Zagadnienia Filozoficzne w Nauce

Philosophical Problems in Science

Cover design: Mariusz Banachowicz

Adjustment and correction: Artur Figarski

Proofreading: dr Roman Krzanowski

Technical editor: Artur Figarski

Typographic design: Piotr Urbańczyk

Typeset in LaTeX

ISSN 0867-8286 (print format)

e-ISSN 2451-0602 (electronic format)

Editorial Office
Zagadnienia Filozoficzne w Nauce
Wydział Filozoficzny UPJPII
ul. Kanonicza 9, 31-002 Kraków
POLAND

e-mail: zagadnienia@upjp2.edu.pl

www.zfn.edu.pl

Copernicus Center PRESS

Publisher: Copernicus Center Press Sp. z o.o.
pl. Szczepański 8, 31-011 Kraków POLAND
tel. (+48) 12 448 14 12
e-mail: marketing@ccpress.pl
www.ccpress.pl

Zagadnienia Filozoficzne w Nauce

Philosophical Problems in Science

LXIX ■ 2020

Artykuły

Articles

Artykuły

Articles

The homunculus brain and categorical logic

Steve Awodey
Departments of Philosophy and Mathematics, Carnegie Mellon University

Michael Heller
Copernicus Center for Interdisciplinary Studies, Jagiellonian University

Abstract

The interaction between syntax (formal language) and its semantics (meanings of language) is one which has been well studied in categorical logic. The results of this particular study are employed to understand how the brain is able to create meanings. To emphasize the toy character of the proposed model, we prefer to speak of the homunculus brain rather than the brain *per se*. The homunculus brain consists of neurons, each of which is modeled by a category, and axons between neurons, which are modeled by functors between the corresponding neuron-categories. Each neuron (category) has its own program enabling its working, i.e. a theory of this neuron. In analogy to what is known from categorical logic, we postulate the existence of a pair of adjoint functors, called Lang and Syn, from a category, now called BRAIN, of categories, to a category, now called MIND, of theories. Our homunculus is a kind of "mathematical robot", the neuronal architecture of which is not important. Its only aim is to provide us with the opportunity to study how such a simple brain-like structure could "create meanings" and perform abstraction operations out of its purely syntactic program. The pair of adjoint functors Lang and Syn model the mutual dependencies between the syntactical structure of a given theory of MIND and the internal logic of its seman-

Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)
No 69 (2020), pp. 7–34

tics given by a category of BRAIN. In this way, a formal language (syntax) and its meanings (semantics) are interwoven with each other in a manner corresponding to the adjointness of the functors Lang and Syn. Higher cognitive functions of abstraction and realization of concepts are also modelled by a corresponding pair of adjoint functors. The categories BRAIN and MIND interact with each other with their entire structures and, at the same time, these very structures are shaped by this interaction.

Keywords

categorical logic, syntax-semantics, mind-brain

1. Introduction: On the Computer Screen

We were preparing a paper for publication. A phase portrait was nicely displayed on the computer screen. The network of trajectories represented a class of solutions to the equation we were interested in. At some points, called critical points, certain trajectories crossed each other. These points were important for our analysis. Some of the diagrams we worked with appeared later as figures in our publication (Woszczyna and Heller, 1990). The figures had to be explained, so we decided to attach appropriate labels to some of the critical points. We attached the label "stable saddle" to one of them. No problem. Then we proceeded to attach the label "unstable saddle" to another one. But the label jumped up. We tried to fix it up, but it jumped down. Then we started laughing. After all, it is an unstable point!

Let us try to understand the situation. We were investigating an equation that (virtually) contains in itself its space of solutions (irrespectively of whether we explicitly know them or not). Through the suitable computer program and some "electronic circuits", which are activated by the program, this space of solutions is mapped into the phase portrait displayed on the computer screen. The diagram we see on the screen is certainly something more than just a picture. It does not simply show stable and unstable critical points; it also does what the abstract equation orders its solutions to do (labels jump up and down at instabilities).

Let us go a step forward. In fact, the phase portrait on the screen is a substitute of the world. For suppose that our equation "describes" (or better – models) a mechanical system (e.g., a pendulum or oscillator).[1] Then the unstable critical points of our equation correspond to physical situations in which the considered mechanical system behaves in an unstable way. We thus have, on the one hand, an equation (or a set of equations) or, more broadly, a mathematical theory and, on the other hand, a domain (or an aspect) of the physical world of which the considered mathematical theory is a model. Between the mathematical model and the domain (or aspect) of the physical world there is a mysterious correspondence – a correspondence in the root-meaning of this word: both sides co-respond to each other. It is an active correspondence, and the activity goes both ways: it looks as if the domain of the world informed the theory about its own internal structure, and the theory answered by prescribing what the domain should do. And the domain does it. The equations prescribe what the world should do, and the world executes this. The equations and the world are coupled with each other and act in unison.

[1] The equations we considered in our publication referred to a cosmological situation.

And the screen on my computer? It is a part of the world. The program we have constructed reads the structure of the equations and executes what the equations tell it. And because of the coupling between the equations and the world, the computer does, in miniature, what does the world on its own scale. This is the reason why computers are so effective in our reading of the structure of the world.

There is another domain in which a formal structure reveals its effective power and produces real effects. Such processes occur in the brain. The formal structure in question consists of electric signals propagating along nerve fibres between neurons across synapses, and the world of meanings should be regarded as a product of this activity. The interaction seems to go both ways: the "language of neurons" (what happens in the brain) produces the meanings related to this language (in the mind), and the meanings somehow influence the architecture of neurons.

It seems that in both these cases (mathematical laws and their effects in the real world, and the brain–mind interactions) we meet two instances of the same working of logic where syntax (a formal structure), by effectively interacting with its semantics, produces real effects. This kind of interaction, although kept strictly on the level of logic (i.e. with no reference to processes in the real world), is well known in the categorical logic. In the present paper, we attempt to employ these achievements of categorical logic to try to understand the brain–mind interaction.

The traditional terminology of brain and mind (irrespective of current trends in cognitive sciences to get rid of their conceptual load) seems especially well adapted to the present context in which general ideas are more important than structural details. Moreover, to avoid too hasty associations with the human brain and to emphasize the toy character of the proposed model, we prefer to speak of the "homunculus brain" rather than just of brain.

The action of our argument develops along the following lines. Section 2 is a reminder on formal language, its syntax and semantics. Sections 3 and 4 briefly review those parts of categorical logic that refer to these concepts. Every category, call it C, has its internal logic, and if this logic is sufficiently rich, the category provides semantics for a certain formal theory T. Moreover, there exists a pair of adjoint functors, called Lang and Syn, from a category, called CATEGORIES, of categories belonging to a certain class (for instance, coherent categories) to a category, called THEORIES, of theories and *vice versa*, which describe mutual dependencies between the syntactical structure of T and the internal logic of its semantics given by C. This is described in section 3. In this way, syntax and semantics are interwoven with each other in a manner corresponding to the adjointness of the functors Lang and Syn. This is explored in section 4. In section 5, we consider a deep categorical duality between the syntactic category of a theory and its individual models and suggest a functional interpretation in terms of abstraction and realization of concepts, in anticipation of the cognitive interpretation to be introduced next. In section 6, the category CATEGORIES becomes the category BRAIN. It constitutes a simple model of a homunculus' brain. Objects of this category are categories (belonging to a certain class); every such category models a neuron. Morphisms of this category model signals propagating along nerve fibres between neurons. The category THEORIES becomes the category MIND. Its objects are "theories of neurons"; more precisely, if $C \in$ BRAIN, then its "theory" is $\mathrm{Lang}(C)$ in MIND. Morphisms of this category are functors between the corresponding syntactic theories; more precisely, if $T_1, T_2 \in$ MIND, then the morphism between them is $\mathrm{Syn}(T_1) \to \mathrm{Syn}(T_2)$. The pair of adjoint functors Lang and Syn model the interaction between the syntax of "theories" and their semantics, i.e. the network of neurons. The categories BRAIN and MIND are indeed somehow related to what their names refer to, at least as far as homunculus' brain and mind are concerned.

Following the seminal paper of McCulloch and Pitts, published as early as in 1943, which proposed using classical logic to model neural processes in the brain, there have been so many papers developing and modifying (with various logical systems) this idea, that to quote even a sample of them would be immaterial (for a relatively recent state of art see a short review Koch, 1997). A. Ehresmann claims that it was R. Rosen who was the first to employ category theory to model biological systems. A series of works followed (a non-representative sample: (Gómez and Sanz, 2009; Healy and Caudell, 2006; Mizraji and Lin, 2011; Tsuchiya, Taguchi and Saigo, 2016)) proposing the use of various parts of category theory to model different aspects of the brain activity. In particular, adjoint functors were suggested to model "a range of universal-selectionist mechanisms" (Ellerman, 2015). However, we have not been able to find anything similar to modeling the interaction between brain's language and its meaning anywhere.

2. Syntax and Semantics

In linguistics, syntax and semantics are regarded as parts of semiotics, the study of signs. Syntax studies relations between signs, and semantics relations between signs and what the signs refer to.[2] Syntactic properties are attributed to linguistic expressions entirely with respect to their shape (or form). Semantics, on the other hand, endows them with meaning by referring signs to what they signify. Logic adapts these ideas to its own needs. Since it is a formal science, the signs it considers should be elements of a formal language, and they cannot refer to anything external. Halvorson put it, "But a formal language is really not a language at all, since nobody reads or writes in a formal language. Indeed, one of the primary features of these so-called formal languages is that the symbols don't have any meaning" (Halvorson, 2016). This is why the meaning should be "artificially" constructed for them. The idea of how this should be done can best be seen in Tarski's prototype of this procedure (Tarski, 1933). If a sentence s, the truth of which we want to define, belongs to a language L then the definition of s should be formulated in a metalanguage M with respect to the language L. And the metalanguage M should contain a copy of s so that anything one can say with the help of s in L, can also be said in M. The definition of "True" should be of the form

$$\text{For all } x, \text{True}(x) \text{ if and only if } \varphi(x)$$

with the condition that "True" does not occur in φ. Here x stands for the copy of the sentence s in the metalanguage L, and $\varphi(x)$ describes, also in M, the state of affairs of which the sentence s in L reports (for more details see Hodges, 2018; Sher, 1999). A metalinguistic copy of s could also be expressed as "s" (taken in quotes). In Tarski's own example:

---

[2] Sometimes one also distinguishes pragmatics which studies relations between signs and their users.

"It snows" is true iff it snows.

For pedagogical reasons, this example is taken from colloquial language, but strictly speaking Tarski's definition refers to formal languages. The formal language L has its own syntax (since it is a formal language), but is lacking its semantic reference. As we have seen, such a reference had to be constructed for it with the help of the metalanguage M.

Now, the idea is to improve the situation by looking for such a conceptual context in which a semantics for a given theory would arise in a more natural (or even spontaneous) way.

3. Categorical Semantics

To do so we must first define precisely what we mean by language. Since the definition must be precise, let us choose as an example the language of mathematics based on standard first order logic (which is enough for most of the usual mathematics). Many other languages may be formalized in a similar way. In such a language we distinguish:

- constants: $0, 1, 2, \dots, a, b, c, \dots,$ and variables: $x, y, z, \dots,$ which can be combined by primitive operations to give
- terms, for example: $x + y$, $x^3, \dots$ which, in turn, can be combined, with the help of primitive relations, such as $=, <, \leq, \dots,$ to produce
- formulae, for example: $x+y = z$, $x \leq y, \dots$ which, in turn can be combined, with the help of the usual logical connectives and quantifiers, into
- more complicated formulae.

To make the language more flexible and more adapted for concrete applications, we diversify its expressions into various types (called also sorts). In mathematics, we might use different letters for natural and real numbers, or different symbols for vectors an scalars. We say that, in both cases, we are using a two-typed language. There may be languages with as many types as is needed.

What we need is not so much a language, but rather a theory. In mathematical logic theory is almost the same as language; it is a formal language aimed at axiomatizing a certain class of sentences. The concept of theory, as it is functioning in modern physics can, in principle, be regarded as the special case of the logical concept of theory, although in scientific practice theories are rarely formulated with the full logical rigor.

Let then T be a theory expressed in a multi-type language. Such a theory is defined as consisting of the following data:

1. A set of types $\{X_1, X_2, \ldots, X, Y, \ldots\}$.
2. A set of variables $\{x, y, z, \ldots, x_1, x_2, x_3, \ldots\}$ with a type assigned to each variable.
3. A set of function symbols with a type assigned to each domain and codomain of every function symbol; for instance, to the term $x_1 + x_2$, with the variable x_1 of type X_1 and the variable x_2 of type X_2, there corresponds the function symbol $f : X_1 \times X_2 \to Y$, and the term $f(x_1, x_2) = x_1 + x_2$ is of type Y.
4. A set of relation symbols with a type assigned to each argument of every relation symbol; for instance, to the formula $x + y = z$, with the variable x of type X_1, the variable y of type X_2 and the variable z of type X_3, there corresponds the relation symbol $R \subseteq X_1 \times X_2 \times X_3$, and $R(x, y, z)$ is an atomic formula.
5. A set of logical symbols.

6. A set of axioms for a given theory built up from terms and relation symbols with the help of logical connectives and quantifiers, respecting types of all terms.

This is, in fact, a purely syntactic definition of theory (for details see Borceux, 1994, pp.344–348; Mac Lane and Moerdijk, 1992, pp.527–530). Now, we want to create a semantics, i.e. a model, for a theory T. This is done by constructing a category C_T which will serve us as such a model. The construction is almost obvious:

1. each type of T is an object of C_T,
2. for each function symbol f in T with types A and B as its domain and codomain, correspondingly, f is a morphism from the object A to the object B in C_T,[3]
3. variables are identity morphisms in C_T,
4. for each relation symbol R in T, its counterpart in C_T is a subobject in C_T. Suppose ϕ is a subobject of an object A in C_T then, by analogy with the usual theory of sets, ϕ can be thought of as a collection of all things of type A that verify ϕ.

This definition must be supplemented with all of the (first order) logic which is used to express axioms in T (for details see nLab, 2017). Roughly speaking, since formulae correspond to subobjects, and all subobjects of a given object are partially ordered by inclusions (they form a poset), the axioms can be expressed in terms of the order relation on the subobject poset in the category C_T. The category, defined in this way, is appropriately called the categorical semantics for a theory T.

[3] Since f is now regarded as being in C_T rather than in T, it should formally be denoted by a different symbol such as $[f]$, but we omit such formalities for present purposes.

We have thus created (almost automatically!) a domain (the category C_T) the theory T refers to. The internal architecture of the category C_T exactly matches the logic involved in the theory T.

Let us also mention that, *vice versa*, having a (sufficiently rich) category C', we can construct the formal theory T' the logic of which matches the internal architecture of the category C'. This can be done by reading the above definition of the categorical semantics "backwards", i.e. we regard objects of C' as types of T', identity morphisms of C' as variables in T', etc. The theory T', reconstructed in this way from the category C', is called internal logic of C. This entire process can be regarded as a functor, called Lang, from a category of categories, call it CATEGORIES, to a category of theories, call it THEORIES,

$$\text{Lang: CATEGORIES} \to \text{THEORIES.}$$

For the time being this definition remains informal since neither CATEGORIES nor THEORIES have been properly defined, but it will be done below.

Let us start with a formal theory T. We now want to organize it into a category Syn(T), called the syntactic category of T. It is done in the following way.

Let Γ be a collection of type assertions, i.e. a collection of rules assigning a type to each term of a given theory, and Φ a collection of all well-defined formulae of T. The pair (Γ, Φ) is called a context. It is a formalization of what in ordinary language one means by this term.

If T is a type theory, its syntactic category, Syn(T), is defined as follows. Its objects are contexts (Γ, Φ) and its morphisms $(\Gamma, \Phi) \to (\Delta, \Psi)$ are interpretations (or substitutions) of variables. The latter means that for each type, prescribed by Δ, we must construct an expression of this type out of data contained in Γ. In general, this is done by substituting terms from Γ for variables in Δ. We must also present, for each assumption required by Δ (if there are any), a proof of this assumption from the assumptions contained in Γ (for details see Fu, 2019; nLab, 2020c).

The category Syn(T), constructed in this way, is also called a category of contexts (for details see Fu, 2019; nLab, 2020b)).

Since from a theory T we have constructed the category Syn(T), we can have a functor,

$$\text{Syn: THEORIES} \to \text{CATEGORIES}$$

provided we define the categories THEORIES and CATEGORIES. We do this in the next section.

4. Syntax–Semantics Interaction

Let us start with objects for both of these categories. It is obvious that they will be categories and theories, respectively. To have workable categories, one must restrict the class of theories as candidates of being objects in THEORIES (and analogously for CATEGORIES). The criterion one follows is the kind of logic that underlines a given theory. It could be what logicians call: finite product logic, regular logic, coherent logic, geometric logic, etc.[4] For our further analysis it is irrelevant which one will be chosen. However, for the sake of concreteness we may think about coherent logic. Roughly speaking, this is a fragment of the first order logic which uses only the con-

[4] As it could be expected, the internal logic of the corresponding semantic category will be of the corresponding kind, i.e. finite product logic, regular logic, etc. (nLab, 2017).

nectives $\wedge$ and $\vee$, and the existential quantifier. Large parts of mathematics can be formalised with the help of this logic. To this logic there correspond coherent theories and coherent categories. They will constitute objects of THEORIES and CATEGORIES, respectively. Morphisms for CATEGORIES are obviously functors between corresponding categories; for instance coherent functors for coherent categories (nLab, 2011). Let now T_1 and T_2 be objects in THEORIES. A morphism $T_1 \to T_2$ is a functor between their corresponding syntactic theories $\mathrm{Syn}(T_1) \to \mathrm{Syn}(T_2)$. Roughly speaking, this means that it is possible to express (to interpret) T_1 in terms of T_2 (for details and discussion see Halvorson and Tsementzis, 2017).[5]

As a side remark let us notice that by studying the category THEORIES, we could learn "how individual theories sit within it, and how theories are related to each other" (Halvorson and Tsementzis, 2017, p.413). This is nicely consonant with a newer trend in the philosophy of science to investigate the so-called inter-theory relations (Batterman, 2016; Rosaler, 2018).

A truly remarkable fact is that the functors Lang and Syn constitute a pair of adjoint functors. Let us explain precisely what this means.

Let us consider any pair of objects: $\mathcal{C}$ of CATEGORIES and T of THEORIES. Adjoint functors serve to compare them. However, they cannot be compared directly since they live in different categories. Adjoint functors serve to move each of them to the correct category so as to enable the comparison. Let us follow this process step by step (Simmons, 2011, pp.148-153).

[5] Strictly speaking, CATEGORIES is a 2-category (since its objects are categories and morphisms are functors), and THEORIES is a 2-category, in this case, called also a doctrine (nLab, 2020a).

Let us first consider the object T which lives in THEORIES. We want to compare it with the object $\mathcal{C}$ which lives in CATEGORIES. We thus move $\mathcal{C}$ to THEORIES with the help of the functor Lang to obtain the object Lang($\mathcal{C}$). We now make the comparison with the help of a suitable morphism,

$$f : \mathrm{Lang}(\mathcal{C}) \to T$$

in THEORIES. We do the same starting with $\mathcal{C}$ in CATEGORIES and T in THEORIES, and compare $\mathcal{C}$ with $\mathrm{Syn}(T)$,

$$g : \mathcal{C} \to \mathrm{Syn}(T)$$

in CATEGORIES. To complete the definition of adjunction we demand that morphisms f and g should constitute a pair of bijections which is natural both in $\mathcal{C}$ and T (see below).

The above definition can be put into a concise form

(1) $$\mathrm{THEORIES}(\mathrm{Lang}(\mathcal{C}), T) \cong \mathrm{CATEGORIES}(\mathcal{C}, \mathrm{Syn}(T)),$$

expressing an isomorphism between the right and left hand sides of this formula that is natural in $\mathcal{C}$ and T. The latter condition says that when $\mathcal{C}$ varies in CATEGORIES and T varies in THEORIES, the isomorphism between morphisms $\mathrm{Lang}(\mathcal{C}) \to T$ in THEORIES and $\mathcal{C} \to \mathrm{Syn}(T)$ in CATEGORIES vary in a way that is compatible with the composition of morphisms in CATEGORIES and THEORIES, correspondingly, and with the actions of Lang and Syn on both these categories (see Awodey and Forssell, 2013; Leinster, 2014, pp.50-51).[6]

We should notice that in the above definition, in fact, we not only compare objects of two different categories, but rather categories

[6] For a full definition of adjoint functors see any textbook on category theory.

themselves (objects $\mathcal{C}$ and T are any pair of objects). Moreover, comparing two categories we are not so much interested in their objects, but rather in morphisms between objects. This is clear from the fact that at the end, we have identified those morphisms of two categories that are pairwise naturally isomorphic among themselves.

As we can see, categorical logic does not simply create a semantics for a given language, but shows that dependencies between them go both ways: in a sense, syntax and semantics create each other. More precisely, they condition each other through the adjointness relation.

5. Realization and Abstraction

There is another aspect of categorical logic that we shall make use of, and it may be seen as a mathematical description of the processes of abstraction and realization of concepts. The category $\mathrm{Syn}(T)$ representing a theory T may be regarded as presenting a general *concept*, of which the theory T is a particular syntactic description. For example, there is a theory T_{Group} consisting of a single basic type X, and function symbols $* : X \times X \to X$ and $(-)^{-1} : X \to X$ and a constant $u : X$, together with the usual equations for groups as its axioms:

$$
\begin{aligned}
x * (y * z) &= (x * y) * z \\
x * u &= x \\
u * x &= x \\
x * x^{-1} &= u \\
x^{-1} * x &= u
\end{aligned}
$$

The syntactic category $\mathrm{Syn}(T_{Group})$ then represents the general concept of a group. This concept can also be represented by another theory T'_{Group} with a different choice of basic equations, or even a different choice of operations,[7] as long as the resulting categories $\mathrm{Syn}(T_{Group})$ and $\mathrm{Syn}(T'_{Group})$ are equivalent.

A general concept may have many individual *instances*; an instance of the concept of a group is, of course, just a particular group: a set G of elements, equipped with functions interpreting the operations of multiplication and inverse, and satisfying the group equations. A logician would call such an instance a *model of the theory of groups*, but we shall avoid this over-worked term and refer to it instead as a *realization* of the theory of groups. A realization of a theory T in any category $\mathcal{C}$ is essentially the same thing as a functor $\mathrm{Syn}(T) \to \mathcal{C}$ that preserves the relevant structure of the theory – in the case of groups, the finite products $X \times X$. (This is in fact the defining universal property of the syntactic category $\mathrm{Syn}(T)$.) The realizations in the category SET, consisting of all sets and functions, are thus exactly what we called the *instances* of the general concept of a group, namely groups.

The standard category GROUP of all groups and their homomorphisms, as usually defined in abstract algebra, is then essentially the same as the category of all such instances, that is, the category $\mathrm{REAL}(\mathrm{Syn}(T_{Group}), \mathrm{SET})$ of all SET realizations, i.e. (structure-preserving) functors, where the morphisms are just natural transformations of such functors (that these correspond exactly to group homomorphisms is not trivial). In this way, for any general concept

[7] For example there is an axiomatization of groups using a single ternary operation in place of the two operations $x * y$ and x^{-1}.

$\mathrm{Syn}(T)$ corresponding to a theory T we can define the category of its SET realizations,

$$\mathrm{REAL}(T) =_{df} \mathrm{REAL}(\mathrm{Syn}(T), \mathrm{SET}),$$

which may be viewed as the category of instances of the concept $\mathrm{Syn}(T)$.

Now an amazing and mathematically deep fact emerges, which can only be seen using the tools of categorical logic: from the category $\mathrm{REAL}(T)$ of all instances of the concept presented by T, one can actually recover the general concept $\mathrm{Syn}(T)$. Indeed, for any structured category $\mathcal{R}$ of the same kind as $\mathrm{REAL}(T)$ (we will say a bit more about the condition "of the same kind" below), one can consider all of the *continuous* functors $f : \mathcal{R} \to \mathrm{SET}$; these may be regarded as "images" or "abstractions" of the (generalized) realizations in $\mathcal{R}$. The category of all such abstractions $\mathrm{ABSTRACT}(\mathcal{R}, \mathrm{SET})$ (again, with natural transformations as morphisms) may be called the *abstract* of $\mathcal{R}$, and written simply

$$\mathrm{ABSTRACT}(\mathcal{R}) =_{df} \mathrm{ABSTRACT}(\mathcal{R}, \mathrm{SET}).$$

A similar construction that the reader may know is the ring $\mathcal{C}(X) = \mathcal{C}(X, \mathbb{R})$ of continuous, real-valued functions on a space X. The noteworthy fact that we mentioned above is this: if for $\mathcal{R}$ we take a category $\mathrm{REAL}(T)$ of realizations of a theory T, then the abstract of $\mathrm{REAL}(T)$, consisting of all "abstractions" $\mathrm{REAL}(T) \to \mathrm{SET}$, will be the associated concept $\mathrm{Syn}(T)$.[8] Thus *the abstraction of the realizations of a concept is the concept itself*. We can even summarize this briefly by saying that *All concepts are abstract*, since every

[8] Under suitable assumptions, and up to the relevant notion of equivalence, of course; see (Awodey, 2019) for the general theory.

concept is the abstraction of its realizations. More generally, for any suitable category $\mathcal{R}$, the category $\mathrm{ABSTRACT}(\mathcal{R})$ of all continuous functors $f : \mathcal{R} \to \mathrm{SET}$ (the "abstractions" of $\mathcal{R}$) is a general concept, of which $\mathcal{R}$ is either the category of realizations, or an approximation thereof.

The general correspondence is given by a (contravariant!) adjunction between the *functors* of Realization and Abstraction which relate these operations; schematically,

$$\begin{tikzcd} \mathrm{CONCEPTS} \arrow[r, bend left, "\mathrm{Realization}"] & \mathrm{INSTANCES}^{op} \arrow[l, bend left, "\mathrm{Abstraction}"] \end{tikzcd}$$

Here CONCEPTS is the category consisting of all "conceptual" categories $\mathrm{Syn}(T)$ and their (relative) "realizations", i.e. functors $\mathrm{Syn}(T) \to \mathrm{Syn}(T')$, and the functor of Realization is defined by taking realizations in SET,

$$\mathrm{Realization}(\mathrm{Syn}(T)) = \mathrm{REAL}(\mathrm{Syn}(T), \mathrm{SET})\,,$$

which we also called the category of "instances" of the concept.

And INSTANCES is the category consisting of all (generalizaed) categories of instances $\mathcal{R}$ (such as the categories GROUP, RING, etc.) with their "continuous" functors $\mathcal{R} \to \mathcal{R}'$, and the functor of Abstraction is defined by taking continuous functors into SET,

$$\mathrm{Abstraction}(\mathcal{R}) = \mathrm{ABSTRACT}(\mathcal{R}, \mathrm{SET})\,,$$

which we called "abstractions" of the category $\mathcal{R}$.

Let us consider a simple example! Propositional logic consists of basic propositional variables $x, y, z, \ldots$ and constants $\top, \bot$, which can be made into formulae using the usual propositional connectives $\neg z\,, x \wedge y\,, x \vee y\,, x \Rightarrow z$, and which are assumed to satisfy the usual logical laws, such as $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)\,, \neg\neg x = x$, etc. A *theory* T in this simplified case is just a set of propositional letters $V = \{p_1, p_2, \ldots, p_n\}$ (regarded a 0-ary relation symbols), and a list of propositional formulae $A = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ built up from these letters, as the axioms of the theory. There are no types, typed variables, or function symbols (or rather, there is a single, implicit type 1), and the logical symbols are just the propositional connectives.

The *syntactic category* $\mathrm{Syn}(T)$, representing the "concept", is then the Boolean algebra $F(V)/A$ obtained as the free Boolean algebra $F(V)$ on the variables V as generators, quotiented by the filter generated by the axioms A. This "concept" associated to the propositional theory $T = (V, A)$ is independent of the particular syntactic presentation (V, A). A *realization* of T is then a boolean homomorphism $F(V)/A \to 2$, where $2 = \{0, 1\}$ is the Boolean algebra of truth values. Thus such a realization is just a truth-value assignment to the variables in V, in such a way that the "conditions" in A are all satisfied, i.e. the elements $a \in A$ are all taken to the value "true" (in other words, a "model" of the propositional theory T). For instance, if the theory T is $V = \{x, y\}$ and $A = \{x \vee y, \neg(x \wedge y)\}$, then a realization would be an assignment of x to an actual sentence p, and y to one q, such that only one of p and q is true (or more formally, a direct assignement of such truth values, by-passing the actual sentences). Under our description above, such a realization is an instance of the general concept $F(V)/A$.

Now, such realizations are exactly the points of the *Stone space* $\mathrm{Stone}(F(V)/A)$, the topological space associated to the Boolean algebra $F(V)/A$ under the celebrated Stone duality theorem (Johnstone, 1982) – which is in fact the "propositional logic" case of the categorical logical duality that we are considering here. Formally, the points of $\mathrm{Stone}(F(V)/A)$ are prime filters in $F(V)/A$, and the topology has basic open sets determined by the elements of $F(V)/A$. The Boolean algebra $F(V)/A$ can be recovered from this space $\mathrm{Stone}(F(V)/A)$ as the algebra of continuous functions $\mathrm{Stone}(F(V)/A) \to 2$ into the discrete space 2, with the pointwise Boolean operations. These *abstractions* of $\mathrm{Stone}(F(V)/A)$ form a Boolean algebra $\mathrm{Bool}(\mathrm{Stone}(F(V)/A))$ which, by Stone duality, is isomorphic to $F(V)/A$,

$$\mathrm{Bool}(\mathrm{Stone}(F(V)/A)) \cong F(V)/A\,.$$

Indeed, for any (not necessarily Stone) space X, we can form the Boolean algebra $\mathrm{Bool}(X)$ of continuous functions $X \to 2$, and the original space X will then map canonically to $\mathrm{Stone}(\mathrm{Bool}(X))$, giving the "best approximation" of X by a Stone space.

In the general case, in categorical logic we consider many other fragments of logic — propositional, equational, coherent, first-order — and for each such subsystem there is an associated Realization–Abstraction adjunction between theories, and the concepts they represent, on the one hand, and their realizations by instances of these concepts, on the other. The propositional theories just considered give rise to Stone duality (Johnstone, 1982); equational theories (like groups) give rise to Lawvere duality (Adámek, Lawvere and Rosický, 2003); coherent and first-order logic are treated by analogous duality theories developed by Makkai and others (Makkai, 1987; Awodey and Forssell, 2013). In each diffferent case, the associated notion of structured category, structure-preserving functor, continuous functor, etc., is suitably adapted to the respective situation. Many of these logical dualities are discussed from the standpoint of categorical logic in the paper (Awodey, 2019).

6. Categories BRAIN and MIND

So far everything that has been said has merely been a reminder of standard and well-known things. From now on, everything will be hypothetical and highly simplified. The bold and maximally simplified hypothesis is that neurons in the brain can be modeled as categories, the internal logic of which is sufficiently complex (yet manageable). Of course, our inspiring motive is the human brain, and in constructing our model we shall try to imitate what is going in it; however, being conscious of our simplified and highly idealized assumptions, we prefer to speak about a homunculus brain. Our homunculus is a kind of "mathematical robot", the aim of which is to provide us with the opportunity to study how such a simple brain-like structure could "create meanings" out of its purely syntactic program. Our other drastically simplifying assumption consists in systematically ignoring all of the brain's functions and processes that are not directly related to the proposed syntax–semantics relationship.

As it is well-known, neurons communicate through signals transmitted via: presynaptic (source) neuron – axon – synapse – dendrite – postsynaptic (target) neuron, and this *via* is unidirectional. In our homunculus model, these transmission processes will be regarded as functors between categories (neurons).

Let us consider the category CATEGORIES, which we now aptly call BRAIN. Its objects are categories modeling neurons, and morphisms are functors between these categories.

We thus assume that each neuron in the homunculus brain is represented by a category (belonging to a certain class of categories; in the following we shall simply say that a neuron is a category). At the moment, we are not interested which biological mechanisms implement this assumption. Everything that counts in this model is the assumption that neurons consist of collections of objects and morphisms satisfying conditions from the category definition. We should have in mind that these simple conditions might lead to highly complicated structures.

Morphisms (arrows) in the category CATEGORIES are functors between object-categories, that is to say axons through which neurons communicate with each other. The crucial thing is that they must satisfy the usual conditions for morphisms: composition of morphisms, its associativity, the existence of identity morphisms. With the latter there is no problem: no output from a neuron counts as its identity morphism. To check whether two other conditions are verified in the human brain would require going deeper into the neural structure of our brain. In the case of the homunculus brain, this is not necessary. Since the homunculus is of our construction, we simply assume that synapses in its brain well-compose and do so in the associative way.

The next step seems obvious. Each neuron (modeled as a category $C \in$ BRAIN) has its own program enabling its working, i.e. an internal logic underlying this program. We thus can define a counterpart of $\mathrm{Lang}(C)$ which is a "theory" of this neuron. It is reasonable to claim that it is an object of the category THEORIES which we now call MIND, and the functor Lang: BRAIN $\to$ MIND is defined in analogy to that between CATEGORIES and THEORIES.

What about the morphisms between such objects? We proceed in strict analogy with what has been done in THEORIES. Let now T_1 and T_2 be objects in MIND, a morphism between them, $T_1, \to T_2$, is a functor between their corresponding syntactic theories, i.e. $\mathrm{Syn}(T_1) \to \mathrm{Syn}(T_2)$, where the functor Syn: MIND $\to$ BRAIN is defined in analogy to that between THEORIES and CATEGORIES.

The analogy is only apparently straightforward. In fact, it is based on a huge extrapolation, and as such highly hypothetical, but it is worth exploring it since the problem at stake deserves even a higher risk. By pursuing this analogy we could claim that also in this case the functors Lang and Syn are adjoint functors. If so, we have a very interesting conjunction between brain and mind; it is interesting even if brain and mind are modeled by such a naive construction.

Neurons, their interactions and programs underlying their working are, in contrast with abstract categories like CATEGORIES and THEORIES, real things, at least in the homunculus world, and we are entitled to suppose that the functors Lang and Syn between Brain and Mind really do what they formally signify (like our phase portrait on the computer screen really did what the program told it to do).

Roughly speaking the functor Lang provides a collection of theories (mind) for a collection of neurons (brain), and the functor Syn transfers the syntax of these theories to the network of neurons. The action of these two functors is adjoint; consequently it determines a strict interaction between BRAIN and MIND. Let C be any object (a neuron) in BRAIN and T any object (the theory of this neuron) in MIND, then equation (1) assumes the form

$$\text{(2)} \qquad \mathrm{MIND}(\mathrm{Lang}(\mathcal{C}), T) \cong \mathrm{BRAIN}(\mathcal{C}, \mathrm{Syn}(T)).$$

The natural isomorphism $\cong$ appearing in this equation is crucial. It states that when we go from neuron to neuron as objects in BRAIN, and their corresponding theories vary in THEORIES, then the isomorphism between morphisms $\mathrm{Lang}(\mathcal{C}) \to T$ in MIND and $\mathcal{C} \to \mathrm{Syn}(T)$ in BRAIN varies in a way that is compatible with the composition of morphisms in BRAIN and MIND, correspondingly, and with the actions of the functors Lang and Syn (Awodey and Forssell, 2013, see[)[][]Leinster.[9] Finally, the "higher" cognitive functions of abstraction and realization of concepts are modelled by a corresponding adjunction between the associated functors $\mathrm{Abstraction}$ and $\mathrm{Realization}$ relating these categories BRAIN and MIND. We could summarise the situation by saying that the categories BRAIN and MIND interact with each other with their entire structures and, at the same time, these very structures are shaped by this interaction.

7. A Comment

The interactions between syntax and semantics are omnipresent both in our everyday conversations and in various forms of practicing science. The world around us is full of meanings and our attempts to decipher them. Science could be regarded as a machine to produce signs, through experimentation and critical reasoning, and extracting from combinations of them information about the structure of the world. Logicians put a lot of effort to make the syntax–semantics interaction precise. As we have seen in section 2, despite the fact that formal languages are lacking any external references, it was possible to create semantical references for them by cleverly exploiting the relation between language and its metalanguage. In categorical logic the situation has improved. Any formal theory T generates via the functor Syn the category $\mathrm{Syn}(C) = C_T$ of which it is a theory, i.e. C_T provides a "natural" semantics for T. And *vice versa*, any (sufficiently rich) category C', via the functor Lang, generates its own theory $\mathrm{Lang}(C') = T'_{C'}$ which constitutes the internal logic of C'. It is interesting to notice that T_{C_T} does not coincide with T, they are

[9] For a full discussion of the role of the naturality condition in the definition of adjoint functors see any textbook on category theory.

only Morita equivalent. Here, we shall not go into technical details; it is enough to say that two Morita equivalent theories could be regarded as two interpretations of the same theory (Halvorson, 2016).

The fact that T_{C_T} does not coincide with T is a consequence of the fact that the functors Lan and Syn are not mutually inverse functors but constitute a pair of adjoint functors. This in turn implies that in categorical logic the interaction between syntax and semantics is skillfully complex, with creative influences coming both ways.

All the above discussed properties of the syntax–semantics interaction can be presumed to be preserved if applied to the categories BRAIN and MIND. There is only one big difference: now "neurons and their theories" are real things (although in a highly idealised, toy version in the homunculus world). Nevertheless, the situation is not so different from the one which we can observe in many empirical sciences, in which some abstract mathematical structures model some real processes (always more or less idealised). We should not be surprised that the method of mathematical modeling works when applied to our cognitive processes, but rather that mathematical structures not only describe the real world (whether it is our brain or the world of physics), but that they are also effectively acting in it (like in the little arrow on the computer screen).

Bibliography

Adámek, J., Lawvere, F. and Rosický, J., 2003. On the duality between varieties and algebraic theories. *Algebra Universalis* [Online], 49(1), pp.35–49. Available at: https://doi.org/10.1007/s000120300002 [visited on 3 November 2020].

Awodey, S., 2019. *Sheaf Representations and Duality in Logic*.

Awodey, S. and Forssell, H., 2013. First-order logical duality. *Annals of Pure and Applied Logic* [Online], 164(3), pp.319–348. Available at: https://doi.org/10.1016/j.apal.2012.10.016 [visited on 3 November 2020].

Batterman, R., 2016. Intertheory Relations in Physics. In: E.N. Zalta, ed. *The Stanford Encyclopedia of Philosophy (Fall 2016 Edition)* [Online]. Stanford: Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/fall2016/entries/physics-interrelate/> [visited on 3 November 2020].

Borceux, F., 1994. *Handbook of Categorical Algebra. 3: Categories of Sheaves*, *Encyclopedia of Mathematics and its Applications* vol. 52. Cambridge: Cambridge University Press.

Ehresmann, A., 2017. Applications of Categories to Biology and Cognition. In: E. Landry, ed. *Categories for the Working Philosopher*. Oxford, New York: Oxford University Press, pp.358–380.

Ellerman, D., 2015. On Adjoint and Brain Functors. *arXiv:1508.04036 [math]* [Online]. Available at: <http://arxiv.org/abs/1508.04036> [visited on 3 November 2020].

Fu, Y., 2019. *Category Theory, Topos and Logic: A Quick Glance* [Online]. Available at: <http:%20//charlesfu.me/repository/topos.pdf> [visited on 28 February 2019].

Gómez, J. and Sanz, R., 2009. *Modeling Cognitive Systems with Category Theory: Towards Rigor in Cognitive Sciences* [Online]. Autonomous Systems Laboratory, Universidad Politécnica de Madrid. Available at: <http://www.aslab.org/documents/controlled/ASLAB-A-2009-014.pdf> [visited on 4 November 2020].

Halvorson, H., 2016. Scientific Theories. In: P. Humphreys, ed. *Oxford Handbook of Philosophy of Science* [Online]. Vol. 1. Oxford: Oxford University Press, pp.402–429. Available at: https://doi.org/10.1093/oxfordhb/9780199368815.013.33 [visited on 3 November 2020].

Halvorson, H. and Tsementzis, D., 2017. Categories of Scientific Theories. In: E. Landry, ed. *Categories for the Working Philosopher*. Oxford, New York: Oxford University Press, pp.402–429.

Healy, M.J. and Caudell, T.P., 2006. Ontologies and Worlds in Category Theory: Implications for Neural Systems. *Axiomathes* [Online], 16(1), pp.165–214. Available at: https://doi.org/10.1007/s10516-005-5474-1 [visited on 3 November 2020].

Hodges, W., 2018. Tarski's Truth Definitions. In: E.N. Zalta, ed. *The Stanford Encyclopedia of Philosophy (Fall 20218 Edition)* [Online]. Stanford: Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/%20archives/fall2018/entries/tarski-truth/> [visited on 3 November 2020].

Johnstone, P.T., 1982. *Stone Spaces*, *Cambridge studies in advanced mathematics* 3. Cambridge: Cambridge University Press.

Koch, C., 1997. Computation and the single neuron. *Nature* [Online], 385(6613), pp.207–210. Available at: https://doi.org/10.1038/385207a0 [visited on 3 November 2020].

Leinster, T., 2014. *Basic Category Theory*, *Cambridge Studies in Advanced Mathematics* 143. Cambridge: Cambridge University Press.

Mac Lane, S. and Moerdijk, I., 1992. *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*, *Universitext*. New York [etc.]: Springer-Verlag.

Makkai, M., 1987. Stone duality for first order logic. *Advances in Mathematics*, 65, pp.97–170.

McCulloch, W.S. and Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* [Online], 5(4), pp.115–133. Available at: https://doi.org/10.1007/BF02478259 [visited on 3 November 2020].

Mizraji, E. and Lin, J., 2011. Logic in a Dynamic Brain. *Bulletin of Mathematical Biology* [Online], 73(2), pp.373–397. Available at: https://doi.org/10.1007/s11538-010-9561-0 [visited on 3 November 2020].

nLab, 2011. *Coherent Functor*. Available at: <https://ncatlab.org/nlab/show/coherent+functor> [visited on 3 November 2020].

nLab, 2017. *Internal Logic* [Online]. Available at: <https://ncatlab.org/nlab/show/internal+logic> [visited on 3 November 2020].

nLab, 2020a. *Doctrine* [Online]. Available at: <https://ncatlab.org/nlab/show/coherent+functor> [visited on 3 November 2020].

nLab, 2020b. *Substitution* [Online]. Available at: <https://ncatlab.org/nlab/show/substitution> [visited on 3 November 2020].

nLab, 2020c. *Syntactic Category* [Online]. Available at: <https://ncatlab.org/nlab/show/syntactic+category> [visited on 3 November 2020].

Rosaler, J., 2018. Inter-Theory Relations in Physics: Case Studies from Quantum Mechanics and Quantum Field Theory. *arXiv:1802.09350 [quant-ph]* [Online]. Available at: <http://arxiv.org/abs/1802.09350> [visited on 3 November 2020].

Rosen, R., 1985. Organisms as Causal Systems Which Are Not Mechanisms: An Essay into the Nature of Complexity. *Theoretical Biology and Complexity: Three Essays on the Natural Philosophy of Complex Systems* [Online]. New York [etc.]: Academic Press, pp.165–203. Available at: https://doi.org/10.1016/B978-0-12-597280-2.50008-8 [visited on 3 November 2020].

Sher, G., 1999. What is Tarski's Theory of Truth? *Topoi* [Online], 18(2), pp.149–166. Available at: https://doi.org/10.1023/A:1006246322119 [visited on 3 November 2020].

Simmons, H., 2011. *An Introduction to Category Theory*. Cambridge [etc.]: Cambridge University Press.

Tarski, A., 1933. *Pojęcie prawdy w językach nauk dedukcyjnych*, *Prace Towarzystwa Naukowego Warszawskiego. Wydział III: Nauk Matematyczno-Fizycznych* 34. Warszawa: nakładem Towarzystwa Naukowego Warszawskiego, z zasiłku Ministerstwa Wyznań Religijnych i Oświecenia Publicznego.

Tsuchiya, N., Taguchi, S. and Saigo, H., 2016. Using category theory to assess the relationship between consciousness and integrated information theory. *Neuroscience Research* [Online], 107, pp.1–7. Available at: https://doi.org/10.1016/j.neures.2015.12.007 [visited on 3 November 2020].

Woszczyna, A. and Heller, M., 1990. Is a horizon-free cosmology possible? *General Relativity and Gravitation* [Online], 22(12), pp.1367–1386. Available at: https://doi.org/10.1007/BF00756836 [visited on 3 November 2020].

Abstract logical structuralism*

Jean-Pierre Marquis
Department of Philosophy, University of Montreal

Abstract

Structuralism has recently moved center stage in philosophy of mathematics. One of the issues discussed is the underlying logic of mathematical structuralism. In this paper, I want to look at the dual question, namely the underlying structures of logic. Indeed, from a mathematical structuralist standpoint, it makes perfect sense to try to identify the abstract structures underlying logic. We claim that one answer to this question is provided by categorical logic. In fact, we claim that the latter can be seen—and probably should be seen—as being a structuralist approach to logic and it is from this angle that categorical logic is best understood.

Keywords

none.

1. Introduction

In their recent booklet *Mathematical Structuralism*, Hellman and Shapiro (2019), give a list of eight criteria to evaluate various

* The author gratefully acknowledges the financial support of the SSHRC of Canada while this work was done. I also want to thank the organizers of the 23rd Kraków Methodological Conference for inviting me and giving me the opportunity to give a talk in such a lovely venue.

Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)
No 69 (2020), pp. 35–78

strands of mathematical structuralism. The very first criterion is the background logic used to express the version of structuralism examined: is it first-order, second-order, higher-order, modal? The claim made by Hellman & Shapiro is that the logic has a direct impact on the philosophical thesis: for instance, the existence of non-standard models in first-order logic seems to affect some of the claims made by a mathematical structuralist. Be that as it may, it is assumed that any variant of mathematical structuralism is based on an underlying logic and the chances are that a change of logic will modify the type of structuralism defended. One then has to weigh the pros and the cons of adopting a specific logic to defend a kind of mathematical structuralism.

This is all well and good, and I don't intend to discuss this assumption in this paper. Let me rather turn the question of the underlying logic on its head. A mathematical structuralist not only believes that *pure* mathematics is about abstract structures, but also that *pure* logic can be seen that way. In other words, a mathematical structuralist aspires to know what are the abstract structures underlying a given logic. Can logic itself be given a structuralist treatment? In other words, is it possible to identify the *abstract mathematical structures* from which the standard logical systems can be derived? In the same way that the natural numbers or the real numbers can be seen as specific structures arising from the combination of particular structures and properties, specific logical systems, e.g. intuitionistic first-order logic, classical first-order logic would result from the combination of particular abstract structures and properties. And important logical theorems—completeness theorems, incompleteness theorems, definability, undecidability, etc.—would be instances of abstract structural features, together with, presumably, singular properties.

Our main claim in this paper is that categorical logic is one way to give precise answers to these questions.[1] Indeed, this is what categorical logic is all about: it reveals the abstract mathematical structures of logic and it relates them to other abstract mathematical structures, revealing yet other structural features. It also shows how the main results of logic are a combination of abstract structural facts and specific properties of logical systems. It is in this sense that one can say that *pure* category theory *is* logic.[2] This is one of the main *points* of categorical logic. Thus, when mathematical structuralism is developed within the metamathematical framework of category theory, it is possible to give a positive answer to our new challenge. And as far as I know, it is at the time the only metamathematical framework that allows us to do it in such generality.

Of course, we are not claiming that the abstract analysis is *superior* in a strong sense to the other presentations of logical systems. It brings a certain perspective, a certain understanding and opens up certain connections that are otherwise unavailable. Furthermore, in as much as a logic is applied, be it in foundational studies or in computer science, one also wants to look at it from a different point of view. But it has to be perfectly clear that these are completely different issues. We are now positioning ourselves in a structuralist framework, and will therefore ignore the aspects of a logic that become prevalent when one look at it for its applications. Ours is a *philosophical* goal, not a technical one.

---

[1] In his (1996), often quoted as representing the categorical perspective on mathematical structuralism, Awodey did include a presentation of the basic features of logic from a categorical point of view, more specifically the internal logic of a topos. Given how he treated logic in his paper, none of his critics saw that he was including logic itself as an object of mathematical structuralism (see, for instance, Hellman, 2001; 2003). We approach the issues from a different angle.

[2] Which is not to say that it is *all* of logic. That is *not* the claim.

It goes without saying that categorical logic, seen as a search for the abstract structures of logic, did not come out of the blue. It was part of a larger mathematical movement, namely the structuralist movement in mathematics, which culminated in Bourbaki's *Éléments de mathématique* (for more on the prehistory and the history of mathematical structuralism, see Corry, 2004; Reck and Schiemer, 2020). It will therefore be worth our while to look briefly at the birth of categorical logic in the 1960s and early 1970s. We will merely indicate the major landmarks of the story, to see how indeed categorical logic was, right from the start, understood as being a part of that movement. We will then briefly look at some of the basic abstract structures that correspond to logic and then how some of the main theorems follow from features of abstract structures. At this point, to call the latter 'mathematical' or 'logical' is a matter of choice.

In the end, this paper should be taken as a challenge. We claim that one should add a ninth criterion of Hellman & Shapiro's list: how does a given form of structuralism treat logic itself? Does it reveal the structures of logic? Are these structures related to other mathematical structures in a natural manner? Are these structures on the same level as the other fundamental structures of mathematics? One could, and we suggest that one should, evaluate different forms of mathematical structuralism according to this criterion.

2. The abstract structures of logic: setting the stage

Although category, functors and natural transformations were introduced explicitly in 1945 by Eilenberg and Mac Lane, category *theory* came to maturity only fifteen years later, thus at the beginning of the

1960s (for more on this history, see, for instance Landry and Marquis, 2005; Krömer, 2007; Marquis, 2009; Rodin, 2014). At that point, the major concepts of category theory were in place, e.g. adjoint functors, representable functors, constructions on categories, including the crucial construction of functor categories, abstract categories like additive categories, abelian categories, tensor or monoidal categories, etc.[3] But it is only in the early 1960s that the connection with logic and the foundations of mathematics was made and it was mostly the work of one person, namely Bill Lawvere. We will not present Lawvere's early work here, for our goal is not to survey the history of the subject, but rather to make a conceptual point.[4]

Of course, by that time, connections between classical propositional logic and Boolean algebras, intuitionistic propositional logic and Heyting algebras, as well as others were known. Already in (Birkhoff, 1940), the main relations are presented.[5] The first links between logic and lattice theory were restricted to propositional logic. The chase for the identification of the abstract structures corresponding to first-order, higher order logics, non-classical logics, as well as algebraic proofs of the main theorems of logic was taken up by, among others, Tarski and his school, Halmos, and the Polish school,

[3] The *locus classicus* of the time is (Mac Lane, 1965).

[4] See (Lawvere, 1963; Lawvere, 1966; 1967; 1969a; Lawvere, 1970; Lawvere, 1971). He was rapidly joined by (Freyd, 1966; Linton, 1966; Lambek, 1968a,b; 1969; 1972). For a detailed history, see (Marquis and Reyes, 2012).

[5] We could argue that already for propositional logics, the structures arising from the logical systems naturally live in categories, e.g. the category of distributive lattices, the category of Boolean algebras, the category of Heyting algebras, the category of S4-algebras, etc., and that the main results are also naturally expressed in categorical language. It is in fact an important point to make, for once the higher order logical systems find their place in this landscape, the fact that we move to bicategories to express and prove the results of first-order logic is easily understood. But we will not dwell on that.

e.g. Łoś, Mostowski, Rasiowa and Sikorski.[6] The main contenders to the title of abstract algebraic structures corresponding to first order logic at the time were cylindric algebras and polyadic algebras. The main problem, so to speak, were the quantifiers $\forall$ and $\exists$. It was not a technical problem. They were treated properly in each case. But their treatment, as algebraic operators, was somewhat *ad hoc*, in the sense that they did not arise as an instance of abstract operators in an algebraic context. The resulting algebras were therefore somewhat *ad hoc* also, in as much as they did not belong to a family of abstract structures that arose naturally in other mathematical contexts. In other words, the abstraction proposed via the concepts of cylindric algebras or polyadic algebras were not genuine mathematical abstractions, for they were merely the algebraic transcription of the quantifiers and solely of the latter.[7] This is in stark contrast with the case of propositional logic, where the abstract algebraic structures capturing the logic and its main properties have instances in a variety of completely different mathematical fields. Distributive lattices, Boolean algebras, Heyting algebras, etc., are genuine mathematical abstractions.

It therefore came as a complete surprise that the quantifiers, as well as the propositional connectives, could be seen as being instances of adjoint functors on very simple categories, the concept of adjoint functors being one of the core concepts of category theory,

[6] The list of references is long, but clearly indicates that it was a very active area of research in the 1950s as well as in the 1960s. See, for instance, (McKinsey and Tarski, 1944; 1946; 1948; Jónsson and Tarski, 1951; 1952; Henkin and Tarski, 1961; Henkin, Monk and Tarski, 1971; Halmos, 1954; 1956b; Halmos, 1956a; Halmos, 1956c,d; 1962; Mostowski, 1949; Rasiowa, 1951; 1955; Rasiowa and Sikorski, 1950; 1953; 1955; Rasiowa and Sikorski, 1963).

[7] The reader might wonder what we mean by "genuine mathematical abstraction". We refer her to (Marquis, 2015; 2016).

introduced by Kan in the context of algebraic topology in 1958. This was one of Lawvere's crucial observations. Three additional crucial facts had been established by Lawvere in his Ph.D. thesis. First, Lawvere showed how algebraic theories, in the standard logical sense of that expression, could themselves be captured by specific categories. Second, the models of algebraic theories, again in the standard logical sense of that expression, could be described in the language of categories, functors and natural transformations. Third, the classical links between the syntax and the semantics of these theories could receive an adequate categorical treatment, and at the core of this treatment one finds adjointness. Thus, it seemed possible to put all the structures of logic in the theoretical framework of categories, the latter being, of course, an abstraction of a central fact of modern mathematics. The overall plan was presented by Lawevere in (1969a). That paper articulates in very broad strokes how the syntax, the semantics and their relationships could be captured in a categorical framework. Here are, in a nutshell, the main ingredients of this ambitious program.

A few words about the philosophical framework underlying Lawvere's program are in order. Lawvere identifies two fundamental aspects to all of mathematics, namely the formal and the conceptual, roughly the manipulation of symbols, on the one hand, and what these symbols refer to, their content. Lawvere is aware of the work done in algebraic logic when he writes his paper. Indeed, he refers to it explicitly in the opening section: "[...] Foundations may conceptualize the formal aspect of mathematics, leading to Boolean algebras, cylindric and polyadic algebras, [...]"(Lawvere, 1969a, p.281) He is also presenting the introduction of categories in the analysis of logic as a structural approach, based on the notion of adjoints: "Specifically, we describe [...] the notion of cartesian closed category, which appears to be the appropriate abstract structure for making explicit [...]. The structure of a cartesian closed category is entirely given by adjointness, as is the structure of a 'hyperdoctrine', which includes quantifiers as well."(Lawvere, 1969a, p.281)

We will now focus on the final section of the paper, which is really programmatic. In this last section, Lawvere is describing what he himself characterizes as a globalized Galois connection, and indeed, it also contains the main ideas that have driven the development of duality theory in a categorical framework. But as far as logic is concerned, we are offered the following picture.

1. Logical operations should arise from an elementary context as adjoint operations. From a categorical point of view, a logical doctrine, that is an abstract mathematical structure encapsulating a logical framework, should be given by adjoint functors.[8]
2. A theory $\mathbf{T}$, in the standard logical sense of the term, should be constructed as a category, in the same way that a propositional theory in classical logic can be turned into a Boolean algebra via the Lindenbaum-Tarski construction. A theory $\mathbf{T}$, seen as a special type of category, is conceived by Lawvere as being the invariant notion of a theory, that is, independent of a choice of primitive symbols or specific axioms. We thus have abstract mathematical structures corresponding to the formal.

[8] We have to point out that categorical logic does cover logical situations in which certain logical operations are not given by adjoint operations. Although they do not constitute logical *doctrines* in the sense of Lawvere, they are part and parcel of categorical logic. We should also mention that the syntactic aspects of logic, which are pushed in the background in Lawvere's early work, occupy nonetheless an important part of categorical logic, for instance via the notion of sketch, introduced by Ehresmann and his school or the various formal graphical languages developed mostly in the context of monoidal categories.

3. The models of $\mathbf{T}$ should form a category. Lawvere, having himself developed the case of algebraic theories earlier in his thesis, generalizes from his work and proposes to make the category of models of a theory a functor category. We will be more specific in later sections. These provide the abstract mathematical structures emerging from the conceptual.
4. Last, but certainly not least, since everything is a category now, the links between the formal and the conceptual should also be given by (adjoint) functors, and we have yet again a new type of abstract mathematical structure, in Lawvere's mind a globalized Galois connection, arising from that situation.

Lawvere was of course guided by his own work on algebraic theory, but also explicitly by Grothendieck's work in algebraic geometry. As I said, at the time, it was a program, some would say a vision. It became a reality in the following decade and is still the basis of important developments in the field.

We have to explain why we claim that we are then in a structuralist framework. It is not only because we are in fact dealing with abstract mathematical structures—this is of course a necessary step—but these abstract mathematical structures can be characterized up to 'isomorphism', where the latter notion is derived from the abstract structures themselves. Each and every one of these abstract structures comes with a notion of homomorphism and, in particular, a notion of 'isomorphism'. Therefore, it becomes possible to develop logic with respect to the structuralist principle: if X is a structure of a given kind, and it has property $P(X)$, then given any other structure Y of the same kind such that X is isomorphic to Y, $X \simeq Y$, in the appropriate sense of isomorphism, we should be able to prove that $P(Y)$. As it can be seen, the key component of this desideratum is the appropriate sense of isomorphism. In some cases, we are dealing with the usual set-theoretical sense of isomorphism, in others, it becomes an equivalence of categories and in still others, it is a 2-categorical equivalence. It is the very possibility of having the appropriate sense of isomorphism that allows us to claim that we are dealing with *abstract* mathematical structures.

I want to emphasize again, at this point, that I am not claiming that categorical logic, as I will present it succinctly below, is the only possible answer for a structuralist nor is it the final answer. But it is one clear answer and one of the very few that provides a comprehensive answer.

3. Categories as abstract logical structures

At this stage, we would have to give a long list of definitions and examples to illustrate how certain categories correspond to the abstract mathematical structures of certain logics. I will assume that the reader knows the notions of category, functor, natural transformation, adjoint functors, etc., for otherwise this paper would be terribly long and boring. We will try to put some flesh on Lawvere's program described in the foregoing section. We assume, however, that the description of the logical connectives, including the quantifiers, as adjoint functors is understood[9]. We will sketch how the other three steps are filled[10]. A warning is necessary. Each following section would require a careful and systematic exposition to be ultimately convincing. It is impossible to do justice to the field in such a short paper. We

[9] Mac Lane's textbook, (1998), is still a good reference. All the standard concepts and examples can also be found in (Riehl, 2017).

[10] A more detailed presentation can be found in (Marquis, 2009), chapter 6. Our exposition here is adapted from the latter, but the philosophical point is different and therefore the presentation found there might not be optimal for our present purpose.

will provide a more detailed presentation in the next section only and merely gloss over the abstract mathematical structures involved in the other sections. We apologize for the opacity that might result from the lack of details and clarifications, but a much longer paper would be required to present and motivate adequately the main mathematical ideas involved.

A theory as a category and a category as a theory

Let us start with the goal that was in the minds of logicians and mathematicians in the 1950s, that is finding the appropriate abstract mathematical structure that correspond to a first-order theory.

Let us fix the logical context first. We consider formal systems with many sorts, which is a simple generalization of the standard first-order logic which is done over a single sort. A *similarity type* or *alphabet* A, often called a language in the literature, is given by:

1. A collection of sorts $S_1, S_2, S_3, \ldots$;[11]
2. A collection of relation symbols $R_1, R_2, R_3, \ldots$, each of which is given with the sorts of its arguments;
3. A collection of function symbols $f_1, f_2, f_3, \ldots$ each of which is given with the sorts of its arguments and the sort of its target; we denote a function symbol f as $f\colon S_1 \times \cdots \times S_n \to S$ if f takes n arguments of sorts $S_1, \ldots, S_n$ respectively to a value of sort S.

[11] Or types, if you prefer. We are dealing here with first-order logic. We will say a few words about type theory later. It does not affect our basic general point. Type theories can also be analyzed as instances of abstract mathematical structures.

4. A collection of constants $c_1, c_2, c_3, \ldots$ each with a specified sort; we denote a constant c by $c\colon 1 \to S_i$ to indicate that the constant c is of sort S_i.

This is the standard definition extended to a many-sorted context. To obtain a *formal system* L_A in the alphabet A, we add the usual elements:

1. Each sort S_i comes with infinitely many variables $x_1, x_2, x_3, \ldots$; we write $x\colon S_i$ to indicate that the variable x is of sort S_i;
2. Each sort has an equality relation $=_S$; notice immediately that this means that equality is not treated as a universal or purely logical relation and that in the interpretation, whatever will correspond to a sort will have to come equipped with a criterion of identity or equality for *its* objects;
3. The usual logical symbols and two propositional constants, $\top$ and $\bot$;
4. The usual deductive machinery for a predicate logic (say, intuitionistic predicate logic; if any other deductive procedures are assumed, they are made explicit).

Terms (of a given sort) and atomic formulas are defined as usual.

Here is the first original result obtained by searching for abstract mathematical structures corresponding to theories in a given logic. Some fragments of first-order logic and some extensions of first-order logic turn out to have significant properties, properties that would not have been identified otherwise[12]. We can immediately identify the following fragments.

[12] We will not be exhaustive here. There are other infinitary fragments that are important, but we will ignore them.

1. A formula φ is said to be *regular* if it is obtained from atomic formulas by applying finite conjunction and existential quantification.
2. A formula φ is said to be *coherent* if it is obtained from atomic formulas by applying finite conjunction, disjunction and existential quantification.
3. A formula is said to be *geometric* if it is obtained from atomic formulas by applying *finite* conjunction, *finite* existential quantification and *infinite* disjunction.

Intuitionistic formulas and classical (or Boolean) formulas are defined in the obvious manner. We point out immediately that many metalogical results about intuitionistic and classical logic follow directly from results about the foregoing fragments. This is a genuine discovery that could not have been foreseen beforehand.

An *implication* of regular (resp. coherent, geometric, etc.) formulas φ and ψ has the form

$$\forall x_1 \ldots \forall x_n (\varphi(x_1, \ldots, x_n) \Rightarrow \psi(x_1, \ldots, x_n))$$

where $\varphi(x_1, \ldots, x_n)$ and $\psi(x_1, \ldots, x_n)$ are regular (resp. coherent, geometric, etc.) formulas. A theory $\mathbf{T}$ in the given language L is said to be a *regular theory* (resp. coherent, geometric, etc.) if all its axioms are implications of regular (resp. coherent, geometric, etc.) formulas. Many mathematical theories can be expressed in the form of regular theories, or coherent theories, etc.

Apart from the fact that we have assumed a many sorted context and cut the fragments of first-order logic in ways that might seem arbitrary, the foregoing presentation is squarely in a standard logical context. We now move to the abstract mathematical context.

Given a theory $\mathbf{T}$ in one of the foregoing languages, we construct a category denoted by $[\mathbf{T}]$, out of it. The latter is sometimes called the category of concepts, sometimes the syntactic category, and it is basically an extension of the Lindenbaum-Tarski construction, but for theories expressed in first-order logic (and others, as the reader can now guess). It is constructed from the language and the axioms of $\mathbf{T}$ as follows.[13]

Remember that we are constructing a category, thus a web of objects connected by morphisms. To get the objects of this category, we start with *formal sets* $[\mathbf{x}; \varphi(\mathbf{x})]$, where $\mathbf{x}$ denotes a n-tuple of distinct variables containing all free variables of φ and φ is a formula of the underlying formal system L. Two such formal sets, $[\mathbf{x}; \varphi(\mathbf{x})]$ and $[\mathbf{y}; \varphi(\mathbf{y})]$ are equivalent if one is the alphabetic variant of the other, that is if $\mathbf{x}$ and $\mathbf{y}$ have the same length and sorts and $\varphi(\mathbf{y})$ is obtained from $\varphi(\mathbf{x})$ by substituting $\mathbf{y}$ for $\mathbf{x}$ (and changing bound variables if necessary). This is an equivalence relation and it is therefore possible to consider equivalence classes of such formal sets. An object of the category of concepts $[\mathbf{T}]$ is such an equivalence class of formal sets $[\mathbf{x}; \varphi(\mathbf{x})]$, where φ is a formula of the formal system L. The objects of $[\mathbf{T}]$ are the *equivalence classes of these formal sets*, for all formulas of L. Notice this last important point: we take *all* formulas of the language, not only those which appear in $\mathbf{T}$. Thus, in a sense, the space of objects is the collection of all possible properties and sentences expressible in that language, thus all possible theories in the given formal system. No logical relationship is considered at this stage. The next step introduces the structure corresponding to the structure of that particular theory $\mathbf{T}$. This is just as one would expect in a categorical framework: the structure of $\mathbf{T}$ is captured by the morphisms we will define and the properties resulting therefrom.

[13] See also (Makkai and Reyes, 1977, chap. 8) or (Mac Lane and Moerdijk, 1994, chap. X, § 5) for more details and proofs or again (Johnstone, 2002).

It is easier to motivate the definition of morphism with an eye on the semantics, although the properties of the morphisms, e.g., that they form a category, have to be proved with the syntactical features of the theory (unless one has a completeness theorem at hand). The basic idea is this: a functor from $[\mathbf{T}]$ to $\mathbf{Set}$ should transform the objects of $[\mathbf{T}]$ into *genuine sets* and the morphisms of $[\mathbf{T}]$ into *genuine functions* automatically. These functions should be functions that are definable in $\mathbf{T}$—i.e., for which we can prove in $\mathbf{T}$ that they are indeed functions. Furthermore, $[\mathbf{T}]$ should contain all of them. By sending a formal set $[\mathbf{x}; \varphi(\mathbf{x})]$ to the set of n-tuples satisfying the formula, i.e. $\{(x_1, \dots, x_n) \mid \varphi(\mathbf{x})\}$, a morphism from $[\mathbf{x}; \varphi(\mathbf{x})]$ to $[\mathbf{y}; \psi(\mathbf{y})]$ should become a genuine function between genuine sets $\{(x_1, \dots, x_n) \mid \varphi(\mathbf{x})\}$ and $\{(y_1, \dots, y_m) \mid \psi(\mathbf{y})\}$ respectively. Such a morphism should simply be given by a formula of the theory $\mathbf{T}$ that defines such a function, that is a formula $\theta(\mathbf{x}, \mathbf{y})$ of $\mathbf{T}$ that is provably functional. The only trick in the construction is to construct a morphism between two (equivalence classes of) formal sets $[\mathbf{x}; \varphi(\mathbf{x})]$ and $[\mathbf{y}; \psi(\mathbf{y})]$ in such a way that, when interpreted, it yields the *graph* of the function, in the standard set-theoretical sense of that expression, between the actual sets $\{(x_1, \dots, x_n) \mid \varphi(\mathbf{x})\}$ and $\{(y_1, \dots, y_m) \mid \psi(\mathbf{y})\}$. Thus, all definable functions in $\mathbf{T}$ will be represented by a morphism in $[\mathbf{T}]$.

Formally, consider a triple $(\mathbf{x}, \mathbf{y}, \gamma)$, where $\mathbf{x}$ and $\mathbf{y}$ are disjoint tuples of distinct variables and γ is a formula with free variables possibly among $\mathbf{x}$ and $\mathbf{y}$. Such a triple defines a *formal function* if the following formulas are provable:

$$\mathbf{T} \vdash \forall\mathbf{x}\forall\mathbf{y}(\gamma(\mathbf{x}, \mathbf{y}) \Rightarrow (\varphi(\mathbf{x}) \wedge \psi(\mathbf{y})));$$

$$\mathbf{T} \vdash \forall\mathbf{x}(\varphi(\mathbf{x}) \Rightarrow \exists\mathbf{y}(\gamma(\mathbf{x}, \mathbf{y})));$$

$$\mathbf{T} \vdash \forall\mathbf{x}\forall\mathbf{y}\forall\mathbf{y}'(\gamma(\mathbf{x}, \mathbf{y}) \wedge \gamma(\mathbf{x}, \mathbf{y}') \Rightarrow \mathbf{y} = \mathbf{y}');$$

where we have used some obvious abbreviations. The underlying motivation should be clear: these formulas will be true in any interpretation of $\mathbf{T}$ in which γ is indeed a morphism.

We now define an equivalence relation $(\mathbf{x}, \mathbf{y}, \gamma) \sim (\mathbf{u}, \mathbf{v}, \eta)$ if

$$\mathbf{T} \vdash \forall \mathbf{x} \forall \mathbf{y} (\gamma \Leftrightarrow (\eta(\mathbf{x}/\mathbf{u}, \mathbf{y}/\mathbf{v}))).$$

The equivalence relation guarantees that for every model M of $\mathbf{T}$, the functions corresponding to γ and to η will coincide. We can now stipulate that a formal function is an *equivalence class* of the foregoing equivalence relation. Given a representative $(\mathbf{x}, \mathbf{y}, \gamma)$ of such an equivalence class, we denote the equivalence class containing it by $\langle \mathbf{x} \mapsto \mathbf{y} \colon \gamma \rangle$. Thus, a *formal morphism* in $[\mathbf{T}]$ is denoted by:

$$\langle \mathbf{x} \mapsto \mathbf{y} \rangle \colon [\mathbf{x} \colon \varphi] \to [\mathbf{y} \colon \psi].$$

We need two more ingredients to get to a category. Firstly, for each formal set $[\mathbf{x} \colon \varphi(\mathbf{x})]$, the identity morphism is provided by the formal morphism $\langle \mathbf{x} \mapsto \mathbf{y} \colon (\mathbf{x} = \mathbf{y}) \wedge \varphi \rangle$. Secondly, given two formal morphisms $\langle \mathbf{x} \mapsto \mathbf{y} \colon \gamma \rangle \colon [\mathbf{x} \colon \varphi] \to [\mathbf{y} \colon \psi]$ and $\langle \mathbf{y} \mapsto \mathbf{z} \colon \eta \rangle \colon [\mathbf{y} \colon \psi] \to [\mathbf{z} \colon \zeta]$, their composition is defined by the formal morphism $\langle \mathbf{x} \mapsto \mathbf{z} \colon \mu \rangle \colon [\mathbf{x} \colon \varphi] \to [\mathbf{z} \colon \zeta]$ where $\mu = \exists \mathbf{y} (\gamma \wedge \eta)$. These two definitions satisfy the usual requirements of a category. Thus, $[\mathbf{T}]$ is a category and we have an abstract mathematical structure corresponding to a given theory.

Notice that $[\mathbf{T}]$ is *not* a category of structured sets and structure-preserving functions! A lot of information about $\mathbf{T}$ is lost when all we have at our disposal is $[\mathbf{T}]$. It is, for instance, impossible to know which atomic formulas are involved in specific formal sets or what were the primitive symbols of the language $L_{\mathbf{T}}$. Furthermore, two different theories $\mathbf{T}$ and $\mathbf{T}'$ can very well yield isomorphic categories of concepts, thus essentially the same category. We are squarely in a structuralist framework: the category $[\mathbf{T}]$ is given up to an isomorphism of categories.[14] As we have seen, the syntactic logical operations, i.e., quantifiers and connectives, become categorical operations in the category and this part of the structure is not lost. Again, moving from a theory $\mathbf{T}$ to its category of concepts $[\mathbf{T}]$ is an *abstraction*: the specific formulas with specific variables are abstracted from when we move to the equivalence classes. The category of concepts is the category of all definable sets and functions of a theory $\mathbf{T}$. Thus, in a sense, it contains all the formally expressible concepts of $\mathbf{T}$, whence its name.

When we start with, for example, a regular theory, the foregoing construction yields a category with additional structures and properties. For instance, it is automatically a category with finite limits.

$\mathbf{T}$ and $[\mathbf{T}]$ are interchangeable in the following sense: for, given a (small) category $\mathcal{C}$, at least with finite limits, it is possible to associate or construct the language $L_{\mathcal{C}}$ of $\mathcal{C}$ as follows. We first have to identify the alphabet of $L_{\mathcal{C}}$. The sorts are given by the objects $X, Y, Z, \ldots$ of $\mathcal{C}$. Every morphism $f\colon X \to Y$ of $\mathcal{C}$ becomes a function symbol of $L_{\mathcal{C}}$. (In particular, a constant $c\colon 1 \to X$ is seen as 0-ary function symbols.) This is called the *canonical language* of $\mathcal{C}$. Notice that $L_{\mathcal{C}}$ is obtained as if we had taken $\mathcal{C}$ and destroyed its categorical structure, retaining only the symbols, and keeping in mind that function symbols are sorted. It is possible to extend this language to reflect the structure of $\mathcal{C}$ more closely. Although subobjects of $\mathcal{C}$ can be denoted naturally by formulas of $L_{\mathcal{C}}$, it is possible to introduce relation symbols for each subobject $R(x_1, \ldots, x_n) \rightarrowtail X_1 \times \cdots \times X_n$ and n-ary function symbol for morphisms $f\colon X_1 \times \cdots \times X_n \to X$. This is

[14] Notice that we are talking about isomorphism here and not an equivalence of categories.

called the *extended* canonical languageof $\mathcal{C}$ (see Makkai and Reyes, 1977, chap. 2, sec. 4)). In order to get the *internal theory* $\mathbf{T}_{\mathcal{C}}$ of $\mathcal{C}$ in its canonical language, $\mathcal{C}$ has to have more structure than just finite limits. It has to be at least a *regular* category, which we will define shortly. In this case, it is possible to give a list of *regular* axioms $\Sigma_{\mathcal{C}}$, that is a set of regular formulas, and prove that $\mathbf{T}_{\mathcal{C}}$ is sound in $\mathcal{C}$ (see Makkai and Reyes, 1977, chap. 3). The internal theory $\mathbf{T}_{\mathcal{C}}$ is related to $\mathcal{C}$ by two expected properties:

1. There is a canonical interpretation G of $\mathbf{T}_{\mathcal{C}}$ in $\mathcal{C}$;
2. For any model M of $\mathbf{T}_{\mathcal{C}}$ in a regular category $\mathcal{D}$, in any reasonable sense of the term 'model', there is a unique regular functor $I\colon \mathcal{C} \to \mathcal{D}$ such that I applied to G is equal to M.

It is of course possible to complete the circle: starting with a regular category $\mathcal{C}$, construct its internal theory $\mathbf{T}_{\mathcal{C}}$ and then move to its category of concepts $[\mathbf{T}_{\mathcal{C}}]$. How are $\mathcal{C}$ and $[\mathbf{T}_{\mathcal{C}}]$ related? They are in fact *equivalent as categories*, which means that they share the same categorical properties. In other words, as abstract mathematical structures, they are indistinguishable. From this, it is possible to conclude a very important result that every (small) regular category is equivalent to a category of concepts for some theory $\mathbf{T}$.

The Architecture of Logical Theories

We now have sketched how a theory in a logical framework can be turned into an instance of an abstract mathematical structure. It should not come as a surprise to learn that a regular theory $\mathbf{T}$ (resp. a coherent, geometric, etc.) yield a specific kind of abstract category, namely a regular category (resp. a coherent, geometric, etc.). We will fill in some blanks here, for we want to emphasize the existence of kinds of abstract mathematical structures. The existence of these abstract mathematical structures explains why we have introduced these fragments of first-order logic. Logic itself can be organized from the perspective of these structures. We thus get what we call the "architecture of logical theories" or the "architecture of logic".[15]

We will simply state the definition without explaining all the technical details. We refer the reader to the literature.

A *regular* category $\mathcal{C}$ is a category with finite limits[16], such that

1. Every morphism has a kernel pair;
2. Every kernel pair has a coequalizer;
3. The pullback of a regular epimorphism along any morphism exists and is a regular epimorphism.

This is the purely abstract mathematical structure corresponding to a regular theory $\mathbf{T}$, but the abstract notion was not abstracted from that construction. It has an independent mathematical existence. The definition does not show automatically how regular logic can be interpreted in a regular category or that a regular theory yields, as its category of concepts, a regular category. But of course, in both cases, it does.

In a structuralist framework, one has to specify the criterion of identity for the abstract structures given. Thus, we first have to specify what a *regular functor* between regular categories is. Of course, it is a functor that preserves the appropriate structure. In this particular case, a functor $F : \mathcal{C} \to \mathcal{D}$ between regular categories is *regular* if

[15] Again, we are being very selective, and our goal is not to be exhaustive. The picture is much more elaborate than what we are presenting here. This is but the tip of the iceberg.

[16] There are various equivalent definitions of regular categories in the literature. We are following (Borceux, 1994).

it preserves finite limits and regular epimorphisms. The criterion of identity for regular categories is given by the notion of *equivalence* of regular categories, that is by a pair of regular functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{C} \to \mathcal{D}$ such that $G \circ F \simeq 1_{\mathcal{C}}$ and $F \circ G \simeq 1_{\mathcal{D}}$.

In the same vein, corresponding to coherent theories, we have:

A *coherent* category $\mathcal{C}$ is a regular category such that

1. Every subobject meet semilattice $S(X)$ is a lattice;
2. Each $f^*\colon S(Y) \to S(X)$ is a lattice homomorphism.

A coherent functor between coherent categories is a functor preserving the coherent structure, and the criterion of identity for coherent categories is extracted from that context.

And, of course, we can add more structure and properties to get other abstract mathematical structures. Let us simply mention a few more abstract structures that are directly related to logic.

A *Heyting category* $\mathcal{C}$ is a coherent category in which each $f^*\colon S(Y) \to S(X)$ has a right adjoint, denoted by $\forall_f$. The last condition is sufficient to entail that each $S(X)$ is a Heyting algebra, that f is a homomorphism of Heyting algebras and that the right adjoint is also stable under substitution. Heyting categories are common: for any small category $\mathcal{P}$, the functor category $\mathbf{Set}^{\mathcal{P}}$ is a Heyting category. They correspond to theories in intuitionistic predicate logic. Heyting functors are defined in the expected manner.

A *Boolean category* $\mathcal{C}$ is a coherent category such that every $S(X)$ is a Boolean algebra, i.e., every subobject has a complement. Boolean functors between Boolean categories are functors preserving the Boolean structure.

A *pretopos* $\mathcal{C}$ is a coherent category having (1) quotients of equivalence relations and (2) finite disjoint sums. Pretopos functors can be defined.[17] The notion of pretopos occupies a central place in the picture, since many of the important theorems about logic can be hooked to that notion. Finally, we have to mention at this stage the notion of Grothendieck topos, which surprisingly sits right at the center of the development of first-order logic and its many variants.[18] There is no need to go on for our purposes.

Let us immediately point out that the category $\mathbf{Set}$ is regular, coherent, Heyting, Boolean, a pretopos and a Grothendieck topos. And it is even more than just those.

There is an interesting and immediate application of the above constructions. In a classical logical framework, the notion of a interpretation or translation of one category into another one is delicate and complicated. Once we move from a theory $\mathbf{T}$ to its category of concepts $[\mathbf{T}]$, there is a very simple and direct way to define it. Indeed, a structure-preserving functor $I\colon [\mathbf{T}] \to [\mathbf{T}']$ between (small) categories of concepts is called an *interpretation* of $[\mathbf{T}]$ in $[\mathbf{T}']$. (When $[\mathbf{T}]$ and $[\mathbf{T}']$ have been constructed from theories, one can verify that it is a legitimate notion of interpretation (see Makkai and Reyes, 1977, chap. 7, p.196).

The models of a theory as a category

We now have to consider how a theory $\mathbf{T}$ can be interpreted in a category $\mathcal{C}$. As it is often the case when facing such a situation, the

[17] As we have already mentioned, for many important results, it is enough to consider weaker functors between some of the categories involved, e.g. coherent functors, for they preserve, in these particular contexts, the additional structure.

[18] As already pointed out by Makkai & Reyes in (1977), Giraud's theorem can be interpreted as giving a *logical characterization* of the notion of a Grothendieck topos (see Makkai and Reyes, 1977, chapter 1, section 4).

easiest solution is to translate what one does in sets, but express it in the language of the category of sets and finally move to an arbitrary category with the adequate structure and properties. It can indeed be done and what we get is a genuine generalization of Tarski's notion of satisfaction or model. The classical notions of satisfaction, interpretation, model and truth transfer directly to this new context. However, instead of presenting the nuts and bolts of these definitions, we will jump immediately to the next step.

Since we have constructed $[\mathbf{T}]$ from $\mathbf{T}$, and since $[\mathbf{T}]$ is a category, we can look directly at the interpretations of the latter. We will illustrate the situation with a coherent theory $\mathbf{T}$, but starting with the (small) coherent category $[\mathbf{T}]$ constructed from it. A coherent functor $M\colon [\mathbf{T}] \to \mathbf{Set}$ is called a (set-)*model* of $[\mathbf{T}]$. Since $\mathbf{Set}$ is a coherent category, this makes sense.

It is natural to consider to category of all such models, that is the functor category $\mathrm{Mod}([\mathbf{T}], \mathbf{Set})$, the *category of all (set-)models* of $[\mathbf{T}]$. The objects of this category are the models of $[\mathbf{T}]$, that is coherent functors $M\colon [\mathbf{T}] \to \mathbf{Set}$, and the morphisms are the natural transformations between models $\eta\colon M_1 \to M_2$. These are the *homomorphisms* of models of $[\mathbf{T}]$ and they are the traditional model-theoretic structure-preserving functions between models. More generally, for any coherent category $\mathcal{C}$, the category $\mathrm{Mod}([\mathbf{T}], \mathcal{C})$ of models of $[\mathbf{T}]$ in $\mathcal{C}$ is defined in the same way. We therefore have a flexibility that was not available previously.

The category $\mathrm{Mod}([\mathbf{T}], \mathbf{Set})$ of models of $[\mathbf{T}]$ in $\mathbf{Set}$ *is* certainly an instance of an abstract mathematical structure.[19] It has, in fact, a lot of structure. It is, among other things, a Grothendieck topos, an important type of abstract mathematical structure.

[19] This is also true when we take category different from $\mathbf{Set}$. But then, the structure of the resulting category depends directly on the structure of $\mathcal{C}$.

A theory and its models: moving up the ladder

We have identified some of the abstract mathematical structures that arise from the traditional logical notions. We now have, on the one hand, abstract mathematical structures corresponding to what Lawvere referred to as the "formal", and, on the other hand, abstract mathematical structures corresponding to what Lawvere referred to as the "conceptual". Of course, these have to be connected and these connections constitute the core of classical logic.

These connections are themselves part of an abstract mathematical structure. Since $[\mathbf{T}]$ is a category—the formal side of mathematics—and $\mathrm{Mod}([\mathbf{T}], \mathbf{Set})$ is a category—the conceptual side of mathematics—, we can investigate the functors between them. But there is more. There are also functors between theories $[\mathbf{T}] \to [\mathbf{T}']$, functors between set-models of theories

$$\mathrm{Mod}([\mathbf{T}], \mathbf{Set}) \to \mathrm{Mod}([\mathbf{T}'], \mathbf{Set}),$$

functors between models of theories in different categories

$$\mathrm{Mod}([\mathbf{T}], \mathcal{C}) \to \mathrm{Mod}([\mathbf{T}], \mathcal{D}),$$

functors between all those and natural transformations between some of these functors![20] In fact, we are in a 2-category, which is a *genuinely new structure*. A 2-category is not merely a category with additional data. Thus, once again, we are in a realm of abstract mathematical structures and many of the results we are interested in will be consequences of this abstract mathematical structure *together with* some specific properties inherent in the situation we are dealing with.

[20] We are not being careful here. Some of these are covariant functors, while others are contravariant. We simply want to point at the possibilities at this juncture. We are not developing the theory as such. Again, these details do not affect our main point.

Here are some questions that can now be investigated. The main point here is that these questions make perfect sense, they are entirely natural, whereas it is hard to imagine how they could have arisen outside this mathematical context.

1. Given an interpretation $I\colon [\mathbf{T}] \to [\mathbf{T}']$ between theories, one can transfer models of $\mathbf{T}'$ to models of $\mathbf{T}$ by composing with I, that is given a model $M\colon \mathbf{T}' \to \mathbf{Set}$, we get by composition with I a model $M \circ I\colon \mathbf{T} \to \mathbf{Set}$. Hence, there is a functor $I^*\colon \mathrm{Mod}(\mathbf{T}', \mathbf{Set}) \to \mathrm{Mod}(\mathbf{T}, \mathbf{Set})$. The natural questions to ask pertain to the relations between I^* and I. More specifically, are there properties of I^* that imply properties of I? In particular, is it possible that I^* being an equivalence of categories imply that I is? In words, what are the properties of the conceptual that affect the properties of the formal?
2. Given a functor $F\colon \mathcal{C} \to \mathcal{D}$ of the right type (that preserves the right kind of structure in each case), we get a functor $F^*\colon \mathrm{Mod}([\mathbf{T}], \mathcal{C}) \to \mathrm{Mod}([\mathbf{T}], \mathcal{D})$ by composing models M with F. One question here focuses on the categories $\mathcal{C}$ and $\mathcal{D}$, more specifically on $\mathcal{C}$ and the abstract mathematical structure both these categories are instances of. Thus, is there an abstract mathematical type of structure such that all the models of $[\mathbf{T}]$ in $\mathcal{D}$ arise from models of $[\mathbf{T}]$ in $\mathcal{C}$ and functors $\mathcal{C} \to \mathcal{D}$?

Other questions can be formulated, but these are not unlike questions that arise in other mathematical domains, thus relating this formulation of logic with comparable frameworks. To be able to identify what is the common abstract core of logic with other mathematical domains and what is specific to logic is one of the gains of the abstract structuralist approach.

4. Metalogical theorems from an abstract structural standpoint

From a structuralist standpoint, once the abstract mathematical structures have been identified, one hopes to be able to prove standard theorems from that vantage. And, indeed, one can. One of the epistemic gains expected from these theorems is the identification of the abstract components involved in various proofs and thus see what is the core structural component upon which these results are grounded. Another expected benefit is the possibility to get genuinely new results which were impossible to get in the classical framework, even impossible to formulate adequately.

Completeness and conceptual completeness

Let us start with what can be considered the pillar of logic in general, namely completeness results. Completeness results for various propositional logics are equivalent to representation theorems for various algebras, e.g., in the case of classical propositional logic, the completeness theorem is equivalent to Stone's representation theorem for Boolean algebras. As we have already mentioned, that most natural context to prove this result is already the context of the category of Boolean algebras and the theorem is done up to isomorphism.

Moving to first-order logic, it is to be expected that the completeness theorems would amount to representation theorems for certain categories, e.g. regular, coherent, pretoposes, Heyting, Boolean, etc. Indeed, the classical (Gödel) completeness theorem is equivalent to a representation theorem for coherent categories, which can be stated thus: for any small coherent category C, there is a (small) set I and a conservative coherent functor $F\colon \mathcal{C} \to \mathbf{Set}^I$. A functor $F\colon \mathcal{C} \to \mathcal{D}$ is said to be *conservative* if it reflects isomorphisms, i.e., if $F(f)$ is an isomorphism in $\mathcal{D}$, then f was already an isomorphism in $\mathcal{C}$. Needless to say, the key property is precisely that of being conservative. For what it amounts to is the fact that for any diagram in $\mathcal{C}$ such that its image under F in $\mathcal{D}$ is a diagram of a universal morphism, then the original diagram was already a diagram of a universal morphism in $\mathcal{C}$.

As we have already mentioned, the category $\mathbf{Set}$ is coherent and so is the functor category $\mathbf{Set}^I$. Since the functor $F\colon \mathcal{C} \to \mathbf{Set}^I$ is conservative, it follows that $\mathcal{C}$ shares all the coherent properties of $\mathbf{Set}^I$, and in fact of $\mathbf{Set}$. The equivalence between the representation theorem and the completeness theorem can be established as follows. Assuming the representation theorem, we start with a coherent theory $\mathbf{T}$ and construct the category of concepts $[\mathbf{T}]$ of $\mathbf{T}$, which is a coherent category. Applying the representation theorem to $[\mathbf{T}]$, we obtain the completeness theorem. To prove the other direction, we assume the completeness theorem and start with a coherent category $\mathcal{C}$. Using the internal language of $\mathcal{C}$, one constructs as above the coherent theory $\mathbf{T}_{\mathcal{C}}$ of $\mathcal{C}$. The models of $\mathbf{T}_{\mathcal{C}}$ are then constructed so that they are identical with functors $\mathcal{C} \to \mathbf{Set}$. The representation theorem then follows from the completeness theorem for $\mathbf{T}_{\mathcal{C}}$.

Two important elements have to be added to the picture. First, the representation theorem for coherent categories is but one representation theorem for a whole collection of relevant categories: regular categories, pretoposes, Heyting categories and Boolean categories. Second, these results in fact follow a general pattern. Indeed, the foregoing representation theorem takes a general, purely categorical form, in other words, there is a crucial part that is purely based on the abstract mathematical structures. Given any categories $\mathcal{S}$ and $\mathcal{C}$, we can always consider the repeated functor category $\mathcal{S}^{(\mathcal{S}^{\mathcal{C}})}$.[21] In this situation, there is a canonical functor, the evaluation functor

$$e\colon \mathcal{C} \to \mathcal{S}^{(\mathcal{S}^{\mathcal{C}})}$$

for which, given any object X of $\mathcal{C}$, and any functor $F\colon \mathcal{C} \to \mathcal{S}$, $e(X)(F)$ is simply $F(X)$, the evaluation of F at X. For any subcategory $\mathcal{D}$ of $\mathcal{S}^{\mathcal{C}}$, the same functor $e\colon \mathcal{C} \to \mathcal{S}^{\mathcal{D}}$ can be defined. It is then possible to show that the representation theorem for coherent categories is equivalent to the claim that the functor $e\colon \mathcal{C} \to \mathcal{S}^{\mathrm{Mod}(\mathcal{C})}$ is conservative. The fact that the evaluation functor is coherent holds on purely general grounds. We therefore have a purely categorical description of the representation theorem. Moreover, in the early seventies Joyal demonstrated that the functor e preserves all existing instances of the Heyting structure in $\mathcal{C}$. This automatically yields a representation theorem for Heyting categories and, in turn, a canonical completeness theorem for intuitionistic logic.

The categorical set-up allows is to consider a stronger claim, called the *conceptual completeness*. Given a functor $I : \mathbf{T} \to \mathbf{T}'$ and an equivalence of categories between $\mathrm{Mod}(\mathbf{T}', \mathbf{Set})$ and $\mathrm{Mod}(\mathbf{T}, \mathbf{Set})$, when is it possible to conclude that I is also an equivalence of categories? From a categorical point of view, the assumption means that the category of models of $\mathbf{T}'$ is indistinguishable from the category of models of $\mathbf{T}$. We can think of the functor I as a translation of $[\mathbf{T}]$ into $[\mathbf{T}']$, thus as a case when the latter theory can in principle be more expressive than the former. In a sense, the conceptual completeness can be interpreted as saying that adding new concepts to $\mathbf{T}$ simply does not modify in any essential way what it can express. This means that $\mathbf{T}$ has some sort of completeness and in this

[21] This is not an unusual construction in mathematics. Think of the double dual of a finite-dimensional vector space, for instance.

context it makes perfect sense to say that it is *conceptually complete*. Thus, we say that $\mathbf{T}$ is *conceptually complete* whenever the following is satisfied: if the functor $I^*\colon \mathrm{Mod}(\mathbf{T}', \mathbf{Set}) \to \mathrm{Mod}(\mathbf{T}, \mathbf{Set})$ is an equivalence of categories, then the functor $I\colon \mathbf{T} \to \mathbf{T}'$ was one already. This literally means that by moving to $\mathbf{T}'$, we did not add anything essentially new to $\mathbf{T}$, although we might have thought we had, and this information was obtained by looking at the *categorical structure of the category of models of the theories*. We can conclude that a certain logical framework, say an equational theory, is enough to characterize a type of structures, from the categorical structure of the category of models. Conceptual completeness is in fact equivalent to a standard result of model theory, namely Beth definability theorem. However, one of the advantages of working in the categorical framework is that categorical methods make it possible to prove results which might not be accessible otherwise, for instance, a constructive proof of this result for intuitionistic logic.[22]

It is possible to strengthen the conceptual completeness theorem. In the latter, we assume as given a functor $I : \mathbf{T} \to \mathbf{T}'$. Is it possible to start with an equivalence of categories $\mathrm{Mod}(\mathbf{T}, \mathbf{Set}) \to \mathrm{Mod}(\mathbf{T}', \mathbf{Set})$ and construct from it an equivalence $\mathbf{T} \to \mathbf{T}'$? This is much stronger theorem, but it can be proved under certain circumstances. It says that a logical theory is completely characterized by the categorical structure of its category of models. In some sense, the conceptual determines the formal, up to equivalence. In the case of propositional logic, it amounts to a form of Stone duality, the latter being formulated entirely *within* the category of Boolean algebras, and not as the existence of an equivalence of categories between the category of Boolean algebras and the category of Stone spaces. The

[22] See (Pitts, 1989) for a categorical proof of conceptual completeness of intuitionistic first-order logic.

strong conceptual completeness asserts that the Lindenbaum-Tarski algebra of a propositional theory can be recovered from its space of models—the ultrafilters on the given Boolean algebra. A theory for which the theorem can be proved is said to be *strongly conceptually complete*. A different way to formulate this result is to say that if $\mathrm{Mod}(\mathbf{T}, \mathbf{Set})$ and $\mathrm{Mod}(\mathbf{T}', \mathbf{Set})$ are equivalent, then $\mathbf{T}$ and $\mathbf{T}'$ are equivalent too. Whereas conceptual completeness is a local phenomenon, since it depends on the interpretation I, strong conceptual completeness is a global phenomenon, since there is no underlying interpretation at hand. The construction of $\mathbf{T}$ can be thought of as a case of abstracting certain data out of another, more "concrete", situation. Finite limit categories of concepts are strongly conceptually complete[23], although the original result applied to (Boolean) pretoposes. If the category of models is adequately enriched in a precise technical sense, then in these circumstances first-order classical logic is strongly conceptually complete (see Makkai, 1988; 1990; for a a different proof which is build with higher-dimensional categories in mind, see Lurie, 2019). Notice that it is hard to see how this theorem could even be formulated outside the context of category theory.

It is impossible not to mention the fact that strong conceptual completeness theorems are closely related to dualities. In fact, they are equivalent in a precise technical sense to dualities. The only thing we want to underline is that these results are proved in the context of 2-categories. Thus, it is not only that the natural set-up involves 2-categories, but that important theorems require 2-categorical (even bicategorical) concepts. We cannot, in such a short paper, present these in any comprehensible manner.

[23] Thus they are the so-called Barr-exact categories. A Barr-exact category is a regular category in which every equivalence relation is a kernel pair (see Makkai, 1990).

Syntax and abstract completeness

From the above considerations, the reader might feel that we have entirely left behind syntactical considerations, more specifically formal deductions. Therefore, it might seem like the categorical completeness results are not quite the same as the classical results which assert that semantical consequences of a theory are provable in a fully specified formal system. This is not the case. For one thing, we have not abandoned the syntax, nor the formal systems in these investigations. But there is an additional point to make, for it brings to the fore a way of dealing with the syntax of theories that emerged naturally from the context of categories, namely the idea of a sketch and its generalizations.

From Lawvere's thesis, category theorists toyed with the idea that a theory could be presented directly in the form of a category or some graphical variant thereof. In this spirit, sketches were introduced by Charles Ehresmann in the early 1960s and developed afterwards by him and his school (see Ehresmann, 1967; 1968; Lair, 2001; 2002; 2003, for instance). A sketch, which is a specific kind of (oriented) graph, is a new kind of syntax, specifically tailored to do categorical logic. It gives directly in a graphical way the syntactical and proof theoretical content of a theory.

We will give one definition of the notion of sketch[24]. A sketch $\mathcal{S} = (G, D, L, C)$ is given by a graph G, a set D of diagrams in G, a set L of cones in G and a set C of cocones in G. We can consider the category of sketches by stipulating that a morphism of sketches is a homomorphism of graphs which preserves the diagrams, the cones and the cocones. It is easy to see that any category $\mathbb{C}$ has an underlying sketch $\mathcal{S}_{\mathbb{C}}$. A *model* of a sketch $\mathcal{S}$ in a category $\mathbb{C}$ takes all the

[24] See (Barr and Wells, 1990; 2005) for an introduction to sketches.

diagrams of $\mathcal{S}$ to commutative diagrams, all the cones of $\mathcal{S}$ to limits of $\mathbb{C}$ and all cocones of $\mathcal{S}$ to colimits of $\mathbb{C}$. A morphism of models is a natural transformation. Thus, we can reproduce what we did above with theories, namely we can construct the category of models $\mathrm{Mod}(\mathcal{S}, \mathbb{C})$ of a sketch $\mathcal{S}$ in a category $\mathbb{C}$.

It is natural to consider sketches in which there are no cocones and only discrete and finite cones, or in which there are no cocones and only finite cones, etc. Sketches organized themselves with respect to these natural choices and they correspond to various logical theories. Thus, there are finite product sketches (a FP-sketch), left exact sketches (a LE-sketch), regular sketches, coherent sketches, etc.

In this framework, it is natural to ask which categories are sketchables: is it possible to characterize categories that are equivalent to categories of models of a type of sketch? There are positive answers to that question and it naturally brings us, when the most general kinds of sketches are considered, to infinitary logic $L_{\infty,\infty}$ (see Lair, 1981; Makkai and Paré, 1989).

Generalization of the notion of sketch has led Makkai to develop a categorical proof theory and establish a completeness theorem along the classical lines, that is proving that a formula is formally provable in a theory if and only if it is true in all models of the theory. Interestingly enough, the set-up still rests upon the categorical representation theorems, but it is enriched with a categorical notion of formal proof in the set-up of (generalized) sketches. (see Makkai, 1997a,b,c).

Incompleteness

We have to say a few words about Gödel's incompleteness theorems. Is it possible to identify abstract mathematical structures that underly these theorems? Is it possible to deduce these theorems from a theorem or theorems about these abstract mathematical structures? There are some pieces in place, although the complete picture—no pun intended—has still to be presented.

First, already in the 1960s, Lawvere presented a categorical analysis of various phenomena related to Gödels's incompleteness theorems. In his (1969b), Lawvere presents what he takes to be the abstract mathematical structure underlying Cantor's theorem that there is no surjection $\mathbb{X} \to 2^{\mathbb{X}}$ and its variants in the heads of Russell, Gödel and Tarski. The starting point here is the notion of a cartesian closed category. A cartesian closed category is a nice example of a categorical doctrine since it can be given entirely by stipulating the existence of certain adjoint functors to elementary, that is first-order, functors. More precisely, a cartesian closed category $\mathbb{C}$ is a category such that

1. The functor $! : \mathbb{C} \to 1$ has a right adjoint;
2. The diagonal functor $\Delta : \mathbb{C} \to \mathbb{C} \times \mathbb{C}$ has a right adjoint, namely the product functor;
3. For each object X of $\mathbb{C}$, the functor $X \times (-) : \mathbb{C} \to \mathbb{C}$ has a right adjoint $(-)^X$.

Given a cartesian closed category $\mathbb{C}$, it is possible to then define what D. Pavlovic has called a "paradoxical structure" on an object of $\mathbb{C}$ that satisfies a fixed-point property. By specifying the adequate cartesian closed category $\mathbb{C}$ and the paradoxical structure on objects of $\mathbb{C}$, it is possible to prove Cantor's theorem, Russell's paradox, Gödel's first incompleteness theorem, Tarski's theorem of the impossibility of defining truth in a theory and many others (see Pavlović, 1992 and Yanofsky, 2003).

In a series of unpublished lectures presented in the 1970s, André Joyal introduced another abstract mathematical structure, in a precise sense weaker than Lawvere's proposal, to pursue the analysis of Gödel's incompleteness results, in particular the second theorem, namely what he called 'arithmetical universes'. Very roughly, an arithmetic universe $\mathbb{U}$ is a pretopos such that the free category object constructed from a graph object in $\mathbb{U}$ exists. This is the abstract structure in which one can do recursive arithmetic and prove versions of the two incompleteness theorems (see Maietti, 2010 for a different definition).

Type theories

We will be very succinct, not because this area is not important, quite the contrary, but simply because there is no need to cover everything in details given the goal of this paper.

The first and well-known result in this area is the correspondence between cartesian closed categories with a natural number object and typed λ-calculus. We finally get to elementary toposes. These were introduced by Lawvere and Tierney in 1970 to provide an elementary treatment of sheaves over a site, thus of Grothendieck toposes. It is another remarkable example of a categorical doctrine. An *elementary topos* $\mathcal{E}$ is a category with finite limits, cartesian closed, and has a subobject classifier. These three conditions do amount to the existence of certain adjoint functors to given (elementary) functors. As is well known now, it is possible to construct an intuitionistic type theory from a given topos $\mathcal{E}$ and, conversely, it is possible to specify an intuitionistic type theory such that its conceptual category is an elementary topos and it can be interpreted in an elementary topos. There is then a correspondance between categorical properties of the topos and logical properties of the type theory (see Boileau and Joyal, 1981; Lambek and Scott, 1988).

The same can be said about homotopy type theory. Homotopy type theory comes form Martin-Löf's intensional type theory (see Program, 2013). It has models in various categories, but a homotopy type theory ought to correspond to a kind of abstract categories. It has been conjectured, by Steve Awodey, that homotopy type theory should correspond to the internal logic of higher-dimensional elementary toposes. As of this writing, the full conjecture has still to be proved, although certain advances have been made (see Kapulkin and Lumsdaine, 2018).

One last thing...

Last but not least, connections between linear logic and category theory appeared almost immediately after the creation of linear logic by Girard in (1987) (see Lafont, 1988; Seely, 1989). It took almost twenty years of research before a consensus emerged as to what constitutes a categorical model of linear logic (see Bierman, 1995; Blute and Scott, 2004; Melliès, 2009; de Paiva, 2014). We will not introduce nor discuss the categorical framework here. It would require defining and explaning a lot of categorical structures, e.g. symmetric monoidal categories, symmetric monoidal adjunctions, etc., as well as an explanation of how the various frameworks proposed converge towards a basic structure. The point is: we may be seeing the beginning of a stable picture that will allow us to start building a conceptual interpretation of linear logic. In as much as homotopy type theory seems to be intimately connected to the basic constituents of spaces, the "atoms of space" to use Baues's expression in (2002), namely homotopy types, linear logic seems to be intimately tied to generalized vector spaces and the mathematics inherent to the latter (see Melliès, n.d.). If this reading is correct, it may lead to new interpretations and developments of conceptual spaces and the categorical structures would naturally find their place in that context. But this is sheer speculation at this point and it does not affect our main point.

5. Conclusion

We insist that this way of framing logic and metalogic is a direct continuation of mathematical logic as it developed in the first half of the 20th century and the rise of the abstract axiomatic method at the same time. Category theory itself is an offspring of this period, and as such, does not constitute a radical methodological change. It is, undoubtedly, a rise in abstraction. It reveals new types of abstract mathematical structures.

We hope we have convinced the reader that it is possible to identify the abstract mathematical structures underlying (fragments of, and extensions of) first-order logic and type theories. It is also possible to see how the important metalogical results correspond to theorems on these abstract mathematical structures. Finally, the invariance property at the core of any abstract mathematical structuralism comes naturally and automatically in this framework. Thus, the standard logical systems—and some non-standard logical systems as well—find a natural place in this structuralist context. Pure logic is seen as a specific type of abstract structure. We can thus answer our own challenge positively and precisely. We leave to other structuralists to provide their answer to our challenge.

Bibliography

Awodey, S., 1996. Structure in mathematics and logic: a categorical perspective. *Philos. Math. (3)* [Online], 4(3), pp.209–237. Available at: https://doi.org/10.1093/philmat/4.3.209.

Barr, M. and Wells, C., 1990. *Category Theory for Computing Science* [Online], *Prentice Hall International Series in Computer Science*. Prentice Hall International, New York, pp.xvi+432. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=1094561>.

Barr, M. and Wells, C., 2005. Toposes, triples and theories. *Repr. Theory Appl. Categ.* [Online], (12), pp.x+288. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=2178101>.

Baues, H.-J., 2002. Atoms of topology. *Jahresber. Deutsch. Math.-Verein.*, 104(4), pp.147–164.

Bierman, G.M., 1995. What is a categorical model of intuitionistic linear logic? *Typed Lambda Calculi and Applications (Edinburgh, 1995)* [Online]. Vol. 902, *Lecture Notes in Comput. Sci.* Springer, Berlin, pp.78–93. Available at: https://doi.org/10.1007/BFb0014046.

Birkhoff, G., 1940. *Lattice Theory* [Online]. American Mathematical Society, New York, pp.v+155. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0001959>.

Blute, R. and Scott, P., 2004. Category theory for linear logicians. *Linear Logic in Computer Science* [Online]. Vol. 316, *London Math. Soc. Lecture Note Ser.* Cambridge Univ. Press, Cambridge, pp.3–64. Available at: https://doi.org/10.1017/CBO9780511550850.002.

Boileau, A. and Joyal, A., 1981. La logique des topos. *J. Symbolic Logic* [Online], 46(1), pp.6–16. Available at: https://doi.org/10.2307/2273251.

Borceux, F., 1994. *Handbook of Categorical Algebra. 2* [Online]. Vol. 51, *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, pp.xviii+443. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=1313497>.

Corry, L., 2004. *Modern Algebra and the Rise of Mathematical Structures*. Second. Birkhäuser Verlag, Basel, pp.xvi+451.

Ehresmann, C., 1967. Sur les structures algébriques. *C. R. Acad. Sci. Paris Sér. A-B* [Online], 264, A840–A843. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=217140>.

Ehresmann, C., 1968. Esquisses et types des structures algébriques. *Bul. Inst. Politehn. Iaşi (N.S.)* [Online], 14(18)(fasc., fasc. 1-2), pp.1–14. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=238918>.

Eilenberg, S. and Mac Lane, S., 1945. A General Theory of Natural Equivalences. *Trans. Amer. Math. Soc.*, 58, pp.231–294.

Freyd, P., 1966. Representations in Abelian Categories. *Proceedings of the Conference on Categorical Algebra, La Jolla, Calif., 1965*. New York: Springer-Verlag, pp.95–120.

Girard, J.-Y., 1987. Linear logic. *Theoret. Comput. Sci.* [Online], 50(1), p.101. Available at: https://doi.org/10.1016/0304-3975(87)90045-4.

Halmos, P.R., 1956a. Algebraic logic. II. Homogeneous locally finite polyadic Boolean algebras of infinite degree. *Fund. Math.* [Online], 43, pp.255–325. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=86029>.

Halmos, P.R., 1954. Polyadic Boolean algebras. *Proc. Nat. Acad. Sci. U.S.A.* [Online], 40, pp.296–301. Available at: https://doi.org/10.1073/pnas.40.5.296.

Halmos, P.R., 1956b. Algebraic logic. I. Monadic Boolean algebras. *Compositio Math.* [Online], 12, pp.217–249. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=78304>.

Halmos, P.R., 1956c. Algebraic logic. III. Predicates, terms, and operations in polyadic algebras. *Trans. Amer. Math. Soc.* [Online], 83, pp.430–470. Available at: https://doi.org/10.2307/1992883.

Halmos, P.R., 1956d. The basic concepts of algebraic logic. *Amer. Math. Monthly* [Online], 63, pp.363–387. Available at: https://doi.org/10.2307/2309396.

Halmos, P.R., 1962. *Algebraic Logic* [Online]. Chelsea Publishing Co., New York, p.271. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0131961>.

Hellman, G., 2001. Three varieties of mathematical structuralism. *Philos. Math. (3)* [Online], 9(2), pp.184–211. Available at: https://doi.org/10.1093/philmat/9.2.184.

Hellman, G., 2003. Does category theory provide a framework for mathematical structuralism? *Philos. Math. (3)* [Online], 11(2), pp.129–157. Available at: https://doi.org/10.1093/philmat/11.2.129.

Hellman, G. and Shapiro, S., 2019. *Mathematical Structuralism* [Online], *Elements in the Philosophy of Mathematics*. Cambridge University Press, Cambridge, p.92. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=3967722>.

Henkin, L., Monk, J.D. and Tarski, A., 1971. *Cylindric Algebras. Part I. With an Introductory chapter: General Theory of Algebras* [Online]. North-Holland Publishing Co., Amsterdam-London, pp.vi+508. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0314620>.

Henkin, L. and Tarski, A., 1961. Cylindric algebras. *Proc. Sympos. Pure Math., Vol. II* [Online]. American Mathematical Society, Providence, R.I., pp.83–113. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0124250>.

Johnstone, P.T., 2002. *Sketches of an Elephant: A Topos Theory Compendium. Vol. 1* [Online]. Vol. 43, *Oxford Logic Guides*. The Clarendon Press, Oxford University Press, New York, pp.xxii+468+71. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=1953060>.

Jónsson, B. and Tarski, A., 1951. Boolean algebras with operators. I. *Amer. J. Math.* [Online], 73, pp.891–939. Available at: https://doi.org/10.2307/2372123.

Jónsson, B. and Tarski, A., 1952. Boolean algebras with operators. II. *Amer. J. Math.* [Online], 74, pp.127–162. Available at: https://doi.org/10.2307/2372074.

Kapulkin, K. and Lumsdaine, P.L., 2018. The homotopy theory of type theories. *Adv. Math.* [Online], 337, pp.1–38. Available at: https://doi.org/10.1016/j.aim.2018.08.003.

Krömer, R., 2007. *Tool and Object* [Online]. Vol. 32, *Science Networks. Historical Studies*. Birkhäuser Verlag, Basel, pp.xxxvi+367. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=2272843>.

Lafont, Y., 1988. The linear abstract machine. Vol. 59, 1-2, pp.157–180. Available at: https://doi.org/10.1016/0304-3975(88)90100-4.

Lair, C., 1981. Catégories modelables et catégories esquissables. *Diagrammes* [Online], 6, pp.L1–L20. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=684535>.

Lair, C., 2001. Éléments de théorie des esquisses. I. Graphes à composition. *Diagrammes* [Online], 45/46, pp.3–33. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=2102295>.

Lair, C., 2002. Éléments de théorie des esquisses. II. Systèmes tensoriels et systèmes enrichis de graphes à composition. *Diagrammes* [Online], 47/48, p.34. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=2104013>.

Lair, C., 2003. Éléments de théorie des esquisses. III. Esquisses. *Diagrammes* [Online], 49/50, p.58. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=2104014>.

Lambek, J., 1968a. A Fixpoint Theorem for Complete Categories. *Math. Z.*, 103, pp.151–161.

Lambek, J., 1968b. Deductive Systems and Categories. I. Syntactic Calculus and Residuated Categories. *Math. Systems Theory*, 2, pp.287–318.

Lambek, J., 1969. Deductive Systems and Categories II. Standard Constructions and Closed Categories. In: P.J. Hilton, ed. *Category Theory, Homology Theory, and their Applications, I.* Vol. 86, Lecture Notes in Mathematics. Berlin: Springer-Verlag, pp.76–122.

Lambek, J., 1972. Deductive Systems and Categories III. Cartesian Closed Categories, Intuitionistic Propositional Calculus, and Combinatory

Logic. In: F.W. Lawvere, ed. *Toposes, Algebraic Geometry and Logic*. Vol. 274, Lecture Notes in Mathematics. Berlin: Springer-Verlag, pp.57–82.

Lambek, J. and Scott, P.J., 1988. *Introduction to Higher Order Categorical Logic* [Online]. Vol. 7, *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, pp.x+293. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=939612>.

Landry, E. and Marquis, J.-P., 2005. Categories in Context: Historical, Foundational and Philosophical. *Philos. Math. (3)*, 13(1), pp.1–43.

Lawvere, F.W., 1966. Functorial Semantics of Elementary Theories. *Journal of Symbolic Logic*, 31, pp.294–295.

Lawvere, F.W., 1967. Theories as Categories and the Completeness Theorem. *Journal of Symbolic Logic*, 32, p.562.

Lawvere, F.W., 1969a. Adjointness in Foundations. *Dialectica*, 23, pp.281–296.

Lawvere, F.W., 1971. Quantifiers and sheaves. *Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 1* [Online], pp.329–334. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0430021>.

Lawvere, F.W., 1963. Functorial semantics of algebraic theories. *Proc. Nat. Acad. Sci. U.S.A.* [Online], 50, pp.869–872. Available at: https://doi.org/10.1073/pnas.50.5.869.

Lawvere, F.W., 1969b. Diagonal arguments and cartesian closed categories. *Category Theory, Homology Theory and their Applications, II (Battelle Institute Conference, Seattle, Wash., 1968, Vol. Two)* [Online]. Springer, Berlin, pp.134–145. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0242748>.

Lawvere, F.W., 1970. Equality in hyperdoctrines and comprehension schema as an adjoint functor. *Applications of Categorical Algebra (Proc. Sympos. Pure Math., Vol. XVII, New York, 1968)* [Online]. Amer. Math. Soc., Providence, R.I., pp.1–14. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0257175>.

Linton, F.E.J., 1966. Some aspects of equational categories. *Proc. Conf. Categorical Algebra (La Jolla, Calif., 1965)*. New York: Springer, pp.84–94.

Lurie, J., 2019. *Ultracategories*. **unpublished**.

Mac Lane, S., 1998. *Categories for the Working Mathematician*. 2nd ed. Vol. 5, *Graduate Texts in Mathematics*. New York: Springer-Verlag.

Mac Lane, S. and Moerdijk, I., 1994. *Sheaves in Geometry and Logic* [Online], *Universitext*. Springer-Verlag, New York, pp.xii+629. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=1300636>.

Mac Lane, S., 1965. Categorical Algebra. *Bull. Amer. Math. Soc.*, 71, pp.40–106.

Maietti, M.E., 2010. Joyal's arithmetic universe as list-arithmetic pretopos. *Theory Appl. Categ.* [Online], 24, No. 3, 39–83. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=2610177>.

Makkai, M., 1997a. Generalized sketches as a framework for completeness theorems. I. *J. Pure Appl. Algebra* [Online], 115(1), pp.49–79. Available at: https://doi.org/10.1016/S0022-4049(96)00007-2.

Makkai, M., 1997b. Generalized sketches as a framework for completeness theorems. II. *J. Pure Appl. Algebra* [Online], 115(2), pp.179–212. Available at: https://doi.org/10.1016/S0022-4049(96)00008-4.

Makkai, M., 1997c. Generalized sketches as a framework for completeness theorems. III. *J. Pure Appl. Algebra* [Online], 115(3), pp.241–274. Available at: https://doi.org/10.1016/S0022-4049(96)00009-6.

Makkai, M., 1988. Strong conceptual completeness for first-order logic. *Ann. Pure Appl. Logic* [Online], 40(2), pp.167–215. Available at: https://doi.org/10.1016/0168-0072(88)90019-X.

Makkai, M., 1990. A theorem on Barr-exact categories, with an infinitary generalization. *Ann. Pure Appl. Logic* [Online], 47(3), pp.225–268. Available at: https://doi.org/10.1016/0168-0072(90)90036-2.

Makkai, M. and Paré, R., 1989. *Accessible Categories: The Foundations of Categorical Model Theory* [Online]. Vol. 104, *Contemporary Mathematics*. American Mathematical Society, Providence, RI, pp.viii+176. Available at: https://doi.org/10.1090/conm/104.

Makkai, M. and Reyes, G.E., 1977. *First Order Categorical Logic* [Online], *Lecture Notes in Mathematics, Vol. 611*. Springer-Verlag, Berlin-New York, pp.viii+301. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0505486>.

Marquis, J.-P., 2009. *From a Geometrical Point of View* [Online]. Vol. 14, *Logic, Epistemology, and the Unity of Science*. Springer, Dordrecht, pp.x+309. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=2730089>.

Marquis, J.-P., 2015. Mathematical abstraction, conceptual variation and identity. In: P.-E.B.G.H.W.H.P. Schroeder-Heister, ed. *Logic, Methodology and Philosophy of Science, Proceedings of the Fourteen International Congress*. London: College Publications, pp.299–322.

Marquis, J.-P., 2016. Stairway to Heaven: The Abstract Method and Levels of Abstraction in Mathematics. *The Mathematical Intelligencer*, 38(3), pp.41–51.

Marquis, J.-P. and Reyes, G.E., 2012. The history of categorical logic: 1963–1977. *Sets and Extensions in the Twentieth Century* [Online]. Vol. 6, *Handb. Hist. Log.* Elsevier/North-Holland, Amsterdam, pp.689–800. Available at: https://doi.org/10.1016/B978-0-444-51621-3.50010-4.

McKinsey, J.C.C. and Tarski, A., 1944. The algebra of topology. *Ann. of Math. (2)* [Online], 45, pp.141–191. Available at: https://doi.org/10.2307/1969080.

McKinsey, J.C.C. and Tarski, A., 1946. On closed elements in closure algebras. *Ann. of Math. (2)* [Online], 47, pp.122–162. Available at: https://doi.org/10.2307/1969038.

McKinsey, J.C.C. and Tarski, A., 1948. Some theorems about the sentential calculi of Lewis and Heyting. *J. Symbolic Logic* [Online], 13, pp.1–15. Available at: https://doi.org/10.2307/2268135.

Melliès, P.A., n.d. *A Functorial Excursion Between Algebraic Geometry and Linear Logic*.

Melliès, P.-A., 2009. Categorical semantics of linear logic. In: P.-L. Curien, H. Herbelin, J.-L. Krivine and P.-A. Méllies, eds. *Interactive models of computation and program behaviour*, *Panoramas et Synthèses* 27. Société Mathématique de France, pp.1–196.

Mostowski, A., 1949. Sur l’interpretation géométrique et topologique des notions logiques. *Library of the Tenth International Congress of Philosophy, Amsterdam, August 11–18, 1948, Vol. I, Proceedings of the Congress* [Online]. publisher unknown, pp.767–769. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=0028277>.

de Paiva, V., 2014. Categorical semantics of linear logic for all. *Advances in Natural Deduction* [Online]. Vol. 39, *Trends Log. Stud. Log. Libr.* Springer, Dordrecht, pp.181–192. Available at: https://doi.org/10.1007/978-94-007-7548-0_9.

Pavlović, D., 1992. On the structure of paradoxes. *Arch. Math. Logic* [Online], 31(6), pp.397–406. Available at: https://doi.org/10.1007/BF01277482.

Pitts, A.M., 1989. Conceptual completeness for first-order intuitionistic logic: an application of categorical logic. *Ann. Pure Appl. Logic* [Online], 41(1), pp.33–81. Available at: https://doi.org/10.1016/0168-0072(89)90007-9.

Program, T.U.F., 2013. *Homotopy Type Theory—Univalent Foundations of Mathematics* [Online]. The Univalent Foundations Program, Princeton, NJ; Institute for Advanced Study (IAS), Princeton, NJ, pp.xiv+589. Available at: <https://mathscinet.ams.org/mathscinet-getitem?mr=3204653>.

Rasiowa, H., 1951. Algebraic treatment of the functional calculi of Heyting and Lewis. *Fund. Math.* [Online], 38, pp.99–126. Available at: https://doi.org/10.4064/fm-38-1-99-126.

Rasiowa, H., 1955. Algebraic models of axiomatic theories. *Fund. Math.* [Online], 41, pp.291–310. Available at: https://doi.org/10.4064/fm-41-2-291-310.

Rasiowa, H. and Sikorski, R., 1950. A proof of the completeness theorem of Gödel. *Fund. Math.* [Online], 37, pp.193–200. Available at: https://doi.org/10.4064/fm-37-1-193-200.

Rasiowa, H. and Sikorski, R., 1953. Algebraic treatment of the notion of satisfiability. *Fund. Math.* [Online], 40, pp.62–95. Available at: https://doi.org/10.4064/fm-40-1-62-95.

Rasiowa, H. and Sikorski, R., 1955. An application of lattices to logic. *Fund. Math.* [Online], 42, pp.83–100. Available at: https://doi.org/10.4064/fm-42-1-83-100.

Rasiowa, H. and Sikorski, R., 1963. *The Mathematics of Metamathematics* [Online], *Monografie Matematyczne, Tom 41*. Państwowe Wydawnictwo Naukowe, Warsaw, p.522. Available at: <https : / / mathscinet.ams.org/mathscinet-getitem?mr=0163850>.

Reck, E.H. and Schiemer, G., eds., 2020. *The Prehistory of Mathematical Structuralism*. Oxford: Oxford University Press.

Riehl, E., 2017. *Category Theory in Context*. Mineola: Dover Publications.

Rodin, A., 2014. *Axiomatic Method and Category Theory* [Online]. Vol. 364, *Synthese Library. Studies in Epistemology, Logic, Methodology, and Philosophy of Science*. Springer, Cham, pp.xii+285. Available at: https://doi.org/10.1007/978-3-319-00404-4.

Seely, R.A.G., 1989. Linear logic, $*$-autonomous categories and cofree coalgebras. *Categories in Computer Science and Logic (Boulder, CO, 1987)* [Online]. Vol. 92, *Contemp. Math.* Amer. Math. Soc., Providence, RI, pp.371–382. Available at: https://doi.org/10.1090/conm/092/1003210.

Yanofsky, N.S., 2003. A universal approach to self-referential paradoxes, incompleteness and fixed points. *Bull. Symbolic Logic* [Online], 9(3), pp.362–386. Available at: https://doi.org/10.2178/bsl/1058448677.

Creating new concepts in mathematics: freedom and limitations. The case of Category Theory

Zbigniew Semadeni
Institute of Mathematics, University of Warsaw

Abstract

The purpose of the paper is to discuss the problem of possible limitations of freedom in mathematics and to look for criteria which would help us to distinguish—in the historical development of mathematics—new concepts which were natural follow-ups of the previous ones from new concepts which opened unexpected ways of thought and reasoning.

The rise of category theory (CT) is analysed, in particular, earlier ideas (which were precursors of the theory) and its initial development. The question of the the origin of the term *functor* is discussed; the presented evidence strongly suggests that Eilenberg and Carnap could have learned the term from Kotarbiński and Tarski.

Keywords

categories, functors, Eilenberg-Mac Lane Program, mathematical cognitive transgressions, phylogeny, platonism.

Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)
No 69 (2020), pp. 79–111

1. Introduction

The celebrated dictum of Georg Cantor that "The very essence of mathematics lies precisely in its freedom" expressed the idea that in mathematics one can freely introduce new notions (which may, however, be abandoned if found unfruitful or inconvenient).[1] This way Cantor declared his opposition to claims of Leopold Kronecker who objected to the free introduction of new notions (particularly those related to the infinite).

Some years earlier Richard Dedekind stated that—by forming, in his theory, a cut for an irrational number—we *create* a new number. For him this was an example of a constructed notion which was a free creation of the human mind (Dedekind, 1872, § 4).

In 1910 Jan Łukasiewicz distinguished *constructive notions* from empirical *reconstructive* ones. He referred (with reservation) to Dedekind's statement and pointed out that a consequence of our "creation" of those notions is the spontaneous emergence of countless relations which no more depend on our will.

Until the discovery of non-Euclidean geometries, geometry was regarded as an abstraction of the spatial reality. The freedom of creation in geometry was limited by this reality. Hilbert advocated a formal point of view, broadening the freedom of choosing the axioms, while Poincaré maintained that the axioms of geometry are conventions.

Clearly, the freedom of mathematics is limited by logical constraints. At the same time, logical inference yields deep meaning to mathematics. As Michał Heller put it, "If I accept one sentence, I

[1] The italics in the original sentence "Das *Wesen der Mathematik* liegt gerade in ihrer *Freiheit*" are Cantor's. The first version was published in 1879, reprinted in (Cantor, 1883, p.34), discussed by Ferreirós Domínguez (1999, p.257).

must also accept another sentence. Why must I? Who forces me? Nobody. Yet, I must. Generally, we bear badly any restrictions of our liberty, but in the case of mathematical deduction inevitability of the conclusion gives us the feeling of safety (I have not deviated from the way) and of the accompanying intellectual comfort, sometimes even great joy" (Heller, 2015, p.21).

A mathematician trying to prove a theorem knows the feeling of an invisible wall which blocks some intended arguments. Also new concepts must be consistent with earlier ones and must not lead to contradiction or ambiguity. Moreover, in mathematical practice, only intersubjective mental constructions are accepted.

The purpose of this paper[2] is to look for restraints and patterns in the historical development of mathematics.[3] Some types of paths will be distinguished, first generally, in the context of the historical development of mathematics, and then they will be used to highlight some features of the rise of category theory (CT).

Michael Atiyah, in his *Fields Lecture* at the World Mathematical Year 2000 in Toronto, expressed his view that "it is very hard to put oneself back in the position of what it was like in 1900 to be a mathematician, because so much of the mathematics of the last century has been absorbed by our culture, by us. It is very hard to imagine a time when people did not think in those terms. In fact, if you make a really important discovery in mathematics, you will then get omitted altogether! You simply get absorbed into the background" (Atiyah, 2002, p.1).

---

[2] The present paper is based on a talk delivered at XXIII Kraków Methodological Conference 2019: *Is Logic a Physical Variable?*, 7-9 November 2019.

[3] The significant question of degree of freedom in mathematical conceptualization of physical reality is not considered here.

This statement may be appear startling, as the mathematical meaning of a text from around 1900 is generally believed to be time-proof. Yet what Atiyah had in mind was not the meaning of published texts—definitions, theorems and proofs— but the way mathematicians thought at that time.

It was 75 years ago when the celebrated paper by Eilenberg and Mac Lane was published.[4] This event marked the rise of category theory (CT). As the development of mathematics accelerated in the second half of the 20$^{\text{th}}$ century, it may be hard to fully imagine how mathematicians thought in 1945. Their definitions, theorems and comments are clear, but one should be aware that a reconstruction of their ideas, their thinking may be specifically biased by our present understanding of the mathematical concepts involved.

2. Background conceptions

The main ideas of this paper—which includes a very wide spectrum of examples, from ancient Greek mathematics to modern, from children's counting to CT—are the following:

- A mathematical concept, no matter how novel, is never independent of the previous knowledge; it is based on a reorganization of existing ideas.
- A radically new mathematical idea never germinate in somebody's mind without a period of incubation, usually a lengthy one.
- There is a long distance to cover between:

[4] Mac Lane earlier in his life (in particular in (Eilenberg and Mac Lane, 1945) and (Mac Lane, 1950) wrote his name as MacLane. Later he began inserting a space into his surname, in particular in (Mac Lane, 1971).

(1) a spontaneous, unconscious use of a mathematical idea or structure in a concrete setting,
(2) a conscious, systematic use of it.

- A person who has achieved a higher level of mathematical thinking is often unable to imagine thinking of a person from another epoch or of a present learner and, consequently, may unconsciously attribute to him/her an inappropriate (to high or too low) level of thinking.

Transgressions

A *mathematical cognitive transgression* (or briefly: a *transgression*) is defined as crossing—by an individual or by a scientific community—of a previously non-traversable limit of own mathematical knowledge or of a previous barrier of deep-rooted convictions. Moreover, it is assumed that:

1. the crossing concerns a (broadly understood) mathematical idea and the difficulty is inherent in the idea,
2. the crossing is critical to the development of the idea and related concepts,
3. it is a passage from a specified lower level to a new specified upper level,
4. the crossing is a result of conscious activity (the activity need not be intentional and purposely orientated towards such crossing; generally such effect is not anticipated in advance and may even be a surprise).

In the history of mathematics there were numerous transgressions, of different importance, some great ones and many "minitransgressions”. Usually they were not single acts—they involved a global change of thinking which matured for years or even generations and they based on the work of many people. In ancient times two transgressions were the most significant:

- The transition from practical dealing with specific geometric shapes to deductive geometry.
- The celebrated discovery (in the 5th century BC) that the diagonal of a square is incommensurable with its side, that they are ἄλογος (*a-logos*), without a ratio, irrational (in modern setting, foreign to Greek thinking, it was the irrationality of $\sqrt{2}$) Baszmakowa, 1975, pp.80-81. The Pythagorean paradigm was undermined by this απορία. Their understanding of mathematics was eroded. However, nothing certain is known about this discovery. Stories presented in popular books are based on doubtful legends from sources written seven or eight centuries later (Knorr, 1975, p.21, 51). This incommensurability could not be a single discovery by an individual. It must have been a lengthy process. Never in the historical development of mathematics such a major change occurred in short time.

In modern history there were many transgressions. Let us list some of the best known:

- Acceptance of negative numbers.
- The transition from potential infinity (infinity at a *process* level) to the *actual* infinity (infinity as an object).
- The emergence of projective geometry.
- The discovery of non-Euclidean geometries.

We will discuss the case of CT, arguing in particular that creating the theory of *elementary topoi* in CT should be regarded as a major transgression.

Phylogeny and ontogeny

The term *phylogeny* refers here to the evolutionary history of mathematics (or rather to its modern reconstructions), from ancient times on. *Ontogeny* means the development of basic mathematical concepts and structures in the mind of an individual person, from early childhood. Phylogeny and ontogeny are in some sense complementary descriptions (Freudenthal, 1984; Piaget and García, 1989, pp.4-29).

In case of mathematics, the oft-quoted phrase: *ontogeny recapitulates phylogeny* implies that one can learn from the history of old mathematics for the sake of present teaching. This sometimes gives useful hints, e.g. one may argue that since the historical process of forming the general concept of a function took centuries (from Descartes, if not much earlier, to Peano and Hausdorff), we should not expect that a secondary school student can grasp it—learning Dirichlet's description—after a few lessons. The general concept of the function was not yet quite clear to mathematicians of the first half of the 19$^{\text{th}}$ century (Lakatos, 1976, Appendix 2; Youschkevitch, 1976; Ferreirós Domínguez, 1999, pp.27–30).

On the other hand, the idea that ontogeny recapitulates phylogeny may be misleading. Piaget always stressed that arithmetic cognition results from logico-mathematical experience with concrete objects, pebbles say, and is educed from *the child's actions* rather than from heard words. It is abstracted from a coordination of intentional motions and accompanying thoughts. Nevertheless, Piaget was in favour of the phylogeny-ontogeny parallelism and reasoned roughly as follows. Since one-to-one correspondence preceded numerical verbal counting in the very early periods of human civilization (evidenced on artefacts such as notched bones and also found in rude unlettered tribes), the same should apply to children. Cantor's theory of cardinal numbers confirmed this thinking (Beth and Piaget, 1966, pp.259–260). Consequently, Piaget and many educators insisted on one-to-one correspondence as a foundation of early school arithmetic, neglecting the fact that nowadays children learn number names early, often together with learning to speak, and moreover counting is now deeply rooted culturally. Research evidence shows that counting, rather than one-to-one correspondence, is a basis of the child's concept of number (Gelman and Gallistel, 1978, pp.77–82). In this way the phylogeny–ontogeny parallelism adversely affected early mathematics education in the time of the 'New Math' movement.

Hans Freudenthal, in the context of mathematics, suggested the converse idea: *What can we learn from educating the youth for understanding the past of mankind?* This reverses the traditional direction of inference in the phylogeny–ontogeny parallelism. In particular, one may ask whether contemporary knowledge of the difficulties in the transition from the concrete to more abstract mathematical reasoning of children may be helpful in better understanding of limitations of our reconstructions of the development of the early Greek mathematics.

In the sequel, certain aspects of the development will be traced both in phylogeny and in ontogeny, inextricably intertwined with the the mathematical questions themselves.

Platonizing constructivism in mathematics

The theoretical framework of the paper is platonizing constructivism in mathematics. It is assumed that:

- each of the three major positions in the philosophy of mathematics from the beginning of the 20th century: *platonism*, *constructivism*, *formalism* describes some inherent, complementary features of mathematics;
- they can be reconciled provided that they are regarded as *descriptive*, as an account of some inherent features of mathematics, and *not normative*, i.e., when one skips the eliminating words (as 'only', 'oppose') which explicitly deny other standpoints. Moreover, various versions of the three positions often overlap.

In the sequel, the term *constructivism* will not be understood as in papers on foundations of mathematics, but rather in a way akin to its meaning in research on mathematics education, related to post-Piagetian psychological versions of constructivism. Briefly, one assumes here that humans construct mathematical concepts in their minds and discover their properties. A concept develops its necessary structure as a consequence of its context and—in the long term—becomes *cristalline* in the sense of David Tall (2013, p.27); then its properties appear independent of our will. This phenomenon may be traced both in phylogeny and in ontogeny. Moreover, in each essential progression, new mental structures are build on the preceding ones and are always integrated with previous ones (Piaget and García, 1989, pp.22–29).

By platonizing constructivism we mean an analysis—in constructivistic terms — of sources and consequences of the platonistic attitude of a majority of mathematicians and contrasting them with the well-known difficulties of consistent platonism in the philosophy of mathematics.

3. Developmental successors

After the introductory examples we now look for ways to distinguish between:

- mathematical concepts which—historically—were natural successors to previous ones,
- concepts which could be conceived and defined only after opening new paths of thought and reasoning.

The following metaphorical labels will be used: *onward development*, *branching-off*, *upward development*, *downward development*, interpreted with examples. We do not expect to find clear criteria, but the ensuing discussion may be illuminating.

The *emergence of numerals* in the Late Stone Age is evidenced by tally marks (in the form of notched bones). Ethnologists have found that early tribes had only two counting words: *one* and *two*, followed by *many*. Also in present Indo-European languages these two numbers and their ordinal counterparts are linguistically different from the following numbers. It has been suggested that the proto-Indo-European number **trei* (three) was derived from the verb **terh* (meaning: *pass*); thus, the word *three* is related to *trans*. This may be a hint of a very ancient mental obstacle between numbers *two* and *three*.

One may conjecture that *after the passage from 2 to 3 there was no notable obstacle to gradual development of unlimited counting*. Of course, the actual development took centuries, if not millennia. Anyway, for present children there is no hurdle between 2 and 3, as they are taught counting very early. Moreover, counting starts to make sense with three items.

Onward development

Onward development of indefinite counting includes its *developmental successors*: simple addition of natural numbers (which develops through a stage called *count all* and then a more advanced stage *count on*), subtraction (as taking away), multiplication, division (originally there are two kinds of it: *equal sharing* and *equal grouping*), and even simple powers, all within some range of natural numbers.

These concepts are included in the onward development of counting, by virtue of the following features:

- *no branching*: each new concept naturally comes after the previous ones;
- *ontological stability*: each concept (e.g., number 17, product 3×6), remains essentially the same object, although the related ideas are enriched after each extension of the scope of arithmetic and—in the historical development—are subject to evolutionary changes.

The conception of developmental successors, outlined here, does not take into account a relative difficulty of concepts; what is crucial is whether they follow the previous lines of thinking.

Branching-off

This conception arises from a negation of the first requirement in the description of an onward development. An example of it are fractions, which *branch off* from natural numbers; it is not onward development, although there are many ties between natural numbers and fractions.

There are two ways of introducing fractions to children. In the first, some idealized whole is divided into m equal parts and then n of them are taken. In the second, n whole things are equally divided into m parts. They are two main aspects of the concept of a fraction. For instance, $\frac{3}{4}$ of pizza may be obtained by cutting it into 4 parts and taking 3 such parts (thus $\frac{3}{4} = 3 \times \frac{1}{4}$). A more advanced way of thinking of $\frac{3}{4}$ is 3 divided by 4; the latter may be explained with the example of 3 pizzas to be divided among 4 persons.

The distinction looks quite elementary. Yet, it was significant in the phylogeny of fractions. First procedure is akin to that of ancient Egyptians, the second – to Greek ratios; both were inherited by Arabic mathematicians. In the ontogeny the two ways are always present, but not necessarily noticed. The following reminiscence by William Thurston (1946-2012), written 8 years after he had received Fields Medal, describes his discovery of the identification of previously different objects.

> I remember as a child, in fifth grade, coming to the amazing (to me) realization that the answer to 134 divided by 29 is $134/29$ (and so forth). What a tremendous labor-saving device! To me, '134 divided by 29' meant a certain tedious chore, while $134/29$ was an object with no implicit work. I went excitedly to my father to explain my major discovery. He told me that of course this is so, a/b and a divided by b are just synonyms. To him it was just a small variation in notation (Thurston, 1990, p.848).

The fraction $\frac{a}{b}$ becomes identified with the result of division $a \div b$ and—in this synthesis—they both form a single mathematical object. Philosophically, however, it is an ontological change: two different beings, results of two different mental constructions, become regarded as a single one.

Onward branching-off can be traced in many parts of mathematics. Calculus branches off from a theory of the field of real numbers (axiomatic or based on a construction). Infinite sequences of real numbers branch off from elementary algebra of real numbers. Limits of sequences branch off from general theory of sequences. These examples vividly show that the question of distinguishing branching-off from onward development is delicate, as the criteria are far from being precise, but it may contribute to better understanding the historical development of mathematics.

Upward development and downward development

By *upward development* of a piece of mathematics we mean passing from some concepts and relations between them to a more abstract version of them. Examples:

- Transition from practical addition (verbalized as, e.g., *two and three make five*) to symbolic version (e.g., 2+3=5) took centuries (the sign $+$ appeared in some 15$^{\text{th}}$ century records; the first occurrence of the equality sign $=$ was found in a text by a Welsh mathematician Robert Recorde from 1557).
- Transition from $\mathbb{R}^n$ to an axiomatically given vector space over $\mathbb{R}$.
- Transition from a vector space over $\mathbb{R}$ to a vector space over a field.
- Transition from group theory to the category **Grp**.
- Transition from a category to a metacategory (in the sense of (Mac Lane, 1971, pp.7–11)).

Downward development is—in some sense—an inverse process, to more concrete questions or to a lower level of abstraction, so the above examples may be used the other way round. Also some typical applications of mathematics may be included here, e.g., the passage from abstract Boolean algebras to a description of certain types of electrical circuits (as conceived by Claude Shannon (1936)).

In the 20$^{\text{th}}$ century the mathematics grew rapidly and the upward development became much easier mentally as a result of both: a general change of the attitude of mathematicians toward abstraction and the routine of expressing all concepts in the language of set theory. Branching-off were so frequent that the above metaphors are of little use. There is, however, a notable exception: a new theory which opened a new direction of thinking, so its beginnings may be discussed in a way akin to that used with respect to distant past.

4. The rise of category theory (CT)

A very special feature of CT is that it has a pretty precise date of its official birth: the publication of the paper by Samuel Eilenberg and Saunders Mac Lane (1945). It was presented at a meeting of the American Mathematical Society in 1942 and published in 1945.

According to the Stanford Encyclopedia of Philosophy, *CT "appeared almost out of nowhere"*. Not quite so. As in any mathematical theory, some CT ideas had been conceived much earlier, particularly in algebraic topology, and some of them can be traced to the 19$^{\text{th}}$ century.

Many conceptual transformations—either explicit and well recognized or used implicitly, without awareness—contributed to the rise of CT. Going back, a significant factor was the historic development of the mathematical concept of a function.

Until the beginning of the 19th century, a general symbol for a function (f or φ) was almost non-existent (Youschkevitch, 1976). A significant step toward CT was the general notion of a mapping introduced by Dedekind in 1888. In his *Enklärung* (*explanation*) he did not define the concept of *Abbildung* φ (literally: *image* or *representation*) from a set (*System*) S into a set S', but interpreted it generally as an arbitrary law (*Gesetz*) according to which to each element s there corresponds (*gehört*) a certain thing (*Ding*) $S' = \varphi(s)$, called the image (*Bild*) of s. He also defined a composed mapping (*zusammengesetzte Abbildung*) $\vartheta(s) = \psi(s') = \psi(\varphi(s))$ of two given ones, denoted as $\varphi.\psi$ or $\varphi\psi$, defined injective mappings (*ähnlich* or *deutlich*), the inverse mapping and proved their main properties (Dedekind, 1888, §2–4; Ferreirós Domínguez, 1999, p.88–90, 228–229).

Emmy Noether in her lectures in the 1920s emphasised the role of homomorphisms in group theory. Before her, groups were understood as generators and relations (in modern terms, as quotients of free groups). She also argued that the homology of a space is a group, is an algebraic system rather than a set of numbers assigned to the space. Her lectures and the lectures of Emil Artin formed a basis for the celebrated book by van der Waerden (1930) on modern algebra. This current of thought led to CT.

Generally, in the symbol of the type $f(x)$, the part f was always understood as fixed and x was a variable. At the end of the 1920s, however, in functional analysis and related fields, a new way of thinking emerged. In certain situations the roles of symbols in $f(x)$ reversed: the point x was regarded as fixed while the function f became a variable (as an element of a function space), e.g., $x \in [0,1]$ was fixed and f was a variable in the space $C([0,1])$ of continuous functions on the interval $[0,1]$. In this new role, the point x became a functional δ_x on $C([0,1])$. Such a change of the roles function-element became crucial, e.g., in the Potryagin duality of locally compact abelian groups (Hewitt and Ross, 1963) and in Gelfand's theory of commutative Banach algebras. It was also used by Eilenberg and Mac Lane in their first example (finite dimensional vector spaces and their dual spaces) motivating the concept of a natural equivalence.

A crucial example of a contravariant functor was the adjoint T^* of a linear operator T on a Hilbert space, introduced in 1932 by John von Neumann.[5]

According to Mac Lane, abstract algebra, lattice theory and universal algebra were necessary precursors for CT. However, he also suggested that certain notational devices preceded the definition of a category. One of them was the fundamental idea of representing a function by an arrow $f\colon X \to Y$, which first appeared in algebraic topology about 1940, probably introduced by Polish-born topologist Witold Hurewicz (Mac Lane, 1971, p.29; 1988, p.333). Originally, it looked as just another symbol, but from a later perspective the use of such symbol was one of the key changes. Thus, a notation (the arrow) led to a concept (category). Such new symbols later got absorbed into the background of mathematical thinking, used as something obvious. Together with commutative diagrams, which were probably also first used by Hurewicz, they paved the way to CT.

[5] Mac Lane, p.330 tells a story how Marshall Stone advised von Neumann to introduce the symbol T^* and how it changed the publication. He also mentions a fact which may interests philosophers: in 1929 von Neumann lectured in Göttingen and presented his axiomatic definition of a Hilbert space, while David Hilbert—listening to it—evidently thought of it as of the concrete space ℓ^2, not in the axiomatic setting.

Mac Lane often accented two features of mathematics: computational and conceptual. He noted that the initial discovery of CT came directly from a problem of calculation in algebraic topology (Mac Lane, 1988, p.333).

Eilenberg and Mac Lane were aware that they introduced very abstract mathematical tools, which did not fit any algebraic system in the Garrett Birkhoff's universal algebra. It might seem too abstract and was certainly off beat and a "far out" endeavour. Although it was carefully prepared, it might not have seen the light of day (Mac Lane, 2002, p.130).

The origin of the term *functor*

Mac Lane has written "*Categories, functors, and natural transformations were discovered by Eilenberg–Mac Lane in 1942*" (Mac Lane, 1971, p.29). The word "discovered" may be regarded as an indication of a hidden Platonistic attitude of Mac Lane, in spite of his verbal declarations against Platonism (Mac Lane, 1986, pp.447–449; Król, 2019; Skowron, 2020). He also wrote:

> Now the discovery of ideas as general as these is chiefly the willingness to make a brash or speculative abstraction, in this case supported by the pleasure of purloining words from the philosophers: "Category" from Aristotle and Kant, "Functor" from Carnap (*Logische Syntax der Sprache*) (Mac Lane, 1971, pp.29–30).

This sentence has been taken very seriously by several authors. However, the way it was phrased suggests that it was rather intended to be a delicate joke.[6] Attributing the origin of the term *category* to Aristotle and Kant is clear, although in 1899 René-Louis Baire (in his *Thèse*) introduced—in another context—the word *category* to mathematics.[7] Concerning the origin of the term *functor* in CT, one may recall the following facts:[8]

- In Rudolf Carnap's book *Abriss der Logistik* (Carnap, 1929) the term "Funktor" does not appear.
- In 1929 the Polish term "funktor" was used in propositional calculus by Tadeusz Kotarbiński in his book (Kotarbiński, 1929).
- Alfred Tarski often emphasised that Kotarbiński had been his teacher (Feferman and Feferman, 2004, part 2).

---

[6] Let us note that the noun *brash* means a mass of fragments; according to *Cambridge International Dictionary of English* (1995), the adjective *brash* is disapproving, referring to people who show too much confidence and too little respect, while *Webster's New World Dictionary* (1984) lists—as meanings of *brash*—also *hasty and reckless*, *offensively bold*. On the other hand, the word *purloining* means *stealing* or *borrowing without permission*. Such a comment (with the word *pleasure*) by Mac Lane concerning CT could not be serious. On the other hand, in 2002 Mac Lane came back to Carnap, adding: "Also the terminology was largely purloined: "category" from Kant, "natural" from vector spaces and "functor" from Carnap. (It was used in a different sense in Carnap's influential book *Logical Syntax of Language*; I had reviewed the English translation of the book (in the Bulletin AMS 1938) and had spotted some errors; since Carnap never acknowledged my finding, I did not mind using his terminology)" (Mac Lane, 2002, pp.130–131).

[7] A subset A of a topological space X is called *a set of first category* (*un ensemble de première catégorie*) in X iff A is the union of a countable family of nowhere dense sets; otherwise it is *a set of the second category* (*Menge erster und zweiter Kategorie*). The celebrated *Baire category theorem* states, in a generalized form, that a complete metric space is not a set of the first category (Hausdorff, 1914, p.328; Kuratowski, 1933, § 10). The clumsy term *set of the first category* was later replaced by the term a *meager set* (Kelley, 1955, p.201). In the 1930s Baire category theorem was a very popular tool in the Warsaw school of topology, so Eilenberg must have known it.

[8] The author is indebted to Professor Jan Woleński for the relevant information.

- Carnap met Tarski in Vienna in February 1930 and visited Warsaw in November 1930; he learned much from Tarski.
- In 1933 Tarski, in the Polish version of his famous paper *On the concept of truth in formal languages* Tarski (1933) used the term “funktor” and mentioned that he owed the term to Kotarbiński.
- Carnap used the term “Funktor” in his book (1934) (quoted by Mac Lane) in a sense more general than that of Kotarbiński and Tarski.
- Eilenberg studied mathematics in Warsaw from 1930. In 1931 he attended Tarski’s lectures on logic (Feferman and Feferman, 2004, part 3 and 12). He left Warsaw in 1939.

This evidence strongly suggests that both Carnap and Eilenberg could have learned, independently, the term from Kotarbiński and Tarski.

Was the original CT an onward development of previous mathematical theories?

Using a metaphor explained above, one may argue that the definition of a category and of a functor were within a major onward development of part of mathematics of the first half of the 20th century, that is set theory, algebra, topology etc. Indeed, for a person working in group theory, say, a natural continuation should be to think of all groups, their homomorphisms, isomorphisms, and the composites as of a single whole: *group theory*. Similarly one could think of vector spaces with linear maps as of another whole. Some analogies between theories were obvious. Moreover, axioms of CT are reminiscent of those of semigroup theory. The concept of a covariant functor was a natural analogue of homomorphisms of algebras. Contravariant functors had been present in various duality theories (e.g., in Pontryagin's duality mentioned above). CT provided general concepts applicable to all branches of abstract mathematics, contributed to the trend towards uniform treatment of different mathematical disciplines, provided opportunities for the comparison of constructions and of isomorphisms occurring in different branches of mathematics, and may occasionally suggest new results by analogy (Eilenberg and Mac Lane, 1945, p.236).

The great achievement of Eilenberg and Mac Lane was the idea that a formalization of various evident analogies was worth systematizing and publishing. The initial neglect of (Eilenberg and Mac Lane, 1945) by mathematicians was very likely a result of the fact that it was regarded as a long paper within onward development of known part of mathematics, with many rather simple definitions and examples, tedious verification of easy facts, and no theorem with an involved proof. Ralf Krömer, in his book on the history and philosophy of CT, has outright stated that Eilenberg and Mac Lane needed to have remarkable courage to write and submit for publication the paper almost completely concerned with conceptual clarification (Krömer, 2007, p.65).

A novelty of (Eilenberg and Mac Lane, 1945, p.272), which at first appeared insignificant, was regarding elements p_1, p_2 of a single quasi-ordered set P as objects of a category, with a unique morphism $p_1 \to p_2$ iff $p_1 \leq p_2$ and no morphisms otherwise. This opened a way to a series of generalizations, in particular regarding certain commutative diagrams as functors on small categories.

One may argue that this achievement was still within onward development of CT as it was within the scope of previous knowledge.

Let us recall that the difficulty and the originality of a theorem are not taken into account; what is crucial is whether the concepts involved are natural extension of the previous knowledge and thinking.

The category axioms represent a very weak abstraction (Goldblatt, 1984, p.25). In spite of this fact, a few years later the conceptual clarification turned out highly effective in the book *Foundations of Algebraic Topology* written by Eilenberg together with Norman Steenrod (1952). The latter admitted in a conversation that the 1945 paper on categories had a more significant impact on him than any other research paper, it changed his way of thinking.

The Eilenberg-Mac Lane Program

This program has been formulated as follows:

> The theory also emphasizes that, whenever new abstract objects are constructed in a specified way out of given ones, it is advisable to regard the construction of the corresponding induced mappings on these new objects as an integral part of their definition. The pursuit of this program entails a simultaneous consideration of objects and their mappings (in our terminology, this means the consideration not of individual objects but of categories). [...]
>
> The invariant character of a mathematical discipline can be formulated in these terms. Thus, in group theory all the basic constructions can be regarded as the definitions of co- or contravariant functors, so we may formulate the dictum: The subject of group theory is essentially the study of those constructions of groups which behave in a covariant or contravariant manner under induced homomorphisms. More precisely, group theory studies functors defined on well specified categories of groups, with values in another such category. This

> may be regarded s a continuation of the Klein Erlanger Programm, in the sense that a geometrical space with its group of transformations is generalized to a category with its algebra of mappings.
>
> [...] such examples as the "category of *all* sets", the "category of *all* groups are illegitimate. The difficulties and antinomies are exactly those of ordinary intuitive *Mengenlehre*; no essentially new paradoxes are involved. [...] we have chosen to adopt the intuitive standpoint, leaving the reader free to insert whatever type of logical foundation (or absence thereof) he may prefer. [...]
>
> It should be observed first that the whole concept of a category is essentially an auxiliary one; our basic concepts are essentially those of a functor and of a natural transformation. [...] The idea of a category is required only by the precept that every function should have a definite class as domain and a definite class as range, for the categories are provided as the domains and ranges of functors. (Eilenberg and Mac Lane, 1945, p.236–237, 246–247)

The quoted comparison of CT to the celebrated program of Felix Klein shows vividly that the authors regarded their work as significant. A important novelty of (Eilenberg and Mac Lane, 1945) was to use the same letter to denote both: the object component of a functor and its morphism component. This had not been a common practice, even when both correspondences were dealt with in a single paper. This novelty and the whole program fit well Atiyah's conception (quoted above) of ideas *absorbed by mathematicians' culture*.

CT is both a specific domain of mathematics and at the same time a conceptual framework for a major part of modern theories.

5. The amazing phenomenon of unexpected branchings-off in CT

Up to this point CT might be regarded are being within onward development of earlier theories: algebra, topology, functional analysis. However, a branching-off in (Eilenberg and Mac Lane, 1945) is the concept of a *natural equivalence* (central in the title of the paper), with an essential use of commutative diagrams.[9] It was a completely new idea, but its initial impact was limited.

After 1945 CT—as a general theory—lay dormant till the emergence of significant new concepts and a breaking series of major branchings-off in the second half of the 1950s. One of their outstanding features was a new type of a definition, formulated in the form of a *unique factorization problem*. First such explicit definition (1948) appeared in the paper by Pierre Samuel, a member of the Bourbaki group, on free topological groups, albeit it was still in the language of set theory, without arrows. In 1950, commutative diagrams were demonstrated as a convenient tool in such problems by Mac Lane. The concept of a *dual category*, formulated in (Eilenberg and Mac Lane, 1945, p.259) and further developed by Mac Lane (1950), had its conceptual roots in various duality theories, particularly in that of projective geometry. The *product* of two categories was an analogue of that for groups and various algebras. Mac Lane analysed the concept of duality, stressed diagrammatic dualities of various pairs of concepts and presented the definitions of *direct* and *free products* in group theory (later generalized to the concepts of categorial *products* and *coproducts*, respectively).

[9] It is not clear why Eilenberg and Mac Lane refrained from setting the concept in the general form of a *natural transformation* (examples abounded). Perhaps they felt they should not pursue a still more general setting without accompanying results.

A metamorphosis from Eilenberg–Mac Lane Program to mature CT

A turning point in the development of CT was the seminal paper by Daniel Kan (1958) on *adjoint functors*. In much the same time period, independently, several closely related concepts and results were worked out: *representable functors*, *universal morphisms*, *Yoneda lemma*, various types of *limits* and *colimits* (Mac Lane, 1988, pp.345–352). Special cases of them had a long earlier history in specific situations in algebra and topology (e.g., Freudenthal's theorems on *loops* and *suspensions* in homotopy theory proved in 1937). This confirms a known phenomenon that mathematicians may use an idea spontaneously, without being conscious of it in a more abstract setting.

Mac Lane (1950) also opened the way to the study of categories with additional structure, which some years later developed to the study of abelian categories. This topic was developed—in a remarkably short time—due to the work of Alexandre Grothendieck, David Buchsbaum, Pierre Gabriel, Max Kelly and authors of two monographs: Peter Freyd (1964) and Barry Mitchell (1965).

Within 20 years CT, originally conceived as a useful language for mathematicians, became a developed, mature theory, something totally unexpected by its founders.

Set theory without elements

The results of the work of William Lawvere turned out to be not only a new branch of CT, but also opened new perspectives in mathematics, logic, foundations of mathematics, and philosophy. In his Ph.D. thesis at Columbia University in New York, supervised by Eilenberg (defended in 1963, known from various copies, with full text published 40 years later) many new ideas were presented, including a categorical approach to algebraic theories (Lawvere, 1963).

Lawvere also tackled the general question as to what conditions a category must satisfy in order to be equivalent to the category **Set**. The idea looked analogous to the so-called *representation theorems*, i.e., propositions asserting that any model of the axioms for a certain abstract structure must be (in some prescribed sense) isomorphic to a specific type of models of the theory or to one particular concrete model.[10] However, Lawvere's case was unique and controversial in the sense that his 'sets' were conceived *without elements*. The theory did not have the primitive notion "element of". And it did work.

Specifically, Lawvere characterized **Set** (up to *equivalence* of categories) as a category $\mathcal{C}$ with the following: an *initial* object **0**; a *terminal* object **1** (which gives rise to *elements* of A defined as morphisms from **1** to A); *products* and *coproducts* of finite families of objects; *equalizers* and *coequalizers*; for any two objects there is an *exponential*; existence of a specific object **N** with morphisms $0\colon \mathbf{1} \to \mathbf{N}$ and $s\colon \mathbf{N} \to \mathbf{N}$ yielding the successor operation s on **N** and a simple recursion for sequences; axiom that **1** is a *generator* (if parallel morphisms f, g are not equal, then there is an element $x \in A$ such that $xf \neq xg$); axiom of choice; three additional elementary axioms of this sort (everything in the language of CT). This was augmented with one non-elementary axiom: $\mathcal{C}$ has products and coproducts for

[10] The oldest theorems of this type are: Cayley's theorem that every (abstract) group is isomorphic to a group of bijections of a set; Kuratowski's theorem that every partially ordered set is order-isomorphic to a family of subsets of a set, ordered by inclusion; theorem that every group with one free generator is isomorphic to $\mathbb{Z}$. Analogous examples are known in many theories.

any *indexing infinite set*. A coproduct of copies of **1** played the role of a set (Lawvere, 1964; Mac Lane, 1986, pp.386–407; 1988, pp.341–345).

The point was not to avoid membership relation completely, but (instead of taking as the starting point the primitive notions of *set*, *element* and membership $\in$) one takes *function* as a primitive notion of the theory (with suitable axioms, using elementary logic, but avoiding any reference to sets) and then one derives membership and most concepts of set theory as a special case from there.

In the second half of the 1960's Lawvere opened a way to a new theory of *elementary toposes* (called also *elementary topoi*, with Greek plural *τόποι* of the noun *τόπος*). Unexpected territories of mathematics were discovered (Lawvere, 1972; Mac Lane, 1988, pp.352–359; Krömer, 2007).

CT became a contender for a foundation of mathematics, although the hope that it undermine the overwhelming role of set theory turned out spurious and most working mathematicians keep away from CT and toposes. CT yields new tools to study many formal mathematical theories and mutual relations between them, from a perspective different from that set theory.

6. Recapitulation of some points

Let us recall Atiyah's remark (quoted in the Introduction) that really important discoveries get later omitted altogether as they become absorbed by the general mathematical culture. This thought fits particularly well with the case of CT. Most of the ideas presented by Eilenberg and Mac Lane in 1945 have been absorbed as a natural language of advanced mathematical thinking. Once mathematicians learnt the definitions of a functor and a natural transformation, these concepts became a major tool of mathematical thinking in many abstract theories of the second half of the 20th century. However, it took several years to realise the scope of the change. Freyd commented as follows:

> MacLane's definition of "product" (1950) as the solution of a universal mapping problem was revolutionary. So revolutionary that it was not immediately absorbed even by most category minded people.
>
> [...] In a new subject it is often very difficult to decide what is trivial, what is obvious, what is hard, what is worth bragging about (Freyd, 1964, p.156).

Mac Lane, however, used the definition of a product and its dual only in the case of groups (general or abelian). He did not formulate it *mutatis mutandis* in the general case of a category, although he had several simple examples at hand. Freyd also told the story of the term *exact sequence*, a technical definition in homological algebra. In the late 1950's, when he was a graduate student at Brown University, he was brought up to think in terms of exactness of maps. This concept seemed to him as fundamental as the notion of continuity must seem to an analyst. And later he was astonished to hear that when Eilenberg and Steenrod wrote their fundamental book (Eilenberg and Steenrod, 1952) (published in 1952) they defined this very notion, recognized the importance of the choice of a suitable name for it, and could not invent any satisfactory word. Consequently, they wrote the word "blank" throughout most of the manuscript, ready to replace it before submitting the book for publication. After entertaining an unrecorded number of possibilities they settled on "exact" (Freyd, 1964, p.157).

One may argue that the 1945 definitions of a category and of a functor were within a major onward development of abstract algebra and other advanced topics. In fact, originally they were not regarded as a novelty. Eilenberg and Mac Lane were not even certain whether their paper will be accepted for publication (it was long and lacked theorems with substantial proofs). However, they were genuinely convinced of the significance of their conceptual clarification and took pains to write the paper clearly and to attract the reader.

After this publication for almost ten years CT appeared dormant. The groundbreaking papers on abelian categories by Buchsbaum and Grothendieck marked a far-reaching change. And then—in the 1960's—CT unexpectedly started to grow rapidly, with astonishing results (Mac Lane, 1988, pp.338–339, 341–361).

Thus, from the present perspective, in spite of the previous arguments, one can say that the emergence of CT was undoubtedly a major transgression in mathematics. It was a crossing of a previously non-traversable barrier of deep-rooted habits to think of mathematics. A vivid argument is the fact that—even after publication of the main ideas—it was so difficult to overcome the previous inhibition and widespread tradition.[11]

The creators of CT and their followers could choose their definitions freely, nobody could forbid that. And yet the previous way of thinking was an obstacle for potential authors and for prospective readers. Great insight of Eilenberg and Mac Lane of what is significant in mathematics turned out a crucial factor.

[11] Many mathematicians, in USA and elsewhere, expressed disinclination to CT. Karol Borsuk, an outstanding topologist, the teacher of Eilenberg in Warsaw and coauthor of their joint paper published in 1936, was later unfavourable to CT and the categorical methods in mathematics (Jackowski, 2015, p.30). Jerzy Dydak, a student of Borsuk, recalled after years: *My own PhD thesis written under Borsuk in 1975 makes extensive use of category theory and I was asked by him to cut that stuff out. Only after I assured him that I spent many months trying to avoid abstract concepts, he relinquished and the thesis was unchanged* (Dydak, 2012, p.92).

Bibliography

Atiyah, M., 2002. Mathematics in the 20th century. *Bulletin of the London Mathematical Society* [Online], 34(1), pp.1–15. Available at: https://doi.org/10.1112/S0024609301008566 [visited on 29 October 2020].

Baszmakowa, I., 1975. Grecja starożytna. Kraje hellenistyczne i imperium rzymskie. *Historia matematyki od czasów najdawniejszych do początku XIX stulecia: w trzech tomach. T. 1: Od czasów najdawniejszych do początku czasów nowożytnych* (S. Dobrzycki, Trans.). Warszawa: Państwowe Wydawnictwo Naukowe, pp.64–168.

Beth, E.W. and Piaget, J., 1966. *Mathematical Epistemology and Psychology*. Dordrecht: D. Reidel Publishing Company.

Cantor, G., 1883. Über unendliche, lineare Punktmannigfaltigkeiten. *Mathematische Annalen* [Online], 21(4), pp.545–591. Available at: https://doi.org/10.1007/BF01446819 [visited on 29 October 2020].

Carnap, R., 1929. *Abriss der logistik: mit besonderer Berücksichtigung der relationstheorie und ihrer Anwendungen*, *Schriften zur Wirtschaftswissenschaftlichen Forschung* Bd. 2. Wien: Verlag von Julius Springer.

Carnap, R., 1934. *Logische Syntax der Sprache*, *Schriften zur Wissenschaftlichen Weltauffassung* Bd. 8. Wien: Verlag von Julius Springer.

Dedekind, R., 1872. *Stetigkeit und irrationale Zahlen* [Online]. Braunschweig: Friedrich Vieweg & Sohn. Available at: <http://www.rcin.org.pl/publication/13029>.

Dedekind, R., 1888. *Was sind und was sollen die Zahlen?* [Online]. Braunschwieg: Friedrich Vieweg & Sohn. Available at: <http://www.digibib.tu-bs.de/?docid=00024927> [visited on 30 October 2020].

Dydak, J., 2012. Ideas and Influence of Karol Borsuk. *Wiadomości Matematyczne* [Online], 48(2), p.81. Available at: https://doi.org/10.14708/wm.v48i2.305 [visited on 30 October 2020].

Eilenberg, S. and Mac Lane, S., 1945. General theory of natural equivalences. *Transactions of the American Mathematical Society* [Online], 58, pp.231–294. Available at: https://doi.org/10.1090/S0002-9947-1945-0013131-6 [visited on 30 October 2020].

Eilenberg, S. and Steenrod, N.E., 1952. *Foundations of Algebraic Topology*, *Princeton Mathematical Series* 15. Princeton: Princeton University Press.

Feferman, A.B. and Feferman, S., 2004. *Alfred Tarski: Life and Logic*. Cambridge [etc.]: Cambridge University Press.

Ferreirós Domínguez, J., 1999. *Labyrinth of Thought: A History of Set Theory and Its Role in Modern Mathematics*, *Science Networks. Historical Studies* vol. 23. Basel: Birkhäuser.

Freudenthal, H., 1984. The Implicit Philosophy of Mathematics: Histtory and Education. In: Z. Ciesielski and C. Olech, eds. *Proceedings of the International Congress of Mathematicians, August 16-24, 1983, Warszawa*. Warszawa, Amsterdam [etc.]: PWN-Polish Scientific Publishers, North-Holland, pp.1695–1709.

Freyd, P., 1964. *Abelian Categories: An Introduction to the Theory of Functors*. New York: Harper & Row.

Gelman, R. and Gallistel, C.R., 1978. *The Child's Understanding of Number*. Cambridge Mass; London: Harvard University Press.

Goldblatt, R., 1984. *Topoi - The Categorial Analysis of Logic*. North-Holland.

Hausdorff, F., 1914. *Grundzüge der Mengenlehre* [Online]. Leipzopg: Veit (Leipzig). Available at: <http : / / gallica . bnf . fr / ark : /12148 / bpt6k9736980b> [visited on 30 October 2020].

Heller, M., 2015. *Bóg i geometria: gdy przestrzeń była Bogiem*. Kraków: Copernicus Center Press.

Hewitt, E. and Ross, K.A., 1963. *Abstract Harmonic Analysis. Vol. 1: Structure of Topological Groups, Integration Theory, Group Representations*, *Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen mit Besonderer Berücksichtigung der Anwendungsgebiete* Bd. 115. Berlin: Springer.

Jackowski, S., 2015. Samuel Eilenberg - wielki matematyk z Warszawy. *Wiadomości Matematyczne* [Online], 50(1), pp.21–43. Available at: https://doi.org/10.14708/wm.v50i1.651 [visited on 1 November 2020].

Kan, D.M., 1958. Adjoint functors. *Transactions of the American Mathematical Society* [Online], 87(2), pp.294–294. Available at: https://doi.org/10.1090/S0002-9947-1958-0131451-0 [visited on 1 November 2020].

Kelley, J.L., 1955. *General Topology*, *University Series in Higher Mathematics*. Princeton, N.J.: D. Van Nostrand Co.

Knorr, W.R., 1975. *The Evolution of the Euclidean Elements: A Study of the Theory of Incommensurable Magnitudes and Its Significance for Early Greek Geometry*, *Synthese Historical Library* vol. 15. Dordrecht: D. Reidel.

Kotarbiński, T., 1929. *Elementy teorji poznania, logiki formalnej i metodologji nauk*. Lwów: Wydawn. Zakładu Narodowego im. Ossolińskich.

Krömer, R., 2007. *Tool and Object: A History and Philosophy of Category Theory*, *Science Networks. Historical Studies* vol. 32. Basel [etc.]: Birkhäuser.

Król, Z., 2019. Category theory and philosophy. In: M. Kuś and B. Skowron, eds. *Category Theory in Physics, Mathematics, and Philosophy* [Online], *Springer Proceedings in Physics* 235. Springer International Publishing, pp.21–32. Available at: https://doi.org/10.1007/978-3-030-30896-4 [visited on 1 November 2020].

Kuratowski, K., 1933. *Topologie. 1, Espaces Métrisables, Espaces Complets*, *Monografje Matematyczne* t. 3. Warszawa ; Lwów: s.n.

Lakatos, I., 1976. *Proofs and Refutations: The Logic of Mathematical Discovery*. Ed. by J. Worrall and E. Zahar. Cambridge [etc.]: Cambridge University Press.

Lawvere, F.W., 1963. Functorial semantics of algebraic theories. *Proceedings of the National Academy of Sciences of U.S.A.* [Online], 50, pp.869–872. Available at: <https://www.pnas.org/content/pnas/50/5/869.full.pdf> [visited on 29 October 2020].

Lawvere, F.W., 1964. An elementary theory of the category of sets. *Proceedings of the National Academy of Sciences of the U.S.A.* [Online], 52(6), pp.1506–1511. Available at: https://doi.org/10.1073/pnas.52.6.1506 [visited on 1 November 2020].

Lawvere, F.W., ed., 1972. *Toposes, Algebraic Geometry and Logic*, *Lecture Notes in Mathematics* 274. Berlin, Heidelberg, New York: Springer.

Mac Lane, S., 1950. Duality for groups. *Bulletin of the American Mathematical Society* [Online], 56(6), pp.485–516. Available at: <https://projecteuclid.org/euclid.bams/1183515045> [visited on 29 October 2020].

Mac Lane, S., 1971. *Categories for the Working Mathematician*, *Graduate Texts in Mathematics* 5. New York: Springer-Verlag.

Mac Lane, S., 1986. *Mathematics: Form and Function*. New York [etc.]: Springer-Verlag.

Mac Lane, S., 1988. Concepts and Categories in Perspective. In: P. Duren, R. Askey and U. Merzbach, eds. *A Century of Mathematics in America, Part I* [Online], *History of Mathematics* 1. Providence, R.I.: American Mathematical Society, pp.323–365. Available at: <https://www.ams.org/publicoutreach/math-history/hmath1-maclane25.pdf> [visited on 29 October 2020].

Mac Lane, S., 2002. Samuel Eilenberg and categories. *Journal of Pure and Applied Algebra* [Online], 168(2-3), pp.127–131. Available at: https://doi.org/10.1016/S0022-4049(01)00092-5 [visited on 29 October 2020].

Mitchell, B., 1965. *Theory of Categories*, *Pure and Applied Mathematics. A Series of Monographs and Textbooks* 17. New York, London: Academic Press.

Piaget, J. and García, R.V., 1989. *Psychogenesis and the history of science* (H. Feider, Trans.). New York: Columbia University Press.

Samuel, P., 1948. On universal mappings and free topological groups. *Bulletin of the American Mathematical Society* [Online], 54(6), pp.591–598. Available at: <https://projecteuclid.org/euclid.bams/1183512049> [visited on 29 October 2020].

Shannon, C.E., 1936. *A Symbolic Analysis of Relay and Switching Circuits* [Online]. Thesis. Massachusetts Institute of Technology. Available at: <https://dspace.mit.edu/handle/1721.1/11173> [visited on 2 November 2020].

Skowron, B., 2020. *Was Saunders Mac Lane a Platonic? [to appear].*

Tall, D., 2013. *How Humans Learn to Think Mathematically: Exploring the Three Worlds of Mathematics*, *Learning in Doing: social, cognitive, and computational perspectives*. Cambridge: Cambridge University Press.

Tarski, A., 1933. *Pojęcie prawdy w językach nauk dedukcyjnych*, *Prace Towarzystwa Naukowego Warszawskiego. Wydział III: Nauk Matematyczno-Fizycznych* 34. Warszawa: nakładem Towarzystwa Naukowego Warszawskiego, z zasiłku Ministerstwa Wyznań Religijnych i Oświecenia Publicznego.

Thurston, W.P., 1990. Mathematical Education. *Notices of the Amer- ican Mathematical Society*, 37(7), pp.844–850.

Waerden, B.L.v.d., 1930. *Moderne Algebra. T. 1*, *Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen mit Besonderer Berücksichtigung der Anwendungsgebiete* Bd. 33. Berlin: Verlag von Julius Springer.

Youschkevitch, A.P., 1976. The concept of function up to the middle of the 19th century. *Archive for History of Exact Sciences* [Online], 16(1), pp.37–85. Available at: https://doi.org/10.1007/BF00348305 [visited on 2 November 2020].

Mathematics as a love of wisdom: Saunders Mac Lane as philosopher

Colin McLarty
Department of Philosophy, Case Western Reserve University

Abstract

This note describes Saunders Mac Lane as a philosopher, and indeed as a paragon naturalist philosopher. He approaches philosophy as a mathematician. But, more than that, he learned philosophy from David Hilbert's lectures on it, and by discussing it with Hermann Weyl, as much as he did by studying it with the mathematically informed Göttingen Philosophy professor Moritz Geiger.

Keywords

naturalism, philosophy, mathematics, Aristotle.

Go, read, and disagree for yourself.
(Mac Lane, 1946, p. 390)

This note describes Saunders Mac Lane as a philosopher, and indeed as a paragon naturalist philosopher. Obviously he approaches questions in philosophy the way a mathematician would. He is one. But, more deeply, he learned philosophy by attending David Hilbert's public lectures on it, and by discussing it with Hermann Weyl, as much as he did by studying for a qualifying exam on it with the mathematically informed Göttingen Philosophy professor

Philosophical Problems in Science (Zagadnienia Filozoficzne w Nauce)
No 69 (2020), pp. 113–128

Moritz Geiger (McLarty, 2007; 2020).[1] Before comparing Mac Lane to Penelope Maddy's created naturalist, the Second Philosopher, we relate him as a philosopher to Aristotle.

This is not to disagree with Skowron's view of Mac Lane as a platonist in ontology (Skowron, 2020). This is about Mac Lane on scientific and philosophic method. He understood *philosophy* very much as Aristotle did: Philosophy is the love of wisdom, and every science pursues wisdom and not mere facts. I do not claim Mac Lane got the ideas from Aristotle. So far as I know, Mac Lane himself had no interest in or opinion of Aristotle, though his teachers Weyl and Geiger certainly talked of Aristotle.[2]

1. Aristotle on wisdom and science

> In general the sign of knowing or not knowing is the ability to teach, so we hold that art rather than experience is scientific knowledge; some can teach, some cannot. Further, the senses are not taken to be wisdom. They are indeed the authority for acquaintance with all individual things, but they do not tell the why of anything, for example why fire is hot, but only that it is hot.
>
> It is generally assumed that what is called wisdom is concerned with the primary causes and principles, so that, as has been already stated, those who have experience are held to

[1] Mac Lane's close contact with Paul Bernays in Göttingen deserves more attention. But my research so far has not identified strong philosophic influence from Bernays. See Section 3.

[2] Browder and Mac Lane (1978) give a beautiful survey of mathematics which mentions Plato and Aristotle on ontology. However, the discussion of Plato and Aristotle summarizes the longer discussion by Browder (1976).

> be wiser than those who merely have any kind of sensation, the artisan than those of experience, the craft master than the artisan. (Aristotle, Metaphysics 981f.)[3]

Here Aristotle says scientific knowledge is better than acquaintance, or even experience, in two related ways:

1. scientific knowledge can be taught, and
2. scientific knowledge gives the why of things.

He says wisdom deals with primary causes and principles. And those are his touchstone for scientific knowledge: "We do not know or have scientific knowledge of objects of any methodical inquiry, in a subject that has principles, causes, or elements, until we are acquainted with those and reach the simplest elements" (Physics 184a).

Throughout his career Mac Lane tied pedagogy to research and research to craft. Among many examples see his early notes on presenting mathematical logic to university students (Mac Lane, 1939), and his impassioned argument that theoretical education prepared men and women well to do the applied mathematics which he supervised in World War II (Mac Lane, 1989; 1997).

Mac Lane also insists mathematical understanding includes knowing the reasons for a given theorem. Some proofs of a theorem may reveal the reason, while other technically sufficient proofs will not reveal the reason. In his book for philosophers, Mac Lane sketches proofs for major theorems from many subjects, like linear algebra, or complex analysis. He often describes several alternative proofs for a single theorem and then eventually singles out one as giving the real reason. That book is (Mac Lane, 1986) and some examples are on pages 145, 189, 427, 455.

---

[3] Translations of Aristotle here use "know" for *oida*, "scientific knowledge" for *episteme*, and "acquaintance" for *gnosis*. Of course "wisdom" is *sophia*.

For me, though, the deepest connection between Aristotle's and Mac Lane's loves of wisdom is how they say we gain this scientific knowledge. Both believe in foundations, or "first principles" if you prefer, but neither believes we start with those. Aristotle's theoretical demand of philosophy was a theoretical and practical demand in mathematics for Mac Lane:

> The natural way of [getting scientific knowledge] is to start from the things which are more knowable and obvious to us and proceed towards those which are clearer and more knowable by nature; for the same things are not 'knowable relatively to us' and 'knowable' without qualification. So in the present inquiry we must follow this method and advance from what is more obscure by nature, but clearer to us, towards what is more clear and more knowable by nature. (Physics 184a)

Aristotle speaks of advancing from what is initially clear to us, towards what is more knowable by nature. I am not sure if he believed there was a final point where the absolutely first principles and simplest elements are known so that they will never change. Mac Lane certainly did not believe it for mathematics.

From his early work on field theory (Mac Lane and Schilling, 1939; 1940) and for the rest of his career Mac Lane often worked to find more basic concepts in some part of mathematics. He and Eilenberg spent over a decade collaborating on ever broader uses of the concepts in their "General theory of natural equivalences" (Eilenberg and Mac Lane, 1945). They meant that paper to be the only one ever needed on this technical concept for group theory and topology, but it became the founding paper of the whole field of category theory.

Only in the 1960s, after meeting graduate student Bill Lawvere, did Mac Lane come to believe category theory could be a foundation for all mathematics. Even then, precisely because of all the concrete mathematics that had gone into developing his ideas, Mac Lane insisted this, and any foundation for mathematics, must be seen as "proposals for the organization of mathematics" (Mac Lane, 1986, p. 406). The optimal organization (i.e. the optimal foundation) will change as mathematics develops, and will help advance those developments. He warned that excessive faith in any "fixed foundation would preclude the novelty which might result from the discovery of new form" (Mac Lane, 1986, p. 455).

2. Mathematics as a love of wisdom

Let us come to cases with one paradigmatically philosophical question, and one paradigmatically mathematical. For Mac Lane these questions are inseparable:

- Q_1 What are mathematical objects, and how do we come to know them?
- Q_2 What are solutions to a Partial Differential Equation (PDE), and how do we come to know them?

For Mac Lane Q_2 can only be a specific case of Q_1. For him, as for Aristotle, basic questions of the special sciences *are* philosophy. They cannot *not* be philosophy.

Let us be clear: A mathematician can learn a textbook answer to Q_2 without ever asking for a philosophy behind it. In just the same way, a philosopher can learn the currently received answers to Q_1 from philosophy books, without ever asking about live mathematics. Admittedly the math textbook answers will be more stable over time than the philosophically received answers. But that is not important. For Mac Lane, both of those ways of learning are failures of understanding. They are failures of *philosophy*. For him, an answer to either one of those questions can only be valuable to the extent that you can see what it is *good for*—for Mac Lane that cannot be either a purely technical mathematical question or a purely academic philosophic one. Think back to his work in World War II.

Mathematicians speak of solutions to PDEs in many ways:

- Smooth (or, sufficiently differentiable) function solutions.
- Symbolic solutions.
- Generalized function solutions (of various kinds…).
- Numerical solutions….

These different senses of solutions are sought in very different ways. There are well understood relations between them, but the relations are not all obvious and in particular cases they may be quite difficult, and important, to find.

Mac Lane's war work certainly involved relating different kinds of solutions to PDEs. Even when an equation has a known exact solution by an easily specified smooth function, applying it also requires numerical solutions. The worker has to choose which aspects are best handled in theory, so as to direct and optimize the calculations, and when best to leave theory and begin calculating. Those choices are rarely textbook work. They are often not clear cut at all. They require exactly what Aristotle called the wisdom of the craft master. Namely, they require grasping the *why* of each kind of solution. They require knowing not only the technical definition of each kind of solution, but *what good* each one is, and especially *the good* of their relations to one another.

The craft master, having wisdom, knows the whys, can teach them, and supervise work with them.

As I write this, I imagine some practice-minded philosopher challenging: "How are philosophies like logicism, formalism, and intuitionism going to help anyone solve or apply a PDE?" Indeed. This is why Mac Lane so often deprecates those philosophies. But just to give one example, Mac Lane argued that formalism in Hilbert's hands was a step towards programmable computers. See Section 3. Those unquestionably help solve PDEs.

3. The philosophy of mathematicians in 1930s Göttingen

> *Wir dürfen nicht denen glauben, die heute mit philosophischer Miene und überlegenem Tone den Kulturuntergang prophezeien und sich in dem Ignorabimus gefallen. Für uns gibt es kein Ignorabimus, und meiner Meinung nach auch für die Naturwissenschaft überhaupt nicht. Statt des törichten Ignorabimus heiße im Gegenteil unsere Losung: Wir müssen wissen – wir werden wissen!*
>
> We must not believe those, who today, with philosophical bearing and deliberative tone, prophesy the fall of culture and accept the ignorabimus. For us there is no ignorabimus, and in my opinion none whatever in natural science. In opposition to the foolish ignorabimus our slogan shall be: We must know – we will know! (Hilbert, 1930, p. 385)[4]

Hilbert's conclusion, *Wir müssen wissen – wir werden wissen!*, is engraved on his tomb in Göttingen.

[4] Before we try to defend or defeat Hilbert's slogan as an assertion in academic epistemology, it is in fact the most important statement in philosophy of mathematics of the past 150 years. I use the translation by Ewald (2005, p. 1164). Hilbert's address was broadcast on the radio and a recording is available at math.sfsu.edu/smith/Documents/HilbertRadio/HilbertRadio.mp3.

Mac Lane arrived in Göttingen just at the time Hilbert was promoting this slogan. I do not recall Mac Lane quoting it in lectures or conversations. He did not have to quote Hilbert. Everything Mac Lane said illustrated this faith. As you can see in Mac Lane (1986; 2005), or McLarty (2007), Mac Lane followed Hilbert's mathematizing scientific optimism, rather than the specific finitist program (often called "formalism," though not by Hilbert) of Hilbert's famous *On the Infinite* (1926). Mac Lane did see specific, productive mathematical value in that program though, and rejected a criticism of it by Freeman Dyson (Mac Lane, 1995).

Dyson supposed Hilbert seriously meant to reduce all mathematics to formal reasoning, and said "the great mathematician David Hilbert, after thirty years of high creative achievement[...] walked into a blind alley of reductionism." Specifically, Dyson claimed that Hlbert "dreamed of" formalizing all mathematics, solving the decision problem for this formal logic, and "thereby solving as corollaries all the famous unsolved problems of mathematics." Mac Lane replied:

> I was a student of Mathematical logic in Göttingen in 1931-1933, just after the publication of the famous 1931 paper by Gödel. Hence I venture to reply [...]. Hilbert himself called this 'metamathematics.' He used this for a specific limited purpose, to show mathematics consistent. Without this reduction, no Gödel's theorem, no definition of computability, no Turing machine, and hence no computers [...]. Dyson simply does not understand reductionism and the deep purposes it can serve.

Mac Lane gives a concise expert review of these issues and places them in the context he knew at the time they arose. He insists Hilbert did not tie the slogan "we must know, we will know" to the decision problem. Rather, Mac Lane says, "[Hilbert] held that the problems of mathematics can all ultimately be solved" without supposing metamathematics will do it. Full disclosure: I admit that after long consideration, drawing on Sieg (1999; 2013), I myself am unsure exactly how Hilbert and/or Bernays intended their work on the decision problem at various times. But however that may be, Mac Lane understood Hilbert this way. And this is Mac Lane's own far from reductionist faith, while recognizing reductionist methods for what they actually have achieved in mathematics.

Mac Lane learned a lot in frequent discussions with Bernays. But for now I have to say I see no larger trace of those discussions in Mac Lane's philosophy than is found in the letter on Dyson. Mac Lane's book on philosophy of mathematics is titled *Mathematics: Form and Function*. But this is clearly "form" as Mac Lane learned about it from talking with Weyl and studying under Geiger (McLarty, 2007). It does not refer to formalism in any sense related to Bernays. Or, at least, so it seems to me. The reader is encouraged to go, read, and disagree if they see something else.

4. Naturalism

The decisive feature marking Penelope Maddy's Second Philosopher as a *naturalist* is that:

> [She] sees fit to adjudicate the methodological questions of mathematics—what makes for a good definition, an acceptable axiom, a dependable proof technique?—by assessing the effectiveness of the method at issue as means towards the goals of the particular stretch of mathematics involved. (Maddy, 2007, p. 359)

Lots of mathematicians, and essentially all leading mathematicians, do the same.[5]

The unusual thing about Mac Lane in this regard is that he was explicitly tasked by the US government to evaluate mathematics research and teaching methods in classified reports during World War II and publicly after that.[6] Those reports were explicitly directed to various different specific short-term and long-term goals. All the variety he saw, and dealt with, left Mac Lane ever more deeply impressed with the actual unity of the whole.

Precisely that background, along with his experience as Chair of the Chicago Mathematics Department, made Mac Lane diverge from another feature of Maddy's Second Philosopher:

> All the Second Philosopher's impulses are methodological, just the thing to generate good science [. . .]. Maddy, 2003, p. 98

All Mac Lane's impulses aim at producing good science and for this reason they are not *all* methodological.

Mac Lane, like Aristotle, knows methods alone generate no science. Besides evaluating methods of reaching goals, at least some mathematicians must evaluate goals. For Aristotle, those should be the craft masters, the wise. In his vivid words: "the wise should not accept orders but give them; nor should they be persuaded, but the less wise should" (Metaphysics 982a). We will see, though, Mac Lane is less focused on command than that. He inclines more

[5] If by axioms Maddy means specifically axioms of set theory then few mathematicians ever learn those, let alone adjudicate them. Mac Lane is famously among those few.

[6] See Mac Lane (1967; 1989) and Fitzgerald and Mac Lane (1977) and Steingart (2011).

to another passage: "those who are more accurate and more able to teach about the causes are the wiser in each branch of knowledge" (Metaphysics 982a).

Because of his broad experience, especially evaluating both methods and goals for mathematics, Mac Lane cannot agree that "the goal of philosophy of mathematics is to account for mathematics as it is practiced, not to recommend reform." (Maddy, 1997, p. 161) Just sticking to the mathematician philosophers we have already named: Hilbert, Weyl, and Mac Lane all knew reform is integral to mathematical practice. You cannot separate reform from practice if you try. And all three made explicitly philosophic arguments for their recommended reforms along with more technically mathematical ones.[7] This is important for philosophy of mathematics.

The paradigm case for anti-revisionism in philosophy of mathematics is Brouwer's intuitionism. Brouwer is by far the favorite illustration of a revisionist, and is the sole example that the *Stanford Encyclopedia of Philosophy* discusses under anti-revisionism in the article "Naturalism in the Philosophy of Mathematics" (Paseau, 2016):

> The mathematician-philosopher L.E.J. Brouwer developed intuitionistic mathematics, which sought to overthrow and replace standard ('classical') mathematics.

So it is important for philosophers to understand that the problem with Brouwer, according to all our exemplars Hilbert, Weyl, and Mac Lane, is not that he had philosophical motives. It is that he was wrong. Actually, for Mac Lane, Brouwer's philosophy was at best wrong. At worst it was "pontifical and obscure" (Mac Lane, 1939).

[7] Hilbert had sweeping success with his reforms. Among many philosophic works by and on him see Hilbert (1923; 1930). Weyl (1918) advocated what Weyl took to be Brouwer's philosophy, while Weyl (1927; 1949) trace his eventual, regretful conclusion that in fact Hilbert was right about this and Brouwer wrong.

Immediately upon completing his doctorate in mathematics at Göttingen, Mac Lane put a philosophy article in *The Monist* (Mac Lane, 1935). Fifty years later he wrote a book describing, as he told me, what he wanted philosophers to know about math (Mac Lane, 1986). There he asks about the large array of mathematics he surveyed: "How does it illuminate the philosophical questions as to Mathematical truth and beauty and does it help to make judgements about the direction of Mathematical research?" (Mac Lane, 1986, p. 409) There is a reason he puts these questions together.

He asks about mathematical truth and beauty knowing very well that few mathematicians want to pursue the question seriously, and knowing philosophers who speak of it rarely know much of the wealth. For Mac Lane both of those are failures of understanding and they are nothing he means to promote.[8] He means to promote mathematically informed philosophic pursuit of the question of mathematical truth and beauty. And so he does of the question on the direction of research. He seriously means to promote philosophic thought on that. Of course he does not see philosophic thought as the sole preserve of those with philosophy degrees. No more does he see philosophy of math as the sole preserve of those with math degrees. Mathematics for Mac Lane, when pursued with full awareness of its worth, is philosophy.

---

[8] À propos, I consider Edna St. Vincent Millay's poem "Euclid alone has looked on Beauty bare" incredibly true to its topic, despite that she apparently studied no mathematics beyond school textbooks based on bits of Euclid's *Elements*.

Bibliography

Browder, F.E., 1976. The relevance of mathematics. *The American Mathematical Monthly* [Online], 83(4), pp.249–254. Available at: https://doi.org/10.2307/2318212 [visited on 9 November 2020].

Browder, F.E. and Mac Lane, S., 1978. The relevance of mathematics. In: L.A. Steen, ed. *Mathematics Today: Twelve Informal Essays*. New York [etc.]: Springer-Verlag, pp.323–350.

Eilenberg, S. and Mac Lane, S., 1945. General theory of natural equivalences. *Transactions of the American Mathematical Society* [Online], 58, pp.231–294. Available at: <https://www.ams.org/journals/tran/1945-058-00/S0002-9947-1945-0013131-6/S0002-9947-1945-0013131-6.pdf> [visited on 29 October 2020].

Ewald, W., 2005. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics. Vol. 2*. Oxford: Clarendon Press.

Fitzgerald, A. and Mac Lane, S., eds., 1977. *Pure and Applied Mathematics in the People's Republic of China: A Trip Report of the American Pure and Applied Mathematics Delegation*. Washington, D.C: National Academy of Sciences.

Hilbert, D., 1923. Die logischen Grundlagen der Mathematik. *Mathematische Annalen* [Online], 88, pp.151–165. Available at: <https://eudml.org/doc/158981> [visited on 9 November 2020].

Hilbert, D., 1926. Über das Unendliche. *Mathematische Annalen* [Online], 95, pp.161–190. Available at: <https://eudml.org/doc/159124> [visited on 9 November 2020].

Hilbert, D., 1930. Naturerkennen und Logik. *Die Naturwissenschaften* [Online], 18(47-49), pp.959–963. Available at: https://doi.org/10.1007/BF01492194 [visited on 10 November 2020].

Mac Lane, S., 1935. A Logical Analysis of Mathematical Structure: *Monist* [Online], 45(1), pp.118–130. Available at: https://doi.org/10.5840/monist19354515 [visited on 10 November 2020].

Mac Lane, S., 1939. Symbolic logic (Club topics). *The American Mathematical Monthly* [Online], 46(5), pp.289–296. Available at: <https://www.jstor.org/stable/2303907> [visited on 10 November 2020].

Mac Lane, S., 1946. Review of E. T. Bell The Development of Math- ematics (New York, McGraw-Hill, 1945). *The American Mathematical Monthly* [Online], 53(7), pp.389–390. Available at: https://doi.org/10.2307/2305857 [visited on 10 November 2020].

Mac Lane, S., 1967. The future role of the federal government in mathematics. *The American Mathematical Monthly* [Online], 74(1 (Part 2: Fiftieth Anniversary Issue)), pp.92–100. Available at: https://doi.org/10.2307/2314871 [visited on 10 November 2020].

Mac Lane, S., 1986. *Mathematics: Form and Function*. New York [etc.]: Springer-Verlag.

Mac Lane, S., 1989. The Applied Mathematics Group at Columbia in World War II. In: P.L. Duren, R. Askey, H.M. Edwards and U.C. Merzbach, eds. *A Century of Mathematics in America. Part III*, *History of mathematics* 3. Providence, R.I: American Mathematical Society, pp.495–515.

Mac Lane, S., 1995. 'A Matter of Temperament'. *New York Review of Books* [Online], 45(15), p.56. Available at: <https://www.nybooks.com/articles/1995/10/05/a-matter-of-temperament/> [visited on 10 November 2020].

Mac Lane, S., 1997. Requiem for the Skillful. *Notices of the American Mathematical Society*, 44(2), pp.207–208.

Mac Lane, S., 2005. *A Mathematical Autobiography*. Wellesley, Mass: A K Peters.

Mac Lane, S. and Schilling, O.F.G., 1939. Zero-dimensional branches of rank one on algebraic varieties. *Annals of Mathematics* [Online], 40(3), pp.507–520. Available at: https://doi.org/10.2307/1968935 [visited on 10 November 2020].

Mac Lane, S. and Schilling, O.F.G., 1940. Normal algebraic number fields. *Proceedings of the National Academy of Sciences* [Online], 26(2), pp.122–126. Available at: https://doi.org/10.1073/pnas.26.2.122 [visited on 10 November 2020].

Maddy, P., 1997. *Naturalism in Mathematics*. Oxford: Clarendon Press.

Maddy, P., 2003. Second Philosophy. *Journal of the Indian Council of Philosophical Research* [Online], 20, pp.73–106. Available at: <http://www.socsci.uci.edu/~pjmaddy/bio/2ndphilosophy.pdf> [visited on 10 November 2020].

Maddy, P., 2007. *Second Philosophy: A Naturalistic Method*. Oxford [etc.]: Oxford University Press.

McLarty, C., 2007. The Last Mathematician from Hilbert's Gottingen: Saunders Mac Lane as Philosopher of Mathematics. *The British Journal for the Philosophy of Science* [Online], 58(1), pp.77–112. Available at: https://doi.org/10.1093/bjps/axl030 [visited on 10 November 2020].

McLarty, C., 2020. Saunders Mac Lane: From Principia Mathematica through Göttingen to the Working Theory of Structures. In: E. Reck and G. Scheimer, eds. *The Prehistory of Mathematical Structuralism* [Online]. Oxford University Press, pp.215–237. Available at: https://doi.org/10.1093/oso/9780190641221.003.0009 [visited on 10 November 2020].

Paseau, A., 2016. Naturalism in the philosophy of mathematics. *The Stanford Encyclopedia of Philosophy* [Online]. winter 2016. Stanford: Metaphysics Research Lab, Stanford University. Available at: <https://plato.stanford.edu/archives/win2016/entries/naturalism-mathematics/> [visited on 10 November 2020].

Sieg, W., 1999. Hilbert's Programs: 1917–1922. *Bulletin of Symbolic Logic* [Online], 5(1), pp.1–44. Available at: https://doi.org/10.2307/421139 [visited on 10 November 2020].

Sieg, W., 2013. *Hilbert's programs and beyond*. Oxford [etc.]: Oxford University Press.

Skowron, B., 2020. *Was Saunders Mac Lane a platonist?*

Steingart, A., 2011. *Conditional inequalities : American pure and applied mathematics, 1940-1975* [Online]. Thesis. Massachusetts Institute of Technology. Available at: <https://dspace.mit.edu/handle/1721.1/84367> [visited on 10 November 2020].

Weyl, H., 1918. *Das Kontinuum: Kritische Untersuchungen über die Grundlagen der Analysis*. Leipzig: Verlag von Veit&Comp.

Weyl, H., 1927. *Philosophie der Mathematik und Naturwissenschaft*, *Handbuch der Philosophie*. München; Berlin: Druck und Verlag von R. Oldenbourg.

Weyl, H., 1949. *Philosophy of Mathematics and Natural Science* (O. Helmer, Trans.). Princeton: Princeton University Press.