare economics tends to be seen as involving value judgements, many economists hold that it is separate from other branches of economics, thereby indirectly upholding the fact-value distinction [cf. Putnam & Walsh (eds.) 2011].

But it is far from obvious that economists are right to claim this, and the issue continues in philosophy of economics (as evidenced, for example, by

Putnam & Walsh (eds.) [2011][1]). What is curious, though, is that this strand of inquiry does not really connect to what is arguably the most prominent area in today's philosophy of economics – research regarding economic models and modelling. Models are rightly considered the most important tool in the arsenal of modern economists – economics is a model-based science [cf. Morgan 2012]. But this means that, as far as it is productive to discuss the role of values in scientific practice in economics, it is also productive to connect this topic to the topic of economic modelling. In other words, when providing an account of what a model is, or how the modelling practice looks like in research, a philosopher of economics might want to consider the role of values in modelling.

The understanding of "values" and "value judgements" in this paper is consistent with Weber's, but it also expands on Weber's thought in a way consistent with more modern accounts. In what follows these terms will be taken to refer to ethical convictions of researchers (such as: "it is desirable to reduce income inequality") or to policy ends assumed in a piece of research (as in: "suppose the government wants to reduce income inequality; this model examines various ways to go about this issue to determine which one is the most effective under our assumptions"), as well as to so called methodological value judgements. The latter understanding of value judgements results in claims that methodological decisions of scientists can constitute value judgements as well.

Yet another possible explication leads to the conclusion that to prefer one research area over another is itself a value judgement, as such preference reveals that the researcher in question thinks some matters more important than others. These various meanings of this paper's central terms will be expounded upon in section 2.

Therefore, this paper proceeds as follows. It begins, in section 2, with a classic take on the role of values in social science, i.e. Weber's understanding of what it means for social science to be objective. In short, the conclusion here is that even though economics in never truly value-free, this does not mean

[1] For discussions of this topic outside of economics see e.g. Kincaid et al. [2007], Forge [2009]. For Polish-language readers, Czarny [2004] is a comprehensive discussion of value judgements in economics.

it is not objective as long as aims and values of a given researcher are clearly articulated. This description can then be taken further and applied to economic science as it is actually practised. In practice, is the Weberian prescription followed? This is an important question, but a different issue is more central to the aim of this paper – if there is a need to be able to identify values and political ends in research in general, and in modelling in particular, a theory of this will be useful. Therefore, section 3 focuses on one of well-known accounts of economic models, by Uskali Mäki [2009, 2011, 2013]. Mäki's "model of a model" goes beyond standard realist descriptions of models as tools for representing some parts of the world. It includes a modeller using the model for a particular purpose, and an audience the model is addressed to. The claim here is that there might be a need to add more detail to the description of the modelling agent – so as to include her value judgements inherent in modelling practice. It will be shown that Weber's account is well suited to become the basis of such action. As will be explained, to proceed in this manner is different than to examine the purpose of the model or its intended audience (which Mäki does). In conclusion, the view advocated here provides a fuller account of economic models.

2. Weberian view on the objectivity of social sciences

As succinctly stated by Dahrendorf [1987, p. 577], Max Weber claimed that "statements of fact are one thing, statements of value another, and any confusing of the two is impermissible." It would seem, then, that Weber advocated that social science be completely value-free. It is rather clear, however, that research in economics (and other disciplines, for that matter) is riddled with value judgements, even if a researcher in question is not aware of that.

On the most general level, to choose your study area over some other is a value judgement in itself, as it demonstrates judgement on what is more, and what is less, important. Also, all scientific reasoning presupposes normative commitments such as primacy of logic and evidence over authority. Furthermore, there are *methodological* value judgements, which concern weighing simplicity, ability to express the model in precise mathematical terms, internal and external consistency, predictive power, etc., against each other. They enter the scene when one is interpreting inferences, data, models, and so on. For example [cf. Shrader-Frechette 1993, ch. 3; 1994, ch. 3]: a researcher committed to the methodological value of predictive power can claim that a model that was not tested against data does not provide sufficient basis for drawing conclusions. On the other hand, someone holding external consistency in high regard might claim that untested models are sufficient basis for drawing conclusions provided that they are at least consistent with data. As we will see, this kind of value judgements is one of the two most important ones from the perspective of this paper.

On a more specific level, to accept cardinal utility measures for the purpose of engaging in welfare economics is a value judgement as it assumes all individuals measure utility on the same scale. Yet another one would be to then engage in interpersonal comparisons of utility (against Robbins [1932], among others). To accept – as is customary in economics – that it is desirable to increase efficiency is a value judgement as well. This last example could be described, in line with the tradition in analytic philosophy, as an *evaluative* judgement – a judgement which corresponds to a statement of value; most commonly, but not exclusively, they correspond to some ethical claim that something is good or bad, just or unjust, desirable or undesirable, etc. [Baujard 2013] Using, and reformulating, an example about income inequality given in the introduction, we could state a judgement such as "current levels of income inequality in country A are unjust". Building on this example we could formulate a *prescriptive* judgement, one that corresponds to a statement of recommendation [*ibidem*] - "we should aim to reduce income inequality in country A."[2] The other example given in the introduction ("suppose the government wants to reduce income inequality; this model examines various

[2] The third traditionally recognised class of judgements are *descriptive* judgements, which correspond to statements of fact. For example: "income inequality in country A took value X in January 2018." There are, of course, well-recognised problems in demarcating descriptive, evaluative, and prescriptive judgements (for a short general discussion see Baujard 2013; for a succinct analysis of evaluative components of descriptive judgements see Sen 1980; for various distinctions within the class of prescriptive judgements see Sen 1967). But these problems should not concern us here as this distinction is presented only to organise thinking and nothing of substance hinges on it.

ways to go about this issue to determine which one is the most effective under our assumptions") can be seen as paving the way for a kind of conditional prescriptive judgement. Suppose the model in question allows a conclusion that method M of reducing income inequality is the most effective under the model's assumptions. Then it can be claimed that "under our assumptions, we should apply method M if we want to reduce income inequality." This is a conditional prescriptive judgement as it prescribes the course of action conditional on: 1) the policy goal/end being reduction of income inequality; 2) assumptions of the model holding[3].

It is this kind of evaluative and prescriptive judgements, concerning ethical choices and policy ends, that forms the second type of value judgements that is important for the purpose of this paper – they are the ones Weber primarily addressed in his writings. Of course, evaluative or prescriptive judgements are not necessarily concerned with ethics or policy – one can have a variety of methodological prescriptions, to name just one example. (In fact, Weber's own prescription that scientists should be clear about their values is a methodological prescription.) However, what Weber was particularly focusing on were values entering consideration when one is analysing things from the perspective of a specific end (see below) – which is why evaluative and prescriptive judgements are presented in such context in current paper. It is also why in some contexts below terms "value" and "end" can be used interchangeably.

Weber, being of course aware that it is impossible for social sciences to be utterly devoid of values, avoided troubling claims that they should be entirely value-free. He did this by applying a two-step approach [Hoenisch 2003]. First, he believed that ultimate values could not in any way be justified through value-free analysis. So, from the point of view of science, it is impossible to choose any, say, ethical or political theory over any other without taking into consideration a value or an end that is to be realised – and this value or end is dictated by convictions of a particular scientist [cf. Lassman & Speirs 1994][4].

[3] The question of what it means for the model's assumptions to "hold" (do they need to be true?; approximately true, whatever that might mean?; is it enough for them to give the model sufficient predictive power?; etc.) is an important one in philosophy of economics, but it falls outside the scope of this paper.

[4] This does not mean, of course, that *no* reasonable discussion about values is possible.

But then comes the second step – once the value or end has been chosen, a value-free analysis is entirely possible of what the most effective means of attaining that end are. Such approach is visible in Weber's writings, where he often declares openly what his standpoint will be in a given essay. At one point in his "The Nation State and Economic Policy" [1895/1980], he clearly states that the situation at Germany's eastern border will be assessed from the standpoint of the German people. Thus, he provides the basis (the end to be achieved) against which the assessment of facts needs to proceed. By doing this, he also paves the way for conditional prescriptive judgements in the form of "if the good of the German people is the value we want to uphold in our policy, then we should engage in action A on Germany's eastern border."

In the same work he underscores that "…the ideals we introduce into the subject matter of our science are not peculiar to it, nor are they produced by this science itself."[5] This means that even though economics necessarily involves value judgements, these value judgements sit above the subject matter of economics and can – and in fact *should* – be disentangled from economic facts. Practical prescription which follows is that a scientist should be open and clear about her values when engaging in an inquiry. Using the terms just introduced, we could say that a scientist should be open and clear about evaluative (and methodological) judgements involved in her inquiry because these evaluative judgements then serve as basis for formulating (conditional) prescriptive judgements. This clarity, then, allows for objective analysis of social phenomena from the perspective of a given end.

A caveat is needed before we go further. For the purposes of this paper the Weberian perspective on value judgements is interesting inasmuch as it generates the practical prescription that social scientists should be open about their values. For Weber, these values are understood as acts of *valuation* (*Werthen*), which mean personal judgements, like taking a stand on some issue [cf. Bruun 2001]. But, strictly speaking, the Weberian approach to value judgements is much broader, and has two facets. Practical valuations of a given scientist, her ethical convictions or political views, are one of them. The other, and logically

[5] This is also underscored by Weber in his famous lecture "Science as a vocation" [1917/1989], as well as in "The meaning of "ethical neutrality" in sociology and economics" [1917/1949].

prior, is the concept of *value relation* (*Wertbeziehung*), taken by Weber from the works of German philosopher Heinrich Rickert and then amended. For Rickert, and then for Weber, value relation was a theoretical concept related, essentially, to choosing the object of scientific research which is *valuable*, where *value* (*Werth*) enters in the sense of being interesting, worth knowing about (*wissenwerth*). For Weber, this value relation was culturally determined and an indispensable part of social science *qua* science. What is interesting is decided not by some universal principles that scientists should know (whereas such universality was argued for by Rickert [cf. Bruun 2001, p. 147-8]), or some metaphysical weight of a given problem, but by the interests of the scientist's public. So, in a way, the very act of selecting the material for research has a kind of a value-related component, and this component is tied to the pressing issues of the times that are in the interest of the public (otherwise, the resulting work of the scientist will itself not be "worth knowing about"). But this component is related to values in the *theoretical* sense of "interest". So it is very different from *practical* value judgements of the scientists, which are their valuations. This means Weber can see a kind of value judgements even in the act of choosing the subject matter for research[6], and yet still say that scientific analysis should be free of values understood as valuations [cf. Weber 1917/1949] (at least in the sense explained above, where Weber's two-step procedure was discussed)[7]. Having said that, in what follows I will be focused on the practical prescription that scientists should be open about their values (valuations).

The above description notwithstanding – and as noted above – the mainstream conviction is that economics is by and large a value-free science. In light of foregoing observations on normative content of economic research and on the Weberian stance, this claim could be understood as correct only if what is being claimed is that economists, when engaging in what is usually called positive economics, are merely trying to establish the most effective

[6] Which goes back to the most general type of value judgements involved in selecting a study area and noted at the beginning of this section.

[7] For an extensive discussion of these matters, see e.g. Bruun [2007]. Bruun [2001, 2007] also demonstrates how the concept of value relation led Weber to the well-known concept of ideal types.

way of achieving a given end. But even then this would be a rather weird way of using the phrase “value-free” - if the end is dictated by values, values are always present in economic analysis. So at best there would be a value-free *component* to an economic theory or model, while the whole piece of research would not be considered devoid of normative statements or presuppositions. In other words, the best we can do is to separate values from facts and be clear about this separation.· Such view also has a long tradition in economics, having been radicalised by Robbins [1932] and endorsed by Musgrave [1999] in his discussions of public policy. A methodological prescription of this sort was also accepted in welfare economics when social welfare function was the main topic [e.g. Bergson 1954]. But the point here is, economists are rarely that clear about the value-fact separation.

One has to remember that this paper is written from the Weberian perspective, and on this account some such value-fact separation is possible. I do not think, though, that this perspective, as it is being used in this paper, assumes we need to be able to *fully* separate these two levels *in each and every case*. To uncritically assume something like this would be to disregard many advances in philosophy of science in the last fifty years [cf. Gonzalez 2013], and it would be just wrong. But if one were to follow writings of Putnam [2004], or authors in Putnam & Walsh (eds.) [2011], and claim that there is no strong ontological difference between values and facts, it would not deem any attempt at separating them futile or unnecessary (similar position is defended in Niiniluoto [2009]). On the contrary, it would rather reinforce the need of such separation inasmuch as we are able to do so, if only to avoid as much confusion as possible. And for the purpose of this paper, such lack of clear distinction would also not nullify, but rather reinforce the need to incorporate value judgements in our view of models and modelling practice in economics, as discussed in section 3. In any case, the Weberian perspective is useful for the purpose at hand, especially given that some version of it seems to underpin thinking of many economists, among them those inclined to view economics as largely value-free[8].

[8] In other words, it is not required that we accept Weber’s strong claim that statements of fact are one thing and statements of value another. The Weberian perspective employed in this paper is Weberian because Weber’s methodological prescription – that we should strive to be as

This raises possibilities for philosophical and methodological scrutiny. One recent example of such endeavour can be found in Desmarais-Tremblay [2014]. The author juxtaposes two theories of public finance. One was proposed by Richard Musgrave, who openly admitted his theory was normative. The other was by James Buchanan, who formulated it in response to Musgrave and stated that his theory was positive. Desmarais-Tremblay provides an explication of the Weberian approach to value judgements, and concludes that Buchanan was closer to Musgrave on the normative-positive spectrum than he admitted. This is evidenced by implicit value judgements that were harder to spot than in Musgrave's theory. The Weberian approach, as explicated by the author, provides two distinct tests of whether a theory is normative, one of which we can invoke here. It goes: a theory is normative when its author does not explicitly formulate value judgements that are implicitly present; or when she does, but at the same time explicitly endorses values expressed in these judgements. It is in this latter sense that Musgrave's theory is classified as normative because Musgrave states that he agrees with the position that some so-called "merit goods" need to be provided publicly. On the other hand, Buchanan failed to clearly articulate some judgements in his theory, so his account is normative in the former sense[9].

In any case, the Weberian approach to the value-fact divide provides at least one prescription when it comes to evaluating a piece of research in social sciences: whatever is being said, see what ends are being served and what values are being endorsed[10].

clear about our values as possible – is deemed an important piece of methodological guidance, as well as useful in formulating the meta-theoretical prescription presented in section 3. This meta-theoretical prescription follows fairly straightforwardly from Weber's methodological prescription.

[9] The second test of normativity is that a theory will not be normative even when value judgements are present, provided that valuations are made solely by the agents in the model/theory. Thus, any external intervention by the theorist will mean it is normative. It is the case with Musgrave's postulate that collective action problems related to public goods will not be solved by intra-theory agents and economist must declare that state intervention is required. Contrastingly, Buchanan sought to show that such problems will be solved by the agents, therefore his theory is not normative in this sense.

[10] As Weber's account of value judgements in scientific practice is used here as a starting point for further ideas, I do not purport to present any novel interpretation of his writings on this

3. Model of a model

As for now, three things can safely be asserted on the basis of the foregoing discussion:

1) economics is a science in which value judgements are commonplace;

2) given that it is so, an evaluation of a piece of research needs to involve some mention of these judgements, especially if the author of the evaluated piece did not explicitly explain her stance (this is where the influence of the Weberian approach is most visible);

3) economics is a model-based science.

Given this, it seems rather curious that discussion of how value judgements appear in scientific practice in economics is not really present in the literature on economic modelling. Let us take one of well-established – although, of course, far from uncontroversial – accounts of what is a theoretical economic model, that by Uskali Mäki [2009, 2011, 2013].

Mäki represents one of the most prominent traditions in thinking about economic modelling. It can be termed "isolationism." Traditionally, isolationists are scientific realists (i.e., in the most basic formulation, they postulate that there are real causal relations occurring in the world that is at least partly independent of the agent's thoughts and states of consciousness). The key terms in their accounts are "isolation" and "representation", meaning that a model is a model by virtue of it being a sort of a "stand-in" for the world (target system). The model takes into account only salient characteristics of the target (so it isolates them from all the unimportant features), at the same time somehow representing these salient features, e.g. by resembling the target in some way. Given this basic setup[11], it feels natural to confine our talk about models to simple dyadic model-target relations. But in his recent research Mäki admits

topic. Additionally, the description provided in the text is necessarily brief. Readers interested in a detailed analysis of Weber's stance on values and value-free science should consult Bruun [2007]. For general Weberian methodology, presented in context of important debates of his day, see Ringer [1997] and Eliaeson [2002]. For a recent argument for the continued relevance of Weber's approach to science, see Ghosh [2014].

[11] There are, of course, other ways of explicating what an economic model is. One of the most prominent alternatives to isolationism is what can be termed "constructivism", as represented, e.g., in influential papers by Sugden [2009] and Grüne-Yanoff [2009].

this is not enough and other components enter the relation between the model and the target, making it more nuanced. Therefore, such richer idea of a model looks as follows [Mäki 2013, p. 91]:

"**[ModRep]**
Agent A
uses (imagined) object M as
a **representative** of (actual or possible) target R
for **purpose** P,
addressing **audience** E,
at least potentially prompting genuine **issues of resemblance** between M and R to arise,
describing M and drawing inferences about M and R in terms of one or more **model descriptions** D,
and applies **commentary** C to identify and coordinate the other components."

Issues of representation and resemblance were briefly mentioned above. Introduction of an agent is meant to demonstrate that nothing is a model without being used as such by some modeller. There are also various purposes one can have in using a model, such as isolating some causal relation, exploring the range of possible causal relations, general explanation of a phenomenon, prediction, influencing policy-making, improving a mathematical structure, etc. A model can also vary depending on an audience – it can be directed at specialists in cutting-edge research in the field, or first-year students, or policy makers, etc. According to Mäki, model description will be influenced by the intended audience – mathematical symbols for specialists, metaphors and real-life examples for laymen.[12] Finally, model commentary identifies and coordinates the other components – it should specify purpose and audience, it should also help in understanding e.g. the role of unrealistic assumptions in the model, etc.

[12] It is far from clear to me that model description can always be so easily decoupled from the model itself. In other words, it is not clear that the same core idea (abstract object) expressed with use of mathematics or with use of natural-language metaphors will always be the same model (or even – if it will really be the exact same idea). Certainly, Mäki's account – as any other – is not without its problems, but it should not detract from the main topic of this study.

Such extended “model of a model” can be used here as it will be instructive to show that even on such broad account an important aspect related to value judgements is missing, and it should not be if economics is indeed so permeated with value considerations.

Consider two very different classes of models. One class consists of models in the Austrian tradition, i.e. representing versions of the Austrian Business Cycle Theory (ABCT). The other class consists of models in the Real Business Cycle (RBC) tradition. Objects in both classes (or at least some of them) have arguably the same general target – cyclical fluctuations in economic performance – and purpose – to explain these cyclical fluctuations by identifying their causes. As done by practising economists, they are also intended to represent, or at least it is not implausible that they are. Their intended audience is on some level the same – this being professional economists – but there are obvious differences in that RBC models are aimed mostly at mainstream mathematical economists, and the Austrian research is largely confined to the narrow Austrian audience. Model description is where the difference is most visible, as RBC models are expressed via mathematics with some added interpretation in English, while most Austrian models are formulated only in English, without the use of mathematics. And of course the explanation itself of why the economy is going through boom-and-bust cycles is completely different.

One could stop here, just noting the differences. But then an important *why* question would go unaddressed, namely: *why* are the content of the model, its intended audience, and, maybe most importantly, its description so different if the target and purpose are by and large similar? In fact this is largely due to normative convictions of both schools. Austrians underscore the importance of capital theory and its complexities (the role of time in production processes, interplays between time preference and the structure of production, etc.) [cf. Garrison 2001]. Modern mainstream macroeconomics, together with the New Classical school in which the RBC theory originated, abstracts from capital theory, thus implicitly treating it as unimportant in explanations of business cycles. In other words, there are differences in the content of the *explanans* between these two classes of models. At least some of them trace back to the most “high-level” type of value judgements mentioned in section 2 – the ones involved in attaching importance to one research area over some other.

Further, Austrians explicitly [Hülsmann 2001, p. 36] differentiate their brand of economics from neoclassical and other mainstream schools by insisting that they are focused on analysis and representations of human action, while mainstream economic thought is confined to analysis of quantities of things that are subject to human action. Yet further, Austrians claim that individual human action, which is the proper subject of economic inquiry, cannot be mathematised, or at least not in the style of modern economics. This is an example of a methodological value judgement, which links back to my statement at the beginning of section 2 that this kind of judgements is of importance in current paper.

All these discrepancies (and the list above is far from exhaustive) do not stem from differences of purpose – both schools are trying to explain how business cycles come about. Differing intended audience cannot be a cause as well, for such conclusion would assume that if the audiences were switched and mainstream economists for some reason started speaking to the Austrian audience, they could just adopt models of the Austrian variety and use them like any other model. But simple observation confirms that this could not be the case. A mainstream economist, whether of New Classical or any other variety, just would not accept the way Austrian economics is being done – and *vice-versa*. There is a gulf separating both ways of thinking about the economy.

Part of this gap is explained by methodological value judgements mentioned above. But it is not implausible to suppose that differences not only in methodological value judgements, but also in what I referred to as evaluative and (conditional) prescriptive judgements concerning ethical choices or policy ends, may be able to explain some of it as well. It is not to say that the pre-existing political stance of some group of economists will completely determine the outcomes of their models. It is to say that the overall political stance (as well as some specific view on a given policy matter) of a given economist does have a chance of influencing their thinking while devising a model[13]. This seems to occur in the case analysed by Desmarais-Tremblay, who ascribes some importance to e.g. Buchanan's libertarian sympathies or

[13] This will apply, of course, only to models whose purpose has actually something to do with policy matters. If the purpose of a model is to work out the complexities of some mathematical structure, then such value judgements will not appear.

Musgrave's conviction that there *just are* some goods that should be treated as merit goods. The free-market orientation of both RBC and Austrian theorists follows from their respective theories as a conclusion, so it might seem hard to view it as an implicitly assumed value judgement. Austrians go even further in their claims that central banks should be abolished [cf. Hayek 1990, Salerno 2010, ch. 19]. Again, this follows from their Misesian theory of business cycles [cf. Mises 1949, ch. XX]. But it is, again, plausible to suppose that these kinds of factors will enter the background assumptions of a modeller whenever a new model is designed e.g., to aid policy-makers. As it is not the aim of this paper to work out the complexities of this case study, I leave this as an intuitive supposition at this stage, and will analyse case studies in more detail in an upcoming paper[14].

The above is a simple, and obviously broadly sketched, example showing how one can work to distinguish models not using the components of Mäki's account of models, but using a separate, additional component – modeller's values and value judgements assumed in the model[15]. It would perhaps be more proper, and more interesting, to compare models from the same intellectual tradition that differ neither in purpose nor in intended audience, nor in the type of model description, and yet are different in important respects that can be traced back to modeller's values. Therefore, a fitting case study will be undertaken as the next research step on the topic. Mainstream monetary economics seems to be a promising avenue, especially models attempting to give theoretical underpinnings to decisions of central banks. They are directly policy-relevant, which means they are formulated with specific policy ends and perspectives in mind. They are situated within one general-equilibrium paradigm and use similar mathematical techniques. Yet they are often different with respect to decision rules upon which a central bank in the model is acting [cf. Clarida et al. 2000, Belke & Klose 2013]. It means, of course, that they are

[14] More intuition for suppositions in this vein is supplied by a recent paper by Salter & Luther [2016], who try to show that it is possible to express the ABCT in the mainstream framework, with equilibrium and rational expectations at its centre. If this is really the case (some Austrians would surely object), then the policy conclusions might result not from methodological judgements, but from *bona fide* ethical value judgements.

[15] Mäki's account is thus refocused – his original aim was only to describe and explain what a model is, while here the aim is to be able to classify models according to their characteristics.

different in their assumptions. In our present context the question is: *why* are they different in their assumptions? Are the differences resulting from normative considerations (such as: "we think the central bank should be operating in this and this way"), or do they sit in the positive realm (as in: "given that the aim of the central bank is this and this, its decision rule should be that and that")? It should be visible that these kinds of questions stem directly from the Weberian approach to understanding value judgements in social sciences.

One potential criticism that could be levied against this approach is that to identify all implicit value judgements in an evaluated model, one might have to guess what the author meant in some statements, and maybe even look at her biography, history of political engagement, opinion pieces in popular press, etc., to understand her broader normative commitments. This would render the whole enterprise daunting to the point of impossibility. Such criticism would be misguided, though. First of all, this being a practical prescription, it does not assume that all value judgements need to be uncovered in the process of understanding and interpreting a model. What is important is the attitude and awareness this prescription elucidates. Second of all, as evidenced by Desmarais-Tremblay's paper [2014], such interpretations of economic research are possible and valuable. The author does use some biographical facts about Musgrave and Buchanan to strengthen his argument, but the biggest weight is lifted by a close reading of books and papers. In general, exercises in interpretation is what any practising scientist is doing whenever reading any piece of research. Whatthe current paper does is it adds an additional dimension to these readings. In this context it is worth remembering that the meaning of any sentence, theory, or model is not an abstract, objective entity, but it is produced at the intersection of the author's intent and reader's own baggage of experience ("what the reader brings to the reading of the text" - Samuels [1990, p. 9]). So there is no one way of going about this issue.

In any case, Mäki's account of models is useful here as it readily admits inclusion of modeller's values into its scheme, giving something like:

[ModRep2]
Agent A,
expressing value judgements contained in set V,

uses (imagined) object M as
a **representative** of (actual or possible) target R
for **purpose** P,
addressing **audience** E,
at least potentially prompting genuine **issues of resemblance** between M and R to arise,
describing M and drawing inferences about M and R in terms of one or more **model descriptions** D,
and applies **commentary** C to identify and coordinate the other components.

Depending on one's view on the possibility of separation of values and facts, it will not always, or maybe not even often (or maybe even never), be possible to specify *all* elements of the set V. But, as mentioned above, this is not a crucial problem for this approach as long as some elements can be specified. Also, the word "expressing" is used here deliberately (instead of, for instance, "making") to allow for the possibility that in some model value judgements are not being made actively or consciously.

4. Conclusion

What has been said here can be summed up thusly:

1) if Weber is right, then there is a need to distinguish and be explicit about normative and positive components of a given piece of research because normative components will impact policy conclusions;

2) if this is so, then a meta-theory of how to identify these components will be useful;

3) economics is a model-based science, so it is desirable that this meta-theory be about models, or be part of a broader theory of models or modelling;

4) Mäki's "model of a model" is an example of such theory of models that is easy to amend and refocus to account for the requirement stated in 2).

Finally, it is worth noting that the Weberian approach is especially suited for such use. It admits of possibility that economists *qua* economists (as opposed to economists *qua* private persons) will be making value judgements or working with some particular ends in mind (only Robbins's version of this approach would claim such activities have no place in economics). Yet it also states that some separation of values and facts is possible in practice – as opposed to what might be termed "strong non-neutrality of economics" thesis [Mongin 2006, Niiniluoto 2009]. (E.g. Myrdal [1958] claimed that value judgements and factual statements cannot be distinguished logically. This also relates to Putnam's position mentioned in footnote 2.) These two Weberian claims give rise to point 1) above and point 2) at the beginning of section 3.

Acknowledgments

The research here presented was financed by a grant from the National Programme for the Development of Humanities (grant no. 0266/NPRH4/H2b/83/2016).

References

Baujard A., (2013), "Value judgments and economics expertise", Working paper GATE 2013-14.

Belke A., J. Klose, (2013), "Modifying Taylor reaction functions in the presence of the zero-lower-bound. Evidence for the ECB and the Fed", Economic Modelling 35, p. 515-527.

Bergson A., (1954), "On the concept of social welfare", The Quarterly Journal of Economics, 68, p. 233-252.

Bruun H.H., (2001), "Weber on Rickert: from value relation to ideal type", Max Weber Studies, 1, p. 138-160.

Bruun H.H., (2007), Science, Values and Politics in Max Weber's Methodology (New Expanded Edition), Aldershot, Ashgate Publishing.

Clarida R., J. Galí, M. Gertler, (2000), "Monetary policy rules and macroeconomic stability: evidence and some theory", The Quarterly Journal of Economics, 115, p. 147-180.

Czarny B., (2004), Pozytywizm i falsyfikacjonizm a sądy wartościujące w ekonomii. Seria Monografie i Opracowania, nr 535, Warszawa, Szkoła Główna Handlowa.

Dahrendorf R., (1987), "Max Weber and modern social science", [in:] Max Weber and his Contemporaries, W.J. Mommsen, J. Osterhammel (eds.), London, Allen & Unwin, ch. 37.

Desmarais-Tremblay M., (2014), "Normative and positive theories of public finance: contrasting Musgrave and Buchanan", Journal of Economic Methodology, 21(3), p. 273-289.

Eliaeson S., (2002), Max Weber's Methodologies: Interpretation and Critique, Cambridge, Polity.

Forge J., (2009), "Science is value-laden: you can count on that!", Metascience, 18, p. 257-260.

Garrison R., (2001), Time and Money: The Macroeconomics of Capital Structure, New York, Routledge.

Ghosh P., (2014), "Beyond methodology. Max Weber's conception of *Wissenschaft*", Sociologia Internationalis, 52(2), p. 157-218.

Gonzalez W.J., (2013), "Value ladenness and the value-free ideal in scientific research", [in:] Handbook of the Philosophical Foundations of Business Ethics, C. Luetge (ed.), Dordrecht, Springer Science+Business Media B.V.

Grüne-Yanoff T., (2009), "Learning from minimal economic models", Erkenntnis, 70, p. 81-99.

Hayek F.A., (1990), Denationalisation of Money – The Argument Refined, London, The Institute of Economic Affairs.

Hoenisch S., (2003), Max Weber's View of Objectivity in Social Science*,* On-line (15.11.2017): http://www.criticism.com/md/weber1.html#s8

Hülsmann J.-G., (2001), „Garrisonian macroeconomics", The Quarterly Journal of Austrian Economics, 4(3), p. 33-41.

Kincaid H., J. Dupré, A. Wylie (eds.), (2007), Value-free Science?: Ideals and Illusions, Oxford, Oxford University Press.

Lassman P. & R. Speirs, (1994), „Introduction", [in:] Weber: Political Writings, P. Lassman, R. Speirs (eds.), Cambridge, Cambridge University Press.

Mankiw N.G., (2011), "Know what you're protesting", New York Times, Dec 3.

Mäki U., (2009), "MISSing the world. Models as isolations and credible surrogate systems", Erkenntnis*,* 70(1)*, p.* 29-43.

Mäki U., (2011), "Models and the locus of their truth", Synthese, 180, p. 47-63.

Mäki U., (2013), "Contested modeling: the case of economics", [in:]Models, Simulations, and the Reduction of Complexity, U. Gähde et al. (eds.), Berlin/Boston, De Gruyter.

Mises L. v., (1949), Human Action New Haven, Yale University Press, ch. XX.

Mongin P., (2006), "Value judgments and value neutrality in economics", Economica, 73, p, 257-286.

Morgan M.S., (2012), The World in the Model: How Economists Work and Think, Cambridge, Cambridge University Press.

Musgrave R.A., (1999), "Public finance and public choice: two contrasting visions of the state", [in:] The Nature of the Fiscal State: The Roots of My Thinking, R.A. Musgrave, J.M. Buchanan (eds.), Cambridge, MA, MIT Press, p. 29-49

Myrdal G., (1958), Value in Social Theory, London, Routledge.

Niiniluoto I., (2009), "Facts and values – a useful distinction", Acta Philos Fennica*,* 86, p, 109-133.

Putnam H., (2004), The Collapse of the Fact/Value Dichotomy and Other Essays, Cambridge, MA, Harvard University Press.

Putnam H., V. Walsh (eds.), (2011), The End of Value-Free Economics, London, Routledge.

Ringer F., (1997), Max Weber's Methodology: The Unification of the Cultural and Social Sciences, Cambridge, Harvard University Press.

Robbins L., (1932), An Essay on the Nature and Significance of Economic Science, London, Macmillan.

Salerno J.T., (2010), Money, Sound and Unsound, Auburn, Ludwig von Mises Institute, ch. 19.

Salter A.W., W.J. Luther, (2016), "The optimal Austrian Business Cycle Theory", [in:] Studies in Austrian Macroeconomics (Advances in Austrian Economics, Volume 20), Steven Horwitz (ed.), Bingley, Emerald Group Publishing Limited, p. 45-60.

Samuels W.J., (1990), "Introduction", [in:] Economics as Discourse. An Analysis of the Language of Economists, W.J. Samuels (ed.), Boston, Kluwer Academic Publishers.

Sen A., (1967), "The nature and classes of prescriptive judgments", The Philosophical Quarterly, 17, p. 42-62.

Sen A., (1970), Collective Choice and Social Welfare, San Francisco, Holden Day.

Sen A., (1980), "Description as choice", Oxford Economic Papers, 32, p. 353-369.

Shrader-Frechette K.S., (1993), Risk and the Case against Geological Disposal of Nuclear Waste, Berkeley, University of California Press.

Shrader-Frechette K.S., (1994), Ethics of Scientific Research, Lanham, Rowman & Littlefield Publishers.

Sugden R., (2009), "Credible worlds, capacities and mechanisms", Erkenntnis, 70, p. 3-27.

Weber M., (1895/1980), "The national state and economic policy (Freiburg address)", Economy and Society, 9(4), p. 428-449.

Weber M., (1904/1949), "'Objectivity' in social social science and social policy", [in:] The Methodology of Social Sciences, E.A. Shils, H.A. Finch (eds.), Glencoe, The Free Press.

Weber M., (1917/1949), "The meaning of "ethical neutrality" in sociology and economics", [in:] The Methodology of Social Sciences, E.A. Shils, H.A. Finch (eds.), Glencoe, The Free Press.

Weber M., (1917/1989), "Science as a vocation", [in:] Max Weber's "Science as a Vocation", P. Lassman, I. Velody, H. Martins (eds.), London, Unwin Hyman, p. 3-32.

Robert Mróz
Faculty of Economic Sciences, University of Warsaw
Correspondence address: Międzynarodowa 64/66a/48, 03-922 Warszawa
E-mail: rmroz@wne.uw.edu.pl

STUDIA METODOLOGICZNE
NR 39 • 2019, 253–296
DOI: 10.14746/sm.2019.39.10

Jarosław Boruszewski, Krzysztof Nowak-Posadzy

From integration to modelling. On a neglected function of the methodology of humanities

Abstract. The aim of this paper is two-fold: to engage the contemporary discussion about the nature of relations between different scientific disciplines, as well as to disentangle the concept of integration of sciences from superstructures of rival proposals. The authors start with a critical analysis of the Polish contribution to the discussion about the nature of integration of sciences from the second half of the XXth century. Such a step is followed by elaborating a refined account of integration and by disentangling the concept of integration from superstructures of rival proposals – unification and interdisciplinarity. On the grounds of such a refined account the authors deliver a reconstruction of a successful scientific integration. In doing this they introduce the idea of connective knowledge as generated by the methodology of humanities. After reconstructing the successful integration trial, in the concluding remarks their account of integration is specified and summarized.

Keywords: methodology of humanities, integration, unification, interdisciplinarity, model of primitive magic syncretism, Poznań Methodological School.

1. Introductory remarks: integration of sciences as a challenge for methodological research

Addressing the question of how to theoretically grasp the relations between different scientific disciplines has a long tradition in the systematic reflection on science across the academia. Among the most prominent approaches that were formulated and advocated for over the years are the following three: unification approach [cf. Kitcher, 1981; Petkow, 2015], interdisciplinarity approach [cf. Klein, 1990, 2010; Lattuca, 2001] and integration approach [cf. Bechtel, 1986a; Mitchell, Dietrich, 2006; Gerson, 2013]. The peculiarity of the discussions that took place in the second half of the XX century in Poland was the special focus on the third approach, namely the integration of sciences. Undoubtedly, merits for thorough study of integration of sciences go to scientific journals such as *Methodological Studies* (pol. *Studia Metodologiczne*), *Problems of Science of Science Quarterly* (pol. *Zagadnienia Naukoznawstwa*) and *Philosophical Studies* (pol. *Studia Filozoficzne*), which created institutional conditions for exchange of scientific information, coordination of research and consultation of results. Among questions that were attempted to address during this discussion the three following are of special importance to our further investigations:

- definition of integration;
- levels of integration;
- structure of integration.

The paper is organized as follows. In the next section we will sketch the Polish contribution to the discussion about the definition, levels and structure of integration of sciences, as well as will offer our refined account of integration. This will be followed by an attempt to disentangle the concept of integration of sciences from superstructures of rival proposals, namely that of unification and interdisciplinarity (section 3). On the grounds of our refined account on integration we will deliver a reconstruction of a successful scientific integration trial. In doing this we will introduce the idea of connective knowledge as generated by the methodology of humanities (section 4). Finally, in the concluding remarks we will specify and summarize our account on integration of sciences.

2. Underlabouring for clarification of the idea of integration

2.1. Problems with definition of integration

As regards the first question – definition of integration, during this discussion it was recognized that the lack of commonly accepted definition of integration was due to the fact that the very term ‘integration’ has been systematically equipped with many different meanings [Lazari-Pawłowska, 1975, p. 32]. To tackle with such a polysemy, some initial conceptual distinctions and partial findings were offered:

> "there are many misunderstandings in the understanding integration as such. First of all, it should be underlined that processes of bringing together various disciplines *need not be related to their merging* (pol. *scalanie*) that is gradual disappearance of some of them; on the contrary – by *participating in integration processes* particular branches of research can gain conditions for a fuller development" [Topolski, 1965, p. 6; emphasis added; authors' translation];
>
> "Let us start with a question – what the integration of science is. This term does not mean *merging* various disciplines but rather aiming at *connecting* (pol. *wiązanie)* them, while keeping their autonomy" [Maisel, 1973, p. 80; emphasis added; authors' translation];
>
> "Organic integration versus external integration. The second element of the (…) opposition can be characterised relatively easily. External, „mechanical" integration, i.e. – to use a somewhat humorous name – „*bookbinder*" integration takes place when one *juxtaposes* (pol. *zestawiać ze sobą*) mechanically, for instance in the frame of a monograph, research results from various disciplines concerning – *prima facie* – the same subject-matter" [Kmita, 1975, pp. 8-9; emphasis added; authors' translation].

The above-mentioned quotes allow us to identify keywords which we will use to explicate the meaning of the term 'integration':

- juxtaposition (pol. *zestawienie*);
- connection (pol. *powiązanie*);
- merging (pol. *scalanie*).

In the light of the foregoing clarifications, integration is about connecting findings of various scientific disciplines; it is not only a juxtaposition and it is not yet a merger. Thus, in the first specification integration is neither about juxtaposition nor about merging. It is more than juxtaposition but less than merging. In this sense these discussions and their partial conclusions and suggestions remain valid and maybe they are even more valid today than then. Let us take the following example of clarification of integration from more recent literature:

> "Integration thus means *more than simple juxtaposition* of efforts in the same location, and more than relationships that consist solely of market relations. Rather, it includes

> coordinated efforts to pose and solve new research problems that can redefine specialty boundaries. (…) Integration of specialties is *almost never complete*, in the sense of a full epistemic and organizational *merger of two specialties*. Instead, both epistemic and organizational integration are *partial*; that is, lines of research in two different specialties conduct their work using common concerns, approaches, or styles, *without merging* or abandoning their other concerns" [Gerson, 2013, p. 516; emphasis added].

Given the above findings, in the first place one should differentiate between integration and juxtaposition. Even though we do not yet have at our disposal a commonly accepted definition of integration, we do have some baseline common vocabulary to discuss this notion and its cognates, such as interdisciplinarity or multidisciplinarity [Holbrook, 2013, p. 1866]. By referring to this vocabulary we can state that there is a minor discord among the commentators involved regarding the latter of the two concepts. Multidisciplinarity is in fact a juxtaposition of findings from two or more disciplines regarding a given problem [e.g. Miller, 1982; Richards, 1996; Klein, 2010]. A multidisciplinary juxtaposition is not about an integrative connection, thus some slightly ironic names, such as 'bookbinder integration' or 'mechanical integration' appear. They are meant to point out that in fact it is not integration, or else that it is some pseudo-integration. Therefore, multidisciplinarity as such is not the subject of our investigations, although we do refer to it while contrasting multidisciplinarity, unification and integration in terms of discipline-specific world-views.

The issue looks completely different in the case of interdisciplinarity: there are clear-cut tendencies "to glue" interdisciplinarity to integration and that interdisciplinarity presupposes or implies integration. Such tendencies are predominant and widespread to the point that, as J. Britt Holbrook put it, "to question whether ID [interdisciplinarity] involves integration is almost heretical" [Holbrook, 2013, p. 1877]. In the following sections we will try to show that it can be legitimately questioned. It should first be noted that scopes of terms 'interdisciplinarity' and 'integration' intersect in a non-empty way. Another issue is to disentangle relations between the concepts of integration and unification, because contemporary discussions about integration of sciences are still burdened with unificatory intentions. It can be illustrated by interchangeable use of terms 'integration' and 'unification' or by excessive use of the expression ‘integration or unification’ [e.g. Grantham, 2004; Kaplan, 2017; Driscoll, 2018].

1.2. Difficulties with levels of integration

As regards the second question – the levels of integration, the main emphasis in the discussion that took place in Poland in the second half of the XXth century was put on distinguishing three levels of integration:

- *epistemic-methodological* – that relates to possible mutual impact of at least two different disciplinary systems of knowledge;
- *practical-institutional* – that relates to possible interaction of at least two types of specialists or groups of specialists with different disciplinary background;
- *ideational* – that relates to possible interplay between at least two types of so-called social methodological consciousness that consists, among others, in different discipline-specific world-views [Łojewska, 1976, pp. 57-61].

Such a proposal finds support in a growing number of works from recent philosophical and methodological literature on integration of sciences. Let us recall here only one case that make use of the concept of levels of analysis. In the first the concept in question has been already clearly explicated and systematically utilized [Mitchell et al., 1997, pp. 103-125; Mitchell, 2009]. According to this proposal, there is no basic level of analysis to which other can be reduced. However, it does not imply that all levels are isolated from one another. The question is rather how to carefully identify connections between them, remembering at the same time not to analytically confuse them.

Let us now go back to the first level of integration – the epistemic one. It relates to possible mutual impact of at least two different disciplinary systems of knowledge. During the discussion in Poland the issue of exploratory potential carried by integration trials was signaled. Let us illustrate this issue by recalling selected quotations from the discussion:

> “We mean here the form of integration of science, which is about(…) *connecting (and not merely mechanically juxtaposing) results* of research from various disciplines, *ex-*

> *ceeding however beyond the task of suggesting ideas*" [Topolski, 1965, pp. 7-8; emphasis added; authors' translation];
>
> "There is an expectation that all humanities (…) should take into account *synchronic connections* between particular forms of social consciousness (…) It therefore imposes (…) the need to *integrate research* on the methodological basis which *respects* – what is especially worth underlining – *a relative autonomy of each of the disciplines*" [Kmita, 1973a, pp. 79-80; emphasis added; authors' translation].

These remarks are supported by contemporary scholars working on the idea of integration of sciences. Let us take the following example of clarification of the exploratory potential of integration from the literature:

> "the linkage between the fields was *discovered only after a critical reconceptualization* occurred in each field separately (…) The integration of research (…) involved identifying relationships between entities that had been studied independently that allowed researchers in each field *to learn new information* about the entities that were of primary interest to them" [Bechtel, 1986a, pp. 45-46; emphasis added].

In a sense this discussions and their partial conclusions and suggestions remain valid and maybe they are even more valid today than then. In recent discussions it is indicated that integration can have an explanatory goal although it is not a necessary condition. Its heuristic goal is more and more often underlined, i.e. that integration in practice is not directly oriented towards providing explanations [Grantham, 2004, p. 144]. Integration is often guided by exploratory questions, for instance of 'what if' type, which do not set a narrowly determined scope of research and therefore play an important integrating function. Exploratory questions are open-ended and they tend to sustain and help integrate research activities. Integration in the exploratory use "can produce novel insights into […] phenomena, stimulate new fields of research, and generally reconfigure expectations of scientific practice" [O'Malley, Soyer, 2012, p. 58]. Novel insights into research phenomena can take the shape of answers to novel questions, models, hypotheses or new research methods. It has to be remembered though that integration is not a goal of science *per se*. Quest for integrated research perspective "can be a useful heuristic, but it should be viewed as a heuristic, not as *the* aim of science" [Waters, 2017, p. 104, emphasis in original]. Heuristic quest for novel insights manifests especially in affecting the shape of theoretical models and the conceptualization of an object of research. In this respect successive integration "might lead to the transformation, or reshaping, of one model or *conceptualization of the system of interest* by another" [Plutynski, 2013, p. 470; emphasis added]. Thus a successful integration, when considered on the level of the system of scientific knowledge, may provide new data, new or refined modeling strategies, as well as new or refined conceptualizations of an object of research. It means that integration carries some exploratory potential.

This is how we arrived at the second level of integration – practical-institutional It relates to possible interaction of at least two types of specialists or groups of specialists with different disciplinary background. During the discussion in Poland the question of social and institutional arrangements, as distinct from theoretical and methodological conditions, that make possible integration of two (or more) kinds of research was raised [Kmita, 1975, p. 8]. Let us illustrate this issue by recalling selected quotations from the discussion:

> "*Institutional integration is an important aspect* of integrating scientific research because cooperation of large teams of specialists requires such an organizational structure of research units which would stimulate research ventures and secure the process of their realization" [Łojewska, 1976, p. 57-58; emphasis added; authors' translation];
>
> "*integration of scientific research from the organizational side* implies that researchers from various specialties take up joint research. This way of understanding integration can be described simply as cooperation which can take the form of a) a team research using the same (or similar) method for various problems b) a multi-faceted investigations of one problem using different ways or c) a joint research in which different approaches intersect" [Maisel, 1973, p. 81; emphasis added; authors' translation];
>
> "integration of sciences is related to the most important cognitive, methodological and *organizational tasks of science*. In this last domain one should equally strictly differentiate between *the autonomy of cognitive processes and organizational matters* as well as properly defined practical postulates" [Czartoryski, 1967, p. 13; emphasis added; authors' translation].

Distinguishing this level may be considered hardly insightful and the sociology of science or studies on technology and science address this question in a detailed way. But still, investigating it in the context of integration of sciences and focusing on relations to other levels may provide new insights. There are some works from the philosophical and sociological literature that clearly distinguish this level of integration from the epistemic one. According to Elihu M. Gerson, there are two kinds of integration – epistemic and organizational. The latter "consists of the ways that the work of laboratories, associations, universities, sponsors, and other organizations mesh and change in forming the system of research institutions" [Gerson, 2013, p. 515]. As there are no simple and clear-cut relationship between these two kinds of integration, what is needed is to identify various intersection in which the epistemic integration interacts with the organizational one. What is more, as Wim J. van der Steen put it, the question is not only about the intersection of the epistemic and practical kinds of integration. According to him, one has to be aware that materialization of only the institutional conditions (e.g. exchange, cooperation, publications) without fulfilling the theoretical-methodological ones is not enough for the successful integration. To avoid the accusation of pseudo-integration both kinds of conditions have to be met [van der Steen, 1993b, p. 349].

As it has been already stated, distinguishing the epistemic and practical levels of analysis of integration of science is quite common in recent literature. What distinguishes, however, the discussion about integration that took place in Poland in the second half of the XXth century was the identification of the third level of analysis of integration. This level relates to possible interplay between at least two kinds of social methodological consciousness that consists, among others, in different discipline-specific world-views. Let us begin the investigation into this level of analysis by shortly clarifying the idea of methodological consciousness. This idea was of major importance to the Poznań Methodological School[1]. As Jerzy Topolski addressed the question to the audience of practicing historians: "[m]ethodological principles, together with the ideal of science and the view of the world and Man (…) constitute what is called methodological consciousness" [Topolski, 1985, p. 149]. Methodological

[1] The Poznań Methodological School is an inherent part of the Polish (and not only) intellectual landscape. It was one of the more unique and creative philosophic-methodological *Denkkollektiv* in the post-war Europe. It was founded in the mid-60s of the XX century by such scholars as Jerzy Topolski (1928-1998), Jerzy Kmita (1931-2012) and Leszek Nowak (1943-2009). For a detailed discussion about the meta-methodological characteristics of the Poznań Methodological School, see: [Boruszewski, Nowak-Posadzy, 2017].

consciousness contains, apart from the cognitive norms and methodological directives, also the researcher's world-view (*Weltanschauung*). All these components are objects of methodological reconstruction. The structure of researchers' methodological consciousness consists of five major components:

- cognitive norms;
- methodological directives;
- meta-scientific attitudes concerning aim of scientific cognition;
- world-view (*Weltanschauung*);
- vision of Man.

This was an interesting attempt to conceptualize the third level of integration in terms of methodological consciousness with special focus on world-view as carried by different disciplines (specialties) engaging the integration trials. As the concept of world-view is not new in discussions from the field of philosophy of science [cf. Mormann, 2018; Rouse, 2015; Aerts, Van Belle, van der Veken, 1999; Aerts, Apostel, De Moor, Hellemans, Maex, Van Belle, Van der Veken, 1994; Cobern, 1991], let us now explicate its meaning and present the way in which it was utilized in the discussion about integration. By the world-view it was meant a system of beliefs which determines both (i) a set of superior positive values (i.e. ultimate values that the person espouses) and (ii) types of connections between those superior positive values and practical values which are either means to achieve those superior positive values, or to prevent their achievement, or are neutral with respect to their achievement [Kmita, 1991, p. 168; Kmita, 1979, p. 299]. From this perspective a science is not only one of the most important factors shaping the world-view belief system, but also its own internal development and research activities are in a sense influenced by a world-view [Łojewska, 1986, p. 211]. This remarks allow us to differentiate respectively between scientific world-view (pol. *światopogląd naukowy*) and world-view in science (pol. *światopogląd w nauce*). Regarding the former, a world-view is scientific when (i) a scientific knowledge determines ways of materializing the superior positive values and (ii) a scientific knowledge guarantees the possibility of doing this [Kmita, 1979, p. 301].

For the purpose of the article we will focus here only on the very concept of world-view in science which needs, however, some further clarification.

As we investigate here the type of world-views that may interplay in integration trials we have to differentiate between: (i) a world-view behind a science that is an presupposed by science (or a given discipline) a general image of the world and (ii) a scientist's world-view which can be attributed to a particular researcher. The difference between these two is not irrelevant: "in the first case it is about such a world-view which will have methodological and theoretical implications and will thus influence the methods applied and conclusions formulated; in the second case, there can be world-views which do not have the above-mentioned implications and therefore cannot be treated as components of the general world-view accepted by science" [Łojewska, 1986, p. 216]. There appear the question of the relation of this level to the epistemic one. It has to be carefully stated that there seems to be no direct and unambiguous connection between results of scientific research and explicitly accepted or implicitly respected system of world-view beliefs [Ibidem, p. 212].

It was recognized in the discussion in question that the important problem in the disciplinary division of labour does not stem from any excessive specialization occurring in sciences, but rather from an insufficient degree of methodological (self)consciousness [Ziembiński, 1966, p. 3; Kula, 1963, pp. 80-81]. A general issue hardly any participants of the debate in Poland contested was that if one take integration seriously, then the dissemination of "a methodological culture" across disciplines was needed for any successful integration trial [Kmita, 1975; Kmita, 1973a; Kmita, 1973b; Lazari-Pawłowska, 1967, p. 34; Łojewska, 1976; Łubnicki, 1967, p. 20]. Some discrepancies appeared in the debate when it comes to the question of consequences of the interplay between two kinds of methodological consciousness behind disciplines engaged in the integration trial to the discipline-specific world-views. Once more, let us illustrate this issue by recalling selected quotations from the discussion:

> "Investigating the domain of culture as constituted by the science via methodology plays a particularly *important role in shaping a coherent, rational world-view* (…), *scientific image of the world*." [Kmita; 1973a, p. 75; emphasis added; authors' translation];
>
> "The increase of mutual connections between particular disciplines (…) proves useful for the development of science in at least two aspects. First of all, it can *enrich* research methods of particular sciences and secondly, it can increase the scope of questions formulated within a given science via inspiring uses of results of one science for others

> (…) Such cooperation aims at *enriching particular sciences* rather than blur differences between them." [Topolski; 1965, p. 3; emphasis added; authors' translation];
>
> "When specialization was not so advanced, theories used to be very general and not very precise. However, they provided a *coherent world-view*. Today a mosaic of thousands of specialist news cannot always be formed into one coherent world-view which determines our specific place in the world and in the society, as well as indicates us global goals in life. *The longing for all-encompassing cognition* constitutes one of additional causes for quandaries one would like to eliminate by integrating sciences." [Koj, 1975, p. 82; emphasis added; authors' translation].

On the grounds of the above suggestions and some more recent findings in the following table we illustrate the differences between three philosophical agendas in terms of consequences of the interplay of different specialties to the discipline-specific world-view:

Table 1: Integration and discipline-specific world views.

Philosophical agenda	**Kind of relation between research results**	**Consequences to the discipline-specific world-view**
multidisciplinarity	juxtaposing	inviolability
unification	merging	reconciliation
integration	connecting	enrichment

This table needs some comments. The presupposition is that each scientific discipline has its own specific world-view [Miller, 1982; Newell, Green, 1982]. Depending on the philosophical agenda taken and respective kind of relation between interplaying specialties, one can draw at least three possibilities regarding consequences to the discipline-specific world-views. Please notice that we are talking here about consequences and not aims. It is so because discipline-specific world-views are to a large degree rather of tacit (or implicit) nature. Let us now clarify these possibilities:

- multidisciplinary juxtaposing of research findings from at least two specialties tends to be accompanied by the inviolability of initial discipline-specific world-views of contributing specialties; is means that multidisciplinary trial is acknowledged to leave initial discipline-specific world-views invariant (intact);

Such a view is supported by some recent works which argue that "*multidisciplinary research* involves low levels of collaboration, does not challenge the structure or functioning of academic communities or hierarchies and *does not lead to any changes in the worldviews* of the researchers themselves" [Lyall, Bruce, Tait, Meagher 2011, p. 13; emphasis added];

- unificatory merging of research findings from at least two specialties tends to be accompanied by the reconciliation of initial discipline-specific world-views of contributing specialties; it means that unificatory trial is acknowledged to arrive at all-encompassing world-view;

However, a difficulty arises for an unification trial to be successful once world-views behind interacting disciplines are radically different (fully incomparable or uninterpretable). As Donald G. Richards put it, "[s]ynthesis, or interdisciplinary *integration* [in its strong or literal sense] as it is often otherwise expressed, may in some contexts be *infeasible due to irreconcilable differences* in epistemological, or value, terms among the world-views of the contributing disciplines, or the variants of these disciplines" [Richards, 1996, p. 126; emphasis added]. In the situation where "the likelihood of a reconciliation of these divergent world-views is remote, or impossible, those who place primary importance on achieving an integrated view of things *will be forced into* adopting one of these fundamental perspectives and excluding the others" [Richards, 1996, p. 123].

- integrative connecting of research findings from at least two specialties tends to be accompanied by the ideal of enrichment of initial discipline-specific world-views of contributing specialties; it means that integrative trial is acknowledged to exceed (transcend), but not beyond necessity, the narrow scope of discipline-specific world-views of contributing specialties.

Such a view is supported by some recent works which argue that "[i]t is clear, then, that the two can yield a *"synthesis" only in a loose* [a weak, or instrumental], "enriched-view-of the-world" sense. There would seem to be unbreachable epistemological barriers preventing genuine integration in such cases. This does not, however, necessarily de-legitimize this type of interdisciplinary cooperation. *An enriched view of the world, or of a particular issue, is a noble academic objective*" [Richards, 1996, p. 122; emphasis added]. In other words, in integration trials there is always a discipline-specific world-view for which some ‘parts’ of other discipline-specific world-view are incomparable or uninterpretable [Kmita, 2000, p. 202]. However, it does not exclude the possibility of achieving an enriched world-view; only the possibility of both full reconciliation and invariance of different world-views are here denied.

Before we go to the preliminary analysis of the problem of the structure of integration, two additional remarks have to be made. Firstly, apart from the main question of the consequences of the interaction of different specialties to the discipline-specific world-views, there arises similar problem whether discipline-specific world-views behind interacting specialties may counteract successful trials. This question has been already addressed within the philosophical literature. It is sufficient here to state, as Lele and Norgaard recently put it, that: “[t]he first kind of (…) barrier (difference in values) is neither directly discernible nor easily separated from the second (*difference in* theories, models, or *worldviews*)” [Lele, Norgaard, 2005, p. 968]. Our second remark concerns the conspicuous trend toward entanglement of the concept of integration in nostalgic pretense to reconciliation of mutually incoherent disciplinary insights [Newell, Green, 1982] or to re-integration of social sciences which in XIX century were united [Szell, Shujiro, 1993]. Firstly, a successful integration trial does not require complete inclusiveness, sometimes it even requires some exclusiveness and it does not assume a lasting state of unity – it can be merely tentative, contingent, occasionally durable and longer-term it can be undurable. Secondly, a successful integration does not require a complete convergence of respective world-views behind the specialized disciplines engaged in integration trial. In other words, due to the fact that the discipline-specific world-views are at best only partially comparable or interpretable, what is sufficient for integration is only to transcend the narrow scope of discipline-specific world-views to achieve a richness of insight without longing for reconciliation.

1.3. Exposition of the structure of integration

As regards the third basic question – the structure of integration, what is important here is the problem of property attributed to the relation of integration. In contemporary discussions concerning the integration of sciences it is strongly emphasized that integrating exhibits the mutual nature which suggests that the relation of integration is a symmetric one. As it has been already pointed out, such a mutuality is equally discernible on the epistemic level of integration where connecting research findings from various disciplines is followed by reciprocal impact, as well as on the ideational level where enriching of discipline-specific world-views encompasses each side engaged in integration trial. Let us now take the following examples from recent literature to support the claim that the relation of integration in symmetric:

> "Partial integration among specialties is embodied in a system of alliances that span multiple specialties. Partial integration depends on several different kinds of stable coordinative arrangements that join specialties without reducing their epistemic integrity. These include, for example, the use of model data systems; conventions for theorizing and for collecting, analyzing, and visualizing data; and the encumbering of one line of work by another as an "obligatory point of passage" or, more simply, obligation (…) *Such coordinative arrangements are integrative in the sense that two or more lines of work become mutually dependent upon one another for success*" [Gerson, 2013, p. 518; emphasis added];
>
> "In many cases, the commitments ("background assumptions") of one of the integrating lines do not coincide with those of the other, and this typically necessitates *mutual adjustments in order to make the intersection work effectively*" [Ibidem, p. 516; emphasis added].

In light of the above considerations, the property of symmetry, suggested by usage of the expressions 'mutually dependent' and 'mutual adjustments', can be legitimately attributed to the relation of integration. Therefore we conceptualize integration as a mutual and not one-sided relation between at least two parts. However, during the discussion that took place in Poland in the second half of the XX century and was continued in the beginnings of XXI century the question of the symmetric nature of the relation of integration was not explicitly thematized and it was rather taken as a default. The claim that the relation of integration is a symmetric one means that if *x* is being integrated with *y*, then *y* is being integrated with *x*. We thus have two integrated sides. However, this question generates a basic problem whether the property of symmetry can be secured when the relation of integration is conceptualized as a binary relation. In classical discussions about this issue, when the problem of reduction was addressed, integrating was considered as such a binary relation but one of the disciplines involved in the integration trial was given a privileged status. Therefore one of the sides of integration provided means for actually carrying out the integration. This, however, implies that in the last instance the relation in question is not mutual but one-sided. When integration is conceptualized as a binary relation, the property of symmetry is significantly violated: *x* is being integrated with *y*, but by means supplied by either *x* or *y*. Moreover, as the second quotation suggests, the question occurs what is the basis for such a mutual adjustment. Thus two issues need to be distinguished and discussed here:

- the question of the substantial parts being integrated: *what is being integrated?*
- the question of what facilitates the integration: *how is it integrated, by means of what is it integrated?*

Let us now turn to the Poznań Methodological School, whose some leading members (Jerzy Topolski, Jerzy Kmita, Anna Pałubicka) were actively engaged in the discussion about integration. The numerous works by this School can be informally classified into two types: "the canonical" (or "core")[2] ones that are the results of systematic research and "the apocryphal" (or "peripheral") ones that are rather the by-products of many intellectual exchanges. Given such a distinction, the contributions of the members of the Poznań Methodological School to the discussion about the idea of the integration of sciences should be labelled as "the apocryphal" ones. There are at least two successful cases when this "peripheral" idea was put into research practice: the project by Jerzy Topolski to integrate economic history and economic theory [Topolski, 1964, 1991, 2009] and the project of integration of archeology and ethnology [Pałubicka, 1979; Kowalski 1997]. In the 80. and 90. of the XX century the discussion about integration of sciences in Poland weakened and the interest in applicatory works diminished. However, the idea of integration reappeared in the Poznanian circle in the beginning of XXI century under the label of "integrated humanities." We will go back to the issue of successful integration of archeology and ethnology via methodology of humanities in one of the following sections of this paper.

[2] By the "canonical" works we mean the works of Jerzy Topolski on methodology of history and theory of non-source-based historical knowledge [Topolski, 1976], Leszek Nowak on the idealizational theory of science [Nowak, 1980] and Jerzy Kmita on historical epistemology [Kmita, 1988] and socio-regulative theory of culture [Kmita, 1996].

It was the Poznań Methodological School that provided some suggestions and partial answers useful in addressing in more detail the two questions enumerated above. The members of this School were aware that there has to be some basis for the mutual adjustment in integration trials and thus focused mostly on the second question. The answer they provide to the question of how at least two disciplines are integrated and by means of what, is that methodology of science can play a facilitating role in such a trial. This is how the discipline of general methodology of sciences entered the scheme of integration. Let us illustrate this issue by recalling selected quotations from the discussion:

> "*General methodology of science can* give a lot in the area of integration of science since in recent years it has been increasingly seeking for contact with the so-called special sciences thus strengthening their inclination toward methodological reflection in their respective domains" [Topolski, 1965, p. 4; emphasis added; authors' translation];
> "regarding (…) the aspect (…) which encompasses theoretical-methodological conditions of the process of integration of science, philosophy, and in particular the branch called *methodology of science, can play a special role* in the analysis of this aspect (…) I am stating here only possibility, not the factual situation" [Kmita, 1975, p. 8; emphasis added; authors' translation];
> "methodology as *means of integrating* scientific research represented by particular humanistic specialties" [Kmita, 1973a, p. 76; emphasis in original].

While the methodology does not enter the scheme of integration as a substantial part, on equal basis with the disciplines being integrated, still its inclusion in the integration trial suggests integration is a trinary relation. In our view, there are three parts of integration: the sides being integrated (x and y) and the integrative side (the integrator, z) *via* which the integration of x and y is taking place, x is being integrated with y via z[3]. However, in the classical, reductionist account of integration this trinary relation collapses to become a binary one, because one of the integrated sides and integrative side are identical, $y = z$. If we take an anti-reductionist account of integration, a symmetry takes place: if x is being integrated with y via z, y is being integrated with x via z. The integration of x and y is symmetric relative to the

[3] We leave open here the question whether integration of x with y by starting from x and via z, will result in the same effect as if we integrated y with x starting from y and via z.

integrator *z*. Therefore, a symmetric anti-reductionist account of integration assumes the existence of two symmetrical integrated sides and an integrative side which differs from them. In that sense, the sides undergoing integration are not privileged, while the integrative side is not a basic discipline, it is a connective one. Therefore, we propose the general methodology of science as a candidate for the integrator (integrative discipline).

Let us now consider what makes general methodology a discipline capable of actively participating in integration trials and successfully playing the role of an integrator. As the Poznań Methodological School did not provide any clear-cut answers in this respect, we make an attempt at filling this gap by using some recent basic findings of Nicky Priaulx and Martin Weinel. The authors offer an "agenda for addressing the kind of extra-disciplinary knowledge that might help to stimulate, enhance and initiate cross-disciplinary collaboration" [Priaulx, Weinel, 2018, p. 15]. They distinguish two kinds of knowledge: 'of-knowledge' which implies a detailed understanding of a given field acquired in the course of collaboration with researchers from other fields and 'about-knowledge' which implies some basic familiarity with information about other fields and issues crucial to them. The latter is not knowledge of connections between different fields or disciplines, but connective knowledge "that makes connections possible" [Priaulx, Weinel, 2018, p. 15]. Although the authors do not explicitly point to methodology as a domain of "about-knowledge", it appears that general methodology of science can meet expectations of being such a domain provided it is viewed in a non-traditional way, i.e. neither as a descriptive discipline ("what are the connections between disciplines in integration trials?") nor as a normative one ("what should be the connections between disciplines to make integration trial successful?"). Instead, one could ask "what might the connections between disciplines be?", focusing on the programming role of methodology, even though with "no guarantee that such connections will be made, or that if made, that they will be successful" [Priaulx, Weinel, 2018, p. 15]. 'About-knowledge' enhances then the awareness of potential connections between different fields[4].

[4] In other words, such awareness consists in researchers' attitudes towards their own research specialization and the place of the latter in the whole system of science, as well as attitudes towards other domains of knowledge and of culture [Łojewska, 1976, p. 58].

To sum up our preliminaries, critical analysis of the discussion about the definition, levels and structure of integration of sciences enables us to clarify our account on integration in the following way:

- *integration is about connecting research findings* – it is more than juxtaposition but less than merging;
- *integration is about enriching discipline-specific world-views* – it transcend the narrow discipline-specific world-views but not beyond necessity, being thus less than reconciliation;
- *integration is a trinary relation* – it involves at least two integrated sides and one integrative (connective) played by the general methodology of sciences.

It is now clear enough that some partial conclusions and suggestions that were generated during the discussion about the definition, levels and structure of integration in the second half of the XX century in Poland remain still valid. However, as the aim of this paper is both to engage the contemporary discussion about the nature of relations between different scientific disciplines, as well as to disentangle the concept of integration of sciences from superstructures of rival proposals (unification and interdisciplinarity), this partial conclusions and suggestions need to be refined and supplemented. On the one hand, the refinement consists in introducing and adapting the concept of connective knowledge, as well as localizing it as belonging to the domain of the methodology of humanities. On the other hand, the supplementation arises from the need to update the Poznań Methodological School's account on integration and to tailor it to the context of scientific practices that to large extent are based on the method of modelling. As there is a multiplicity of types of models and their functions in scientific investigations, we argue that integration of sciences may be conducive to a certain type of modelling, namely the exploratory one. Finally, our investigations aims also at meeting some recent postulate raised by the community of philosophers of social sciences that "[i]ntegration is another popular term desperately in need of analysis" [Mäki, 2016, p. 338].

3. Disentanglement of the concept of integration from unification and interdisciplinarity

3.1. Integration / unification

We have already said that in contemporary discussions terms 'integration' and 'unification' are often used interchangeably or the terms or their derivatives are too often put together with an 'or'. Undoubtedly, it is not conducive for the clarity of discussion on problems of integration of sciences. However, one can point to some non-trivial examples of differentiating between these concepts. Some of them are contextual and rather vague. They are in a way a by-product of considerations which are not directly oriented to these topics although they are related to them. For instance, Jaakko Hintikka and Ilpo Halonen by questioning the relation between unification and explanation and by making reference to the example of the special theory of relativity, noted that this theory "was an attempt to integrate the laws of electrodynamics with the laws of mechanics, in the first place with the laws of motion. Einstein's theory is not an explanation of either set of laws; it is a synthesis of the two. (…) Historically speaking, it nevertheless is somewhat dubious to call that integration process unification" [Halonen, Hintikka, 1999, p. 38]. Also in contemporary broad discussions concerning the reduction of psychology to neuroscience, a similar problem is noted: "psychology and neuroscience are and should be connected and perhaps integrated, but not unified" [Schouten, Looren de Jong, 2007, p. 21]. What is more, a growing number of scholars distinguish these concepts explicitly. Instead of using the misleading phrase "unification or integration' we get the following:

- "integration without unification" [Mitchell, Dietrich, 2006];
- "unification beyond integration" [Marquis, Wibler, 2008];
- "unification versus integration" [Miłkowski, 2016].

The most important distinctions that have been pointed out include:

- globality / locality [Bechtel, 1993, pp. 277-278; Wylie, 1999, p. 300; Mitchell, 2003, p. 190; Brigandt, 2010, pp. 306-307];
- non-exclusionarity / exclusionarity [van der Steen, 1993, p. 273; Marquis, Wibler, 2008, p. 351; O'Malley, 2013, p. 559; Breitenbach, Choi, 2017, p. 397];

- simplicity / complexity [Mitchell, 2003, p. 190; Brigandt, 2010, pp. 306-307; Gerson, 2013, p. 517; Miłkowski, 2016, p. 16].

The first option, integration without unification, is first of all related with the criterion of globality / locality. There can be successful integration trials without general theoretical frameworks. In this sense general theoretical frameworks are not required for integration. It is particularly visible in sciences in which fundamental theories are not available. We can then paraphrase integration without unification option in the way that unification is not a *sine qua non* condition of integration, which, as a result, allows us to make integration trials but not unification trials. What is more, as Wim van der Steen noted: „ideal of unification as such is not *sufficient* as a warrant for integrationism" [van der Steen, 1990, p. 34; emphasis in original]. We then get a very informative statement on the relations between unification and integration:

Unification is neither a necessary nor sufficient condition of integration.

The second option, unification beyond integration, is typical for proponents of unification, who are often opponents of integration. One of them is for instance David Trafimov, who discusses the issues which are of interest for us in psychology. He notices that integration tendencies in psychology do occur although they insufficiently take up unification efforts: „integration falls well short of unification. Psychologists should unify but, at best, they integrate" [Trafimov, 2012, p. 702]. Thus, for proponents of unification, integration appear to be excessively eclectic and they consider it results not in a unified whole but in heaps. In order to move towards a holistic approach, one needs to go beyond integration. Therefore, the unification beyond integration option is tightly related to the criterion of exclusionarity / non-exclusionarity, because unification understood in such a way is a "radically nonexclusionary approach" [Marquis, Wibler, 2008 p. 351]. The non-exclusionarity criterion is particularly strongly exposed in contemporary unification accounts, which are intended to be in line with the stance of pluralism in science. The claim that the goal of unification is the pursuit of the final state of unified science is rejected, while unification means a common, continuous cooperation on given problems. This creates room for the cognitive value of pluralism – when pluralism is linked to a joint, collective engagement, thus allowing for overcoming cognitive limitations of particular individuals, it contributes to cognitive progress. A considerable consequence of such an approach is that an important limitation can be established:

> "Unified pluralism embraces a wide range of pluralisms but imposes one important, governing limitation: it rules out ways of proceeding that undermine the continued cooperation and collaboration necessary to make a virtue out of pluralism. Unified pluralism thus *excludes only but all exclusionary projects*" [Breitenbach, Choi 2017, p. 397; emphasis added].

Proponents of unified pluralism consider as an exclusionary approach for instance treating evidence against a given concept as evidence in favor of it. It is characteristic of, for instance, the so-called conspiracy theories. The notion of unification, as it is stated in the above-mentioned extract, *excludes only but all exclusionary projects*. In this respect integration opposed to unification is less inclusive or at least integration can in particular cases exclude more than just exclusionary projects. As Maureen O'Malley put it: "integration does not always mean greater inclusiveness of data, methods or explanation […]. *Integration may involve considerable exclusiveness* to achieve the desired integrative aim" [O'Malley, 2013, p. 559; emphasis added]. Of course, one should not understand it in a way that exclusivity is a necessary condition of integration but rather than:

Non-exclusionarity is not a necessary condition of integration.

The option unification versus integration strongly opposes two convergence tendencies. In this sense, we can treat it as an exclusive disjunction. It is particularly visible when we take into account the simplicity / complexity opposition. In this aspect integration is related to complicating, gradually increasing the complexity of the object of research and it is opposed to simplification and idealization, which are typically linked to unification. Instead of providing simplification, we then recognize connections between various parameters and this way we increase the complexity of the object of research. The following statements in this respect are symptomatic:

> "Integrative models would have to be very complex" [van der Steen, 1990, p. 29];

> "Without (…) a unified theoretical framework (…) one is left with a piecemeal approach to integration. This view recognizes (…) the nonindependence of at least some of the contributions to complex combinations" [Mitchell, 2003, p. 207];
> "partial integration connects specialties in complex and occasionally durable ways without leading to unification of them" [Gerson, 2013, p. 517];
> "greater integration leading away from simplicity toward greater complexity" [Plutynski, 2013, p. 470];
> "the results of integration need not be simple, beautiful, or general, (…) [T]he resulting scientific representation may be highly redundant, violate parsimony considerations, and so forth" [Miłkowski, 2016, pp. 18-19].

Taking into account the above statements, increasing the simplicity reduces redundancy. In this respect, Marcin Miłkowski warns that reducing redundancy always comes at a cost. One should pay attention to the fact that increasing simplicity beyond necessity (as the Occam's razor principle puts it) can lead to undesired consequences:

> "first, it may make the representation more susceptible to error (as redundancy helps error detection); second, it requires more computational effort to handle non-redundant representation. For this reason, models of mechanisms should be as simple and parsimonious *only* as far as it aids their uses" [Miłkowski, 2016, p. 26; emphasis in original].

Unification, contrary to integration, leads to formulating general, simple and global theoretical approaches of the investigated phenomena. Integration, on the other hand, is focused on tackling more local problems, taking into account their complexity – their multi-aspect character and plurality of relations. Unification and integration are therefore guided by different aims. One could thus risk to say that:

Simplicity and integration are inversely correlated with one another.

2.2. Integration / interdisciplinarity

In many discussions which took place in 70-90. of the XXth century interdisciplinarity was strictly associated with integration. It was a common conviction that integration is the necessary condition for the success of interdisciplinary projects. The relation between integration and interdisciplinarity is however questioned. William Bechtel pointed out that if researchers from various disciplines exchange information and share ideas but remain closely attached to their own disciplines, we then have interdisciplinarity but without integration [Bechtel, 1993, p. 295]. What is more, if the exchange does not concern some uncharted territory which can be explored thanks to cooperation and working out new approaches, methods or research instruments, then we have interdisciplinary clusters but not integration of sciences [Bechtel, Hamilton, 2007, p. 405]. Although, as we have signaled in the introduction to this article, while questioning the relation between integration and interdisciplinarity could be viewed as a heresy, it contributes to the clarity of the discussion and makes possible to better conceptualization of the former. One should not infer that the approach already considered as traditional, that is defining interdisciplinarity via integration becomes entirely questioned. It just becomes one of the three options we have in this context:

- "interdisciplinary integration" [Klein, 2010];
- "(successful) interdisciplinarity without integration" [Grüne-Yanoff, 2016];
- "integration without (much) interdisciplinarity" [Brigandt, 2013].

Conceptual oppositions, which serve to explicate the difference between integration and interdisciplinarity are then as follows:

- distal / proximal [Karlqvist, 1999, p. 382];
- loose / strong links [van der Steen, 1990, p. 34].

When it comes to the first option – interdisciplinary integration, it is considered a classical one, because this is how interdisciplinarity is the most often being distinguished from multidisciplinarity. Successful interdisciplinary integrations, which have resulted in the emergence of new scientific disciplines, also fit into this area. Classical example in this respect is biochemistry, which should not be treated as applied organic chemistry but as a discipline with its own domain [Bechtel 1986b, pp. 97-98].

When it comes to the second option – (successful) interdisciplinarity without integration, the relation between integration and interdisciplinarity was firmly questioned by Wim J. van der Steen, who in a way denounced certain interdisciplinary projects as pseudo-integration. He qualified as such the extension of sociobiology to humanities, biological theories of culture and some research from the area of biological psychiatry. The key to denounce the pseudo-integrational character is the excessive use of overgeneral, "diluted," concepts and failing to see the distinctive conceptual character of objects of research in various disciplines. An example of this first flaw are overgeneralizations of concepts from the area of evolutionary biology (for instance, adaptation), which is how they become uninformative. An example of the second pseudo-integration is the failure to notice distinct conceptualizations of culture (biology versus anthropology) or altruism (sociobiology versus ethics). We then have to do with only a superficial and illusory similarity of the objects of research. The conclusions the author draws here are clear:

> „*The use of overgeneral concepts tends to suggest that there is theoretical coherence, within or among disciplines, where none in fact exists*" [van der Steen, 1990, pp. 24-25; emphasis in original];
>
> "Pseudo-integration is common in science" [van der Steen, 1993, p. 272].

While van der Steen focused on denouncing pseudo-integration in interdisciplinarity, Till Grüne-Yanoff presented case studies in which the successful interdisciplinary interaction neither is based nor leads to the integration of disciplines. The conclusion is that the integration is not the necessary condition for successful interdisciplinarity [Grüne-Yanoff, 2016, pp. 358-359]. One of such case studies is the example of evolutionary game theory, which is the effect of interdisciplinary exchange between biology and economics. An important argument here is that it was an authentic interdisciplinary interaction and not only a multidisciplinary juxtaposition. First of all, there was a mutual exchange, "mutual adoption". First, the biologists imported the game theory from economics and then economists re-imported it from biology. What is more, as a result of this exchange there was a double impact of these disciplines on one another. The disciplines were affected by one another and in this sense they both changed their identity:

> "the involved disciplines are substantially affected. (…) the interdisciplinary exchange lead to epistemic success – to more detailed explanations, better control, and higher scientific activity. Crucially, the interdisciplinary exchange was an important causal factor in the production of this success. (…) From either transfer, the importing discipline came out considerably affected" [Grüne-Yanoff, 2016, p. 349].

Although the example of the evolutionary game theory is not a juxtaposition of results of biology and economics, the author claims it is not an example of integration either. Even though we had a situation of mutual exchange and mutual impact, there was no convergence of knowledge. Attempts of integrating these two sciences met with difficulties of ontological and methodological kind. What is interesting, "disintegrating" occurred during the attempts of overcoming these difficulties: "The real changes, instead, arose from attempts to deal with these obstacles. In trying to overcome them, scientists from both discipline worked out discipline-specific concepts and methods, and in that process moved their discipline away from the other" [Grüne-Yanoff, 2016, p. 349]. Therefore, we get interdisciplinarity but without integration. In this respect, stating that interdisciplinarity "presupposes as a minimum that some sort of *inter*-action and *integration* between at least two relevantly different *disciplines* take place" [Hvidtfeldt, 2017, p. 38; emphasis in original] is inadequate because this minimal condition is too narrowly defined. It is the interaction condition that should be treated as a minimal condition and in this sense as a *sine qua non* condition [Lattuca, 2001, p. 14; Holbrook, 2013, p. 1874]. On the other hand, interaction understood as a symmetric relation, interchange or in short just exchange, implies a mutual impact between the disciplines. In this respect we have a "genuine interdisciplinary episode that left *both disciplines considerably transformed*, but not integrated" [Grüne-Yanoff, 2015, p. 708; emphasis added]. From the above, we can draw a following thesis:

Interaction and not integration is a necessary condition of interdisciplinarity.

The above-mentioned considerations should be accompanied by an important comment. Formula "interdisciplinarity without integration" should not be treated in an absolute manner, i.e. that there is no trace of connections whatsoever. It should be understood in a way that there occur interconnections in the non-integration interdisciplinarity but they are loose or weak [van der Steen, 1990, p. 34; Wylie 1999, p. 301]. A good case in point here are sustainability sciences. They are also a good example of interdisciplinary exchange as we have a transfer between the disciplines which are very distant from one another, for instance, oceanography and political sciences. In this respect, the integration is very unlikely. For sustainability sciences the key to interdisciplinarity is the problem-feeding – the transfer of problems and their solutions between different, sometimes pretty distant disciplines. Integration is then unlikely and is not a necessary condition but cooperation and the transfer of problems do occur:

> "the theoretical interconnections that must be in place in order for problem-feeding to ensue can be comparatively weak. *Some* connection needs to be in place, but nothing as substantive as, say, an interfield theory needs to exist. (…) Most philosophical treatments concentrate on disciplinary fields that are in many respects proximate. They share much at the outset, and this makes the sharing and shifting of problems a lot smoother. Within sustainability science this is decidedly not the case –at least, when it comes to integrating the natural and social dimensions of sustainability. However (…) this does not undermine the recognition that problems *need* to be transferred" [Thorén, Persson, 2013, pp. 351-352; emphasis in original].

We could therefore risk the following thesis:

*Interdisciplinarity is impossible without even loose interconnection*s.

We are facing an analogical problem when we consider the third option – the opposite possibility i.e. integration without interdisciplinarity, because to be precise we have a situation in which "such integrative accounts do not involve much interdisciplinarity" [Brigandt, 2013, p. 461]. We then consider neighboring research fields or we stay within the realm of one discipline. It does not mean though that integration remains a trivial task here. It is about integration of knowledge pertaining to different levels of organization or research on different levels of analysis. There occur epistemological and methodological problems specific to integration, which concern relations between levels and, what is particularly important, relations between different levels but concerning the same phenomena.

In the level-of-analysis approach from the field of biological sciences it is assumed that each hypothesis or model comes within one of four distinguished levels: evolutionary, functional, ontogenetic and mechanistic. Lack of differentiation between these levels leads to terminological misunderstandings and unwarranted polemics. Another important issue of levelism is that competition between alternatives occurs only within the realm of particular levels and not between them [Sherman, 1988, pp. 616-617]. Sandra Mitchell, advocate of integrated pluralism, accepts the first element of levelism but rejects the second one. According to the author, accepting the second element of levelism leads to isolationism which is unjustified from the point of view of research practice of biological sciences:

> "While the levels-of-analysis approach correctly recognizes the diversity of questions that can be raised, it fails to acknowledge that the answers at one level may well influence what can be a plausible or probable answer at another. (…) The view of pluralism that I endorse is not "anything goes" or "winner takes all" or "levels of analysis" but rather *integrative pluralism*, which attempts to do justice to the multilevel, multicomponent, evolved character of complex systems. But, one may reasonably ask, what kind of integration?" [Mitchell, 2009, pp. 112-114; emphasis in original].

Answering to the above question we could state that:

- *integration is not isolationist* – interrelatedness between particular levels is assumed; questions formulated on one of the levels cannot be answered in a satisfactory manner without considerations on other levels; it is especially important in a situation when considerations on one of the levels limit the scope of possible answers on the remaining levels;
- *integration is not reductionist* – it does not call for a privileged level, which should be targeted by all the proposed explanations;
- *integration is not formalistic* – there is no a purely formal procedure or algorithm of interlevel integration.

The author of integrative pluralism illustrated her account with examples from biological sciences. However, as she underlined: "[b]oth the ontology and the representation of complex systems recommend adopting a stance of integrative pluralism, *not only in biology, but in general* [Mitchell, 2004, p. 81; emphasis added]. The stance of integrative pluralism is also possible in the social sciences and humanities. Communication studies can be an example, where we have three levels of analysis defined by Claude Shannon: transmissional, semantic and effectiveness. In this respect we can also have isolationist or reductionist approaches. However, an integrative approach is possible in communication studies, where interrelatedness between levels of analysis can be seen [Boruszewski, 2017, pp. 22-24]. Of course, an interlevel integration generates epistemological and methodological problems of sort – above all it requires caution and big awareness of the levels of analysis and interrelatedness between them: „Integrating across levels of analysis is tricky business. (…) Although there is risk of confusion, careful consideration of one level of analysis can benefit at the other" [MacDougall-Shackleton, 2011, p. 2083].

In the light of the above considerations, we could therefore state the following thesis:

Distality as such makes integration more difficult, while proximity as such does not make it easier.

Let us now sum up this section. In the following diagram we illustrate the way in which we conceptualize the relations of integration to unification and interdisciplinarity:

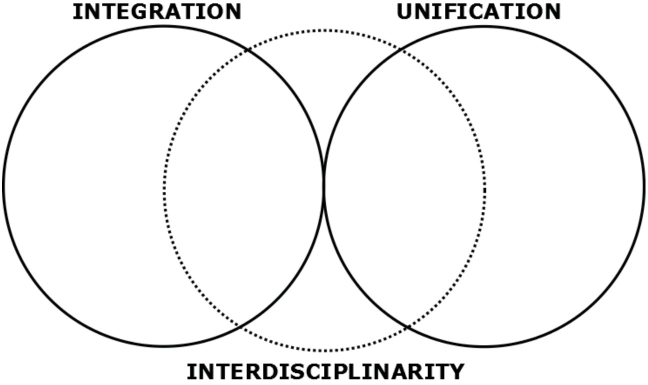


Diagram 1: Disentanglement of integration from unification and interdisciplinarity.

We consider the scopes of concepts of integration and unification as disjunctive although adjacent. In other words, these concepts are mutually exclusive although some transitions ("leaps") between them are possible, namely unification beyond integration and integration without unification. In this sense we treat the overused phrase ‘unification or integration’ as denoting the union of scopes of these two concepts with an empty intersection. Meanwhile, the scopes of the concepts of integration and interdisciplinarity intersect in a non-empty way. What is more important, however, are the differences of scopes of these two concepts and the fact that they also are not empty sets. In this respect, phrases such as ‘interdisciplinary integration’ are not trivial as it was the case in traditional approaches. We propose an analogous solution in the case of unification and interdisciplinarity although as it is not the subject of this article, we leave the question open.

4. A successful integration – a case study

As we have already specified our account of integration of sciences and disentangled it from the superstructures of unification and interdisciplinarity, let us now support it with a case study. We take into consideration a model of magical archaic culture with its central formula in the form of the idea of primitive magic syncretism to show a successful integration trial, which meets the main characteristics of integration presented in our account. This theoretical model may be considered as the effect of an encounter of groups of specialists with different disciplinary background including archeologists (Henryk Mamzer, Andrzej P. Kowalski), ethnologists (Wojciech Burszta, Michał Buchowski) and methodologists of humanities (Jerzy Kmita, Anna Pałubicka). This meeting resulted in connecting various autonomous research findings concerning interpretations of material remnants (archeology), indigenous knowledge of contemporary small-scale traditional societies (ethnology), as well as relicts of magical thinking in speech (linguistics). Establishing this connection was possible only when accompanied by a critical reconceptualization of the empirical basis of historical sciences about culture (epistemic level). In order to go beyond the traditional view that archeological remnants are merely remainings of material culture to be catalogized, the empirical basis was reconceptualized so as to include the idea that such remnants also played a sacral or ritual function which has to be reconstructed. The idea was to reconstruct the place of a magical world-view in the past, but also in various domains of contemporary culture, e.g. art, religion or, as Jerzy Kmita put it, even in science (ideational level). Last but not least, such an integration trial would be hardly possible without the methodology of humanities entering the scheme of integration as the integrator. It was possible, because, as Kmita has recognized, methodology "belongs (…) to the sphere of the historical sciences about (symbolic) culture" [Kmita, 1974, pp. 47-48]. Let us now go to the case study.

The project of a theoretical account of magic was worked out by Anna Pałubicka and then developed in an milieu of researchers with different disciplinary background. The basis of this project is *a model of magical archaic culture* [Burszta, 1991, p. 101; Kowalski, 1997, p. 166; Bińczyk, 2007, p. 102]. The basic problem for us is how to investigate archaic preliterate cultures, when we are faced with unsurmountable epistemological difficulties in their discovering? The complexity of this problem manifests itself already at the basic, conceptual level – how the term 'culture' is understood in prehistorical research. Pałubicka paid attention to the dual understanding of the term. In the first sense we speak about archeological culture, a defined set of material artifacts located in time and space, while in the second understanding the culture takes the form of beliefs shared in a given community and remnants are its symptoms and a base to reconstruct the culture: „the starting point in this process of reconstruction is made by the data from archeological culture, and the point of arrival – aim – data which make culture" [Pałubicka, 1979, p. 60]. Here appears the first cognitive barrier in investigating of magical cultures – it is practically impossible to reconstruct a magical culture solely based on the archeological sources. It is not possible to present the past in a purely physical way and certainly one cannot attribute a cultural status to material objects as such without making a reference to a given model of culture. Also, one cannot limit oneself to describing the purely physical qualities of objects because presumably these qualities were linked to their symbolic function: „classes of objects like Neolithic polished stone implements may have been, in the culture of that age, a metamorphic presentation of sky gods or a sacral force manifesting itself through the lustre, for example" [Kowalski, 2009, p. 36]. Without a reference to a theoretical model of culture objects were under a certain illusion of direct contact with the past – an illusion of cultural transparency of objects.

On the other side of the spectrum, the "purely mental" one, there appears the second cognitive barrier – the impossibility to understand archaic cultures in the sense of antinaturalistic methodology of humanities. More precisely, the way of thinking characteristic of primary societies is uninterpretable or hard to comprehend to contemporary people from the Euro-Atlantic cultural circle. The reality of archaic culture is usually incomprehensible to us:

> "This is why the cultures alien to our culture, that is, the 'primitives' cultures which accept the states of affairs and the 'mystical' relations between them that we take to be unacceptable are unintelligible to us. The states and the relations can be expressed in our concepts, and therefore comprehended by us in this sense. Still, we cannot recognize their objectivity. This is the obstacle that makes our attempt to understand them vain. Although it is possible for us to find out what the states of affairs and the relations between them are present in the 'primitive' cultures, we are not able to understand how one can believe in such things" [Pałubicka, 1998, p. 186].

The model of archaic culture is therefore meant to enable an approximative description of this culture, its conceptual articulation. At its basis, three historical forms of world-view valorization are distinguished, that is ways of combining practical activities with convictions concerning world-views. This way we single out *magical*, *religious* and *modern* valorization [Pałubicka, 1985]. Then the above-mentioned problem is how to characterize communities belonging to magical cultures from the modern culture perspective. Subjects participating in modern culture usually distinguish three spheres of culture: technical-instrumental, symbolic-communicative and world-view. According to the central hypothesis of the model, spheres of consciousness in primeval societies made up a united syncretic whole. For primeval people particular activities were a technical, communicative and world-view act at once. In other words, in primeval minds metonymical (cause-effect, whole-part) and metaphorical (symbolization) relations coexisted. Therefore, the primeval mind does not perform a mutual transformation of these relations on one another because it would mean that it differentiates between them. The differentiation between metonymy and metaphor appeared only later on:

> "The formula: *primitive magic syncretism* – is supposed not only to show that primitive magic is involved, but, in the first place, *that syncretism is stressed* which does

> not mean (as it is generally understood) combining already distinguished elements, but – on the contrary – the combined appearance of elements which only from a later point of view are considered to be different" [Kmita, 1989, p. 157; emphasis added].

The central formula of the model, that is the primitive magic syncretism, could be also developed into a full hypothesis concerning participation in the primitive magical culture: "*Each* participant of the production *is at the same* time the producer, the sender of the message and the advocate of world view (magical) valorization" [Pałubicka, 1985, p. 65; emphasis added]. The simultaneous coexistence of metonymy and metaphor in primitive magic syncretism is sometimes called a *palimpsest* [Leach, 1976, p. 25]. This term is instructive in the sense that it reveals additional difficulties in translating statements taken from magical cultures, for instance 'A stag is a feather', 'We are Araras', into a language understandable for us. Each attempt of translation and interpretation can be a partial paraphrase at most. What is more, if we have many of such partial paraphrases, their sum also won't be an adequate rendition of the unique sense of expressions from magical cultures. A translation of a metonymic-metaphorical palimpsest is in principle impossible because in the magical thinking "there is neither metaphor (…) nor statement about metonymy (…); it consists of both of them at once, and at the same time it has none of these meanings in pure form" [Buchowski, 1996, pp. 307-308]. Therefore, one cannot attribute a purely symbolical status to magical linguistic statements because it would be an imputation on the side of the researcher. Similarly, one cannot attribute a purely instrumental, technical character to material objects (for instance, a piece of pottery or biface) without taking into account their symbolic sense. It is also a bias on the side of the contemporary western culture: "through our own research method, we imply the priority of the technological role of culture over its other, especially symbolic, aspects" [Mamzer, 2009, pp. 96-97].

The model of primitive magic syncretism is the result of integration in a twofold way. The first one concerns the interdisciplinary and integrative character. In developing this model mainly the output of historical sciences, ethnology and archeology is used, with addition of research work in the field of linguistics and religious studies. It does not come as a surprise because the archaic culture, which is the particular and specific subject-matter, is the object of interest of many disciplines. What is much more important here is that methodology of humanities serves as the integrating means (the integrator). Secondly, we have to do with integration of two levels of analysis – the physical level and the mental level. The archeological findings can be described in two ways – as a physical object and as an object used in a given way in a certain community. On the first level we have a description of the portion of the matter occurring in given time and taking a specific space, while in the second description we place a given find in a cultural context adequate to it. The context equips the find with a cultural sense, thanks to which the object described in a physical way becomes a cultural object. In this respect one can distinguish two different account of the empirical basis in the research on preliterate culture. The first of them, the traditional one, means cataloguing – making descriptions of physical objects using terms of their observed properties. In the second one the empirical base is: "interpretatively construed within the perspective of readiness-to-hand. Hence, the tools interpreted by meaning ascription make up the empirical basis for further investigation of culture" [Pałubicka, 2009, pp. 63-64].These descriptions cannot be inferred from one another and cannot be reduced to one another, in particular a mental description is irreducible to a physical description. The integration of two levels of analysis in researching old cultures enables to show the particular interaction between them. As Andrzej P. Kowalski noted, the model of primitive magic syncretism is useful in prehistorical research although not as a basis of classification of archeological matter, because archeology has worked out its own methods in this respect. However, this model makes it possible *to re-interpret and to correct* archeological interpretations, to which the researchers attributed a cultural character. It allows us to explain the cultural dimension of the presence of given objects in a given culture [Kowalski, 1997, p. 166]. In this respect taking into account of the mental description imposes a certain requirement of interpretive criticism on spontaneous acts of ascribing a cultural sense to old tools. This spontaneity is expressed in that the researcher to some extent unintentionally ascribes a sense from their own culture although "since spontaneous recognition of meaning may in many cases be misguided, it is necessary to explore further research techniques like methodical and critical interpretation which aim at meaning-ascription" [Pałubicka, 2009, p. 61]. The other way round, physical descriptions can serve as basis of selecting mental descriptions and thus reduce the speculative character of humanistic considerations. This speculativity is expressed in arbitrarily ascribing subjective intentionality, which can lead to a situation, in which scientific analysis of old cultures is driven out by eseistic visions of the researcher. The physical description therefore represents a criterion of selecting mental descriptions although it does not form the foundation for investigating culture. We therefore have mutual interaction between the levels of description: *correction of spontaneity* (mental-physical relation) and *selection of speculation* (physical-mental relation). In this respect we can point to a significant cognitive value of the integrated model of primitive magic syncretism.

5. Concluding remarks: methodology of science as an integrator in integration trials

In considerations regarding the problems of integration two of its levels are clearly underlined: the practical-institutional one and the epistemic-methodological one. Analysis of the notion of integration show they both play an important role. Van der Steen, when denouncing the pseudo-integration, he paid attention to the fact that even though certain institutional conditions are met (exchange, cooperation, publications), adequate methodological conditions were not met. According to the author, integration "would be possible only if certain institutional resources would be available *and* if certain methodological criteria could be satisfied" [van der Steen, 1993b, p. 349; emphasis in original]. The author expressed a strong disappointment with the then (first half of 1990s) philosophy of sciences in this respect:

> "Philosophers who do not care for *elementary reconstructions of live science* will end up with a very biased view of science" [van der Steen, Sloep, 1993, p. 23; emphasis added];
> "In order to make philosophy of science more practical, one had better start with application of *elementary*, relatively uncontroversial philosophy (especially logic) to problem of science" [van der Steen, 1990, p. 24; emphasis in original];
> "It is unfortunate that conceptual analysis is not very popular in the philosophy of science nowadays" [van der Steen, 1993a, p. 265].

The three quotations mentioned above illustrate this disappointment and their content undoubtedly remains relevant. One can therefore speak of a certain methodological deficit in the analysis of integration of science. It is visible for instance in a specific renunciation of the analysis on the basic, conceptual level or even in setting aside some inconvenient problems:

> "it can also take the form of a denial of the persistent heterogeneity of the conceptual bases of science in progress and, on the whole, this heterogeneity is seldom analyzed. (…) But, for a cross cutting critique to be possible, and for real progress to be made in the integration of scientific knowledge, it is necessary to build analytical frameworks that allow take into account that every research programme claims its own domain of demonstration and its own validation criteria. That is why it seems necessary, at a meta level, to construct analyses which on common bases explicate the hypotheses of heterogeneous research programmes, their criteria of scientificity, the part of arbitrary of each approach, and the limits of empirical validity of their results" [Laurent, 2012, p. 230; emphasis added].

The above comments should not be understood in a way that a need emerges to work out a super-science or a meta-discipline which will automatically deal with the above-mentioned problems. It is also not about constructing a new high-level formal theoretical apparatus. Classically such hopes have been pinned on the General Systems Theory and integration of science remains an evergreen challenge for this theory [Solem, 1993]. However, the success in this respect is at the very least doubtful. Still, it is a perspective of a different kind – the GST can play a certain role in integration but it does not constitute rudimentary knowledge. And this science has to be more implicit and rather focused on know-how because it is to constitute the basis for commitments in integration. Instead of taking up the evergreen challenge for the high-level general theory, it seems much more fruitful to take up the actual challenge by using the good old tools from the methodological toolbox. Having a common specialist knowledge is one thing, but it is about having some common basic knowledge. The latter is treated as a condition of successful integration because it facilitates coordination without explicit communication, common knowledge acts as a kind of infrastructure for the social world that it supports. Common knowledge is built, modified, and used jointly; by its nature, no one can have exclusive use, possession, or control of it. In this respect, common knowledge is analogous to the system of public roads, or utility networks [Gerson, 2013, p. 518]. In this respect it is not about having knowledge of another field or discipline, but knowledge about it. Integration by sharing common knowledge-of is possible only in rare cases when we have specialists from the field of at least two disciplines. In this respect the requirement of the specialistic knowledge-of can prove impractical or excessive because it is important to have the knowledge-about the other integrated part, i.e. connective knowledge, which is:

> „a range of fairly simple facts and information about the sorts of problem domains and approaches that populate different fields and specialisms" [Priaulx, Weinel, 2018, p. 8];
> "a far lower-level and rudimentary knowledge about the kinds of work and approaches that populate a range of academic fields and specialisms" [Ibidem, p. 12];
> "Connective knowledge in this sense is not knowledge *of* connections, but the kind of knowledge that makes connections possible" [Ibidem, p. 15; emphasis in original].

Considerations on connective knowledge suggest that in the scheme of integration a third element should appear, that the relation of integration is in fact a trinary relation. This third element can be called an integrator. In order to explain how much adequate such an approach is, let us recall similar considerations concerning unification, where issues pointing to the existence of a third element appeared. It can be for instance ethics of science understood as unifier of science [Agassi, 1969, p. 470]. A more important proposal in this respect can be the correction in the understanding of unificatory relation of reduction proposed by Adam Grobler. Let us remind that in the classical unificationist approach, reduction is a binary (there is the reduced and the reducing side) and asymmetric relation. The correction is aimed at the following extremely important remark: „problem does not consist in which (present-day or future) science is being reduced to which one, but *by means of which one*" [Grobler, 1982, p. 91; emphasis in original]. In this proposal the relation of reduction is a trinary relation – apart from the reduced and the reducing theories, there is also a third theory, via which the reduction is realized – a reducer. The author of the notion of reducer paid attention to the fact that some branches of mathematics can act as reducers. We can therefore paraphrase a passage from Grobler: the problem does not consist in what to integrate with what but *via* what. We then get the following *scheme of integration*:

***x** is being integrated with **y** via integrator i*

Therefore, the candidate we propose as integrator or integrative discipline is the general methodology of science. This role of methodology of science is different from the ones traditionally associated with it (descriptive methodology, normative methodology). It is about the already mentioned programming role ("how it can be"), which was well expressed by Jerzy Kmita. What is important then is "to identify the kinds of potential connections that *might* be made by combining one's own expert knowledge with other fields" [Priaulx, Weinel, 2018, p. 9; emphasis in original]. In this respect the project of pragmatic methodology by Kazimierz Ajdukiewicz, which was the basis of the Poznań Methodological School, remains valid. Pragmatic methodology deals not so much with the products of scholars' activity but with science understood as profession of researchers, whose three main tasks are:

- analysis of types of activities carried out in research work;
- descriptions of research procedures;
- finding out the goals for which researchers in the various fields strive [Ajdukiewicz, 1974, p. 188].

In the analysis of integration projects, specialized methodology obviously plays the most important role but this does not mean that general methodology does not play any important role. It is visible for instance when we have "cognitive operations which occur in all disciplines, even though *they may play different roles in different sciences*" [Ibidem, p. 186; emphasis added].

Therefore, as we can see, such an understanding of integrator is coherent with the analysis of integration projects, where the problem of methodological deficit has been identified. This problem, however, does not concern insufficient advance when it comes to the pragmatic methodology but its omission. In this respect we have the possibility to "identify the lowest level of cognition about other fields" [Priaulx, Weinel, 2018, p. 12]. This level does not "yet" concern the explanation, which is a much higher level. On the lowest level we have conceptualization problems with *conceptualizing the units of analysis* being the most basic of them [O'Malley et al., 2014, p. 823]. In this respect in particular cases of integration trials one should take into account local diversity of conceptualizations, parametrizations and thought styles. In a sense this task may seem to be elementary and rather unambitious. However, even if it lacks high-level abstractness, it does not mean it cannot take a sophisticated analytical form: "[l]ocal analyses are as large as they are made. If one can find a way to make connections between disparate events, one will have an extended analysis. (…) there is no theoretical limit to the size of the analytical network that one creates" [Stump, 1996, p. 285]. Our account on integration of sciences as a trinary relation, as elaborated in this paper, is an attempt to rethink the function of the general methodology of science and humanities so it can play the role of integrator in integration trials.

Acknowledgments

The research here presented was financed by a grant from the National Programme for the Development of Humanities (grant no. 0266/NPRH4/H2b/83/2016).

Bibliography

Aerts D., Van Belle H., van der Veken J. (eds.) (1999), *World Views and the Problem of Synthesis*, Springer-Science+Business Media, B.V., VUB University Press.

Aerts D., Apostel L., De Moor B., Hellemans S., Maex E., Van Belle H., Van der Veken J. (eds.), (1994), *World Wievs*, VUBPRESS.

Agassi J., (1969), "Unity and Diversity in Science", [in:] *Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968*, eds. R.S. Cohen, M. Wartofsky, Dordrecht, Reidel Publishing Company, pp. 463-522.

Ajdukiewicz K., (1974), *Pragmatic Logic*, Dordrecht, Reidel Publishing Company.

Bechtel W., (1986a), "The Nature of Scientific Integration", [in:] *Integrating Scientific Disciplines*, ed. W. Bechtel, Dordrecht, Martinus Nijhoff Publishers, pp. 3-52.

Bechtel W., (1986b) "Biochemistry: A Cross-Disciplinary Endeavor That Discovered a Distinctive Domain, [in:] *Integrating Scientific Disciplines*, ed. W. Bechtel, Dordrecht, Martinus Nijhoff Publishers, pp. 77-100.

Bechtel W., (1993), “Integrating Sciences by Creating New Disciplines: The Case of Cell Biology”, *Biology and Philosophy* 8, pp. 277-299.

Bechtel W., Hamilton A., (2007), “Reduction, Integration, and the Unity of Science: Natural, Behavioral, and Social Sciences and the Humanities”, [in:] *Handbook of the Philosophy of Science: General Philosophy of Science – Focal Issues*, ed. T. Kuipers, Elsevier, pp. 377-430.

Bińczyk E., (2007), “Language in Archaic, Pre-referential Cultures. The Emergence of Dualism”, [in:] *Cultures. Conflict – Analysis – Dialogue*, eds. C. Kanzian, E. Runggaldier, Heusenstamm: Ontos Verlag, pp. 101-109.

Boruszewski J., (2017), “On Reductionism in Communication Studies”, *Lingua Posnaniensis* 59, pp. 15-25.

Boruszewski J., Nowak-Posadzy K., (2017), “Methodology through a Cultural Lens. The Poznań Approach to Philosophy of Humanities Against Alternative Meta-Methodological Orientations”, *Sensus Historiae* 29, pp. 15-37.

Breitenbach A., Choi Y., (2017), “Pluralism and the Unity of Science”, *The Monist* 100, pp. 391-405.

Brigandt I., (2010), “Beyond Reduction and Pluralism: Toward an Epistemology of Explanatory Integration in Biology”, *Erkenntnis* 75, pp. 295-311.

Brigandt I., (2013), “Integration in Biology: Philosophical Perspectives on the Dynamics of Interdisciplinarity”, *Studies in History and Philosophy of Biological and Biomedical Sciences* 44, pp. 461-465.

Buchowski M., (1996), “Metaphor, metonymy, and cross-cultural translation”, *Semiotica* 110, pp. 301-310.

Burszta W.J., (1991), “Mowa magiczna jako przejaw synkretyzmu kultury” (“Magical speech as a manifestation of a syncretism of culture”), [in:] *Język a kultura,* t. 4, eds. J. Bartmiński, R. Grzegorczykowa, Wrocław, pp. 93-104.

Cobern W.W., (1991), *World View Theory and Science Education Research*, Scientific Literacy and Cultural Studies Project, 44, http://scholarworks.wmich.edu/science_slcsp/44 (access: 27.09.2018).

Czartoryski P., (1967), „Ankieta Wydziału I PAN w sprawie rozwoju nauk społecznych i humanistycznych w Polsce do r. 1985”, Część V („Survey of the Faculty I PAN concerning the development of social sciences and humanities in Poland untill 1985”, Part V), Polska Akademia Nauk, Warszawa.

Driscoll C., (2018), “Cultural Evolution and the Social Sciences: A Case of Unification?”, *Biology & Philosophy* 33, https://doi.org/10.1007/s10539-018-9618-2.

Gerson E.M., (2013), “Integration of Specialties: An Institutional and Organizational View”, *Studies in History and Philosophy of Biological and Biomedical Sciences* 44, pp. 515-524.

Grantham T.A., (2004), “Conceptualizing the (Dis)unity of Science”, *Philosophy of Science* 71, pp. 133-155.

Grobler A., (1982), “Towards New Approach to the Concept of Reduction”, *Reports on Philosophy* 6, pp. 83-97.

Grüne-Yanoff T., (2015), "Models of Temporal Discounting 1937-2000: An Interdisciplinary Exchange between Economics and Psychology", *Science in Context* 28, pp. 675-713.

Grüne-Yanoff T., (2016), "Interdisciplinary Success without Integration", *European Journal for Philosophy of Science* 6, pp. 343-360.

Halonen I., Hintikka J., (1999), "Unification – it's Magnificent but is it Explanation?", *Synthese* 120, pp. 27-47.

Holbrook J.B., (2013), "What is Interdisciplinary Communication? Reflections on the Very Idea of Disciplinary Integration", *Synthese* 190, pp. 1865-1879.

Hvidtfeldt R., (2017), "Interdisciplinarity as Hybrid Modeling", *Journal for General Philosophy of Science* 48, pp. 35-57.

Kaplan D.M, (2017), "Integrating Mind and Brain Science: A Field Guide", [in:] *Integrating Mind and Brain Science*, ed. D.M. Kaplan, Oxford University Press, pp. 1-28.

Karlqvist A., (1999), "Going beyond Disciplines: The Meanings of Interdisciplinarity", *Policy Sciences* 32, pp. 379-383.

Kitcher P., (1981), "Explanatory unification", *Philosophy of Science* 48, pp. 507-531.

Klein J.T., (1990), *Interdisciplinarity: History, Theory, and Practice*, Wayne State University Press, Detroit.

Klein J.T., (2010), "A Taxonomy of Interdisciplinarity", [in:] *The Oxford Handbook of Interdisciplinarity*, eds. J. T. Klein, C. Mitcham, Oxford University Press, pp. 15-30.

Kmita J., (1973a), "O integracyjnej roli marksistowskiej metodologii nauk" („On the integrative role of the marxian methodology of sciences"), *Człowiek i Światopogląd* 10/73, pp. 67-80.

Kmita J., (1973b), "O potrzebie i warunkach integracji nauk" ("On the need and conditions of integration of sciences"), *Nurt* 8/73, pp. 10-11.

Kmita J., (1974), "Methodology of Science as a Theoretical Discipline", *Soviet Studies in Philosophy* 12, pp. 38- 51.

Kmita J., (1975), "W poszukiwaniu modelu integracji nauk (Sprawozdanie z dyskusji)" ("In search of model of integration of sciences – Report from the discussion"), *Studia Filozoficzne* 4/75, pp. 7-18.

Kmita J., (1979), „Światopogląd nauki – światopogląd naukowy" („Science's worldview – Scientific world- view"), [in:] *Nauka i światopogląd*, ed. J. Lipiec, Warszawa, pp. 299-305.

Kmita J., (1988). *Problems in Historical Epistemology*. Dordrecht.

Kmita J., (1989), "The Legacy of Magic in Science", [In:] *Visions of Culture and the Models of Cultural Science*, eds. J. Kmita, K. Zamiara, "Poznań Studies in the Philosophy of the Sciences and the Humanities", Vol. 15,Rodopi, Amsterdam-Atlanta, pp. 153-169.

Kmita J., (1991), *Essays on the Theory of Scientific Cognition*, Dordrecht.

Kmita J., (1996), "Towards Cultural Relativism 'with a small *r*'", [in:] *Epistemology and History. Humanities as Philosophical Problem and Jerzy Kmita's Approach to it*, ed. A. Zeidler-Janiszewska, "Poznań Studies in the Philosophy of the Sciences and the Humanities", Vol. 47, Amsterdam-Atlanta, pp. 541-613.

Kmita J. (2000), *Wymykanie się uniwersaliom* (*Escaping Universals*), Warszawa: Oficyna Naukowa.

Koj L., (1975), "Uwagi o integracji nauk" ("Some remarks on integration of sciences"), *Studia Filozoficzne* 9/75, pp. 37-93.

Kowalski A.P., (1997), "Synkretyzm kultury pierwotnej a interpretacje archeologiczne" ("The syncretism of a primitive culture and archeological interpretations"), [in:] *Kulturowe konteksty idei filozoficznych*, ed. A. Pałubicka, Poznań, pp. 151-167.

Kowalski A.P., (2009), "'Thing' in the Perspective of Anti-dualistic Ontology and the Problem of Archaeological Objects", *Analecta Archaeologica Ressoviensia* 4, pp. 35-43.

Kula W., (1963), *Problemy i metody historii gospodarczej* (*The problems and methods of economic history*), PWN, Warszawa.

Lattuca L.R., (2001), *Creating Interdisciplinarity*, Nashville, Vanderbilt University Press.

Laurent C., (2012), "Plurality of Science and Rational Integration of Knowledge", [in:] *Special Sciences and the Unity of Science*, eds. O. Pombo, J.M. Torres, J. Symons, S. Rahman, Dordrecht, Springer, pp. 219-231.

Lazari-Pawłowska I., (1975), "W poszukiwaniu modelu integracji nauk (Sprawozdanie z dyskusji)" ("In search of model of integration of sciences – Report from the discussion"), *Studia Filozoficzne* 4/75, pp. 32-33.

Lazari-Pawłowska I., (1967), "Ankieta Wydziału I PAN w sprawie rozwoju nauk społecznych i humanistycznych w Polsce do r. 1985", Część I ("Survey of the Faculty I PAN concerning the development of social sciences and humanities in Poland untill 1985", Part I), Polska Akademia Nauk, Warszawa, pp. 32-37.

Leach E., (1976), *Culture & Communication. The Logic by which Symbols are Connected*, Cambridge University Press.

Lele S., Norgaard R.B., (2005), "Practicing Interdisciplinarity", *Bioscience* 55, pp. 967-975.

Lyall C., Bruce A., Tait J., Meagher L., (2011), *Interdisciplinary Research Journeys. Practical Strategies for Capturing Creativity*, New York, Bloomsbury Academic, London.

Łojewska M.I., (1976), "Integracja nauki a naukoznawstwo" ("Integration of sciences and the science of science"), *Studia Filozoficzne* 2/76, pp. 57-67.

Łojewska M.I., (1986), *Filozofia nauki (Philosophy of science)*, Warszawa.

Łubnicki N., (1967), „Ankieta Wydziału I PAN w sprawie rozwoju nauk społecznych i humanistycznych w Polsce do r. 1985", Część I („Survey of the Faculty I PAN concerning the development of social sciences and humanities in Poland untill 1985", Part I), Polska Akademia Nauk, Warszawa, pp. 17-20.

MacDougall-Shackleton S.A., (2011), "The Levels of Analysis Revisited", *Philosophical Transactions of The Royal Society B* 366, pp. 2076-2085.

Maisel W., (1973), "Problematyka integracyjna dyscyplin prawno-historycznych" ("The problems of interdisciplinary relations between history and law"), *Studia Metodologiczne* 10, pp. 79-93.

Mamzer H., (2009), "Archaeological Sources: Artifacts or Linguistic Statements Concerning Them?", *Analecta Archaeologica Ressoviensia* 4, pp. 77-99.

Marquis A., Wilber K., (2008), "Unification Beyond Eclecticism and Integration: Integral Psychotherapy", *Journal of Psychotherapy Integration* 18, pp. 350-358.

Mäki U., (2016), "Philosophy of interdisciplinarity. What? Why? How?", *European Journal for Philosophy of Science* 6, pp. 327-342.

Miller R.C., (1982), "Varieties of Interdisciplinary Approaches in the Social Sciences", *Issues in Integrative Studies* 1, pp. 1-37.

Miłkowski M., (2016), "Unification Strategies in Cognitive Science", *Studies in Logic, Grammar and Rhetoric* 48, pp. 13-33.

Mitchell S.D., (2003), *Biological Complexity and Integrative Pluralism*, Cambridge University Press.

Mitchell S.D., (2004), "Why Integrative Pluralism?", *E:CO* 6, pp. 81-91.

Mitchell S.D., (2009), *Unsimple Truths. Science, Complexity, and Policy.* The University of Chicago Press.

Mitchell S.D., Dietrich M., (2006), "Integration without Unification: An Argument for Pluralism in the Biological Sciences", *The American Naturalist* 168, pp. S73-S79.

Mitchell S.D., Gigerenzer G., Daston L., Sesardic N., Sloep P.B, (1997), "The Whys and Hows of Interdisciplinarity", [in:] *Human by Nature: Between Biology and the Social Sciences*, eds. P. Weingart, S.D. Mitchell, P.J. Richerson, S. Maasen, Mahwah, Lawrence Erlbaum Associates, pp. 103-150.

Mormann T., (2018), "Scientific worldviews as promises of science and problems of philosophy of science", *Centaurus. An International Journal of the History of Science and its Cultural Aspects* 59, pp. 189-203.

Newell W.H., Green W.J., (1982), "Defining and Teaching Interdisciplinary Studies", *Improving College and University Teaching* 30, pp. 23-30.

Nowak, L. (1980). *The Structure of Idealization*, Dordrecht.

O'Malley M.A., (2013), "When Integration Fails: Prokaryote Phylogeny and the Tree of Life", *Studies in History and Philosophy of Biological and Biomedical Sciences* 44, pp. 551-562.

O'Malley M.A., Brigandt I., Love A.C., Crawford J.W., Gilbert J.A., Knight R., Mitchell S.D., Rohwer F., (2014), "Multilevel Research Strategies and Biological Systems", *Philosophy of Science* 81, pp. 811-828.

O'Malley M.A., Soyer O.S., (2012), "The Roles of Integration in Molecular Systems Biology", *Studies in History and Philosophy of Biological and Biomedical Sciences* 43, pp. 58-68.

Pałubicka A., (1979), „The positivist and instrumentalist concept of the so-called archeological cultures", *Poznan Studies in the Philosophy of the Sciences and the Humanities* 5, pp. 55-66.

Pałubicka A., (1985), "O trzech historycznych odmianach waloryzacji światopoglądowej" ("On three historical forms of valorization concerning point of view"), *Studia Metodologiczne* 24, pp. 51-76.

Pałubicka A., (1998), "Frazer's and Levy-Bruhl's Conceptions of Magic and the Problem of Cognition in the Humanities", [in:] *Theory and Practice of Archaeological Research*, Vol. III, ed. S. Tabaczyński, Warszawa, pp. 177-189.

Pałubicka A., (2009), "Tool and Thing in Archeological Interpretation", *Analecta Archaeologica Ressoviensia* 4, pp. 51-65.

Petkov S., (2015), "Explanatory unification and conceptualization", *Synthese* 192, pp. 3695-3717.

Plutynski A., (2013), "Cancer and the Goals of Integration", *Studies in History and Philosophy of Biological and Biomedical Sciences* 44, pp. 466-476.

Priaulx N., Weinel M., (2018), "Connective knowledge: what we need to know *about* other fields to 'envision' cross-disciplinary collaboration", *European Journal of Future Research* 6, https://doi.org/10.1186/s40309-018-0150-z.

Richards D.G., (1996), "The Meaning and Relevance of 'Synthesis' in Interdisciplinary Studies", *The Journal of General Education* 45, pp. 114-128.

Rouse J., (2015), *Articulating the World. Conceptual Understanding and the Scientific Image*, The University of Chicago Press Chicago and London.

Schouten M., Looren de Jong H., (2007), "Mind Matters: The Roots of Reductionism", [in:] *The Matter of the Mind. Philosophical Essays on Psychology, Neuroscience, and Reduction*, eds. M. Schouten, H. Looren de Jong, Blackwell Publishing, pp. 1-28.

Sherman P.W., (1988), "The Levels of Analysis", *Animal Behaviour* 36, pp. 616-619.

Solem O., (1993), "Integrating Scientific Disciplines: An Evergreen Challenge to Systems Science", [in:] *Systems Science. Addressing Global Issues*, eds. F.A. Stowell, D. West, J.G. Howell, Springer, pp. 593-598.

Stump D.J., (1996), "From Epistemology and Metaphysics to Concrete Connections", [in:] *The Disunity of Science: Boundaries, Context, and Power*, eds. P. Galison, D. Stump, Stanford University Press, pp. 255-286.

Szell G., Shujiro Y., (1993), "The Re-integration of Social Sciences: Methodological and Epistemological Foundations of Integrated Social Sciences", *Hitotsubashi Journal of Social Studies* 25, pp. 103-114.

Thorén H., Persson J., (2013), "The Philosophy of Interdisciplinarity: Sustainability Science and Problem-Feeding", *Journal for General Philosophy of Science* 44, pp. 337-355.

Topolski J., (1964), "Historia gospodarcza a teoria ekonomii" ("Economic history and theory of economics"), *Kwartalnik Historyczny* 1, pp. 85-89.

Topolski J., (1965), "Integracyjny sens materializmu historycznego" ("The integrative sense of historical materialism"), *Studia Metodologiczne* 1, pp. 5-21.

Topolski J., (1976), *Methodology of History,* Dodrecht.

Topolski J., (1985), "On the Concept and Structure of Methodological Consciousness of Historians", [in:] *Consciousness: Psychological and Methodological Approaches,* "Poznan Studies in the Philosophy of the Sciences and the Humanities" Vol. 8, ed. J. Brzeziński, pp. 148-156.

Topolski J., (2009), "The Model and Its Concretization in Economic History", [in:]Idealization XIII: Modeling in history, "Poznan Studies in the Philosophy of the Sciences and the Humanities" Vol. 97 ed. K. Brzechczyn, pp. 159-172.

Trafimov D., (2012), "The Role of Mechanisms, Integration, and Unification in Science and Psychology", *Theory & Psychology* 22, pp. 697-703.

van der Steen W.J., (1990), "Interdisciplinary Integration in Biology? An Overview", *Acta Biotheoretica* 38, pp. 23-36.

van der Steen W.J., (1993a), "Towards Disciplinary Disintegration in Biology", *Biology and Philosophy* 8, pp. 259-275.

van der Steen W.J., (1993b), "Additional Notes on Integration", *Biology and Philosophy* 8, pp. 349-352.

van der Steen W.J., Sloep P., (1993), "Philosophy, Education and the Explosion of Knowledge", *Interchange* 24, pp. 19-28.

Waters C.K., (2017), "No General Structure", [in:] *Metaphysics and the Philosophy of Science, New Essays*, eds. M.H. Slater, Z. Yudell, Oxford University Press, pp. 81-107.

Wylie A., (1999), "Rethinking Unity as a 'Working Hypothesis' for Philosophy of Science: How Archaeologists Exploit the Disunities of Science", *Perspectives on Science* 7, pp. 293-317.

Ziembiński Z., (1966), "O niektórych przyczynach dezintegracji nauk prawniczych" ("On some reasons of disintegration of legal sciences"), *Studia Metodologiczne* 2, pp. 3-17.

Jarosław Boruszewski
Philosophical Faculty
Adam Mickiewicz University
Szamarzewskiego St. 89C, 60-569 Poznań
e-mail: borjar@amu.edu.pl

Krzysztof Nowak-Posadzy
Philosophical Faculty
Adam Mickiewicz University
Szamarzewskiego St. 89C, 60-569 Poznań
e-mail: k_nowak@amu.edu.pl