

Stairway to Heaven

The abstract method and levels of abstraction in mathematics

Jean-Pierre Marquis*
Département de philosophie
Université de Montréal
Montréal, Canada
Jean-Pierre.Marquis@umontreal.ca

Abstract

In this paper, following the claims made by various mathematicians, I try to construct a theory of levels of abstraction. I first try to clarify the basic components of the abstract method as it developed in the first quarter of the 20th century. I then submit an explication of the notion of levels of abstraction. In the final section, I briefly explore some of main philosophical consequences of the theory.

1 Introduction

One of the distinctive features of 20th century mathematics is the rise and systematic use of the abstract method. The latter changed dramatically both the very object of mathematics and its methods. Mathematics suddenly referred to groups, rings, fields, metric spaces, topological spaces, vector spaces, Banach spaces, manifolds, lattices, categories, etc. and it relied on the abstract method for its presentation and development. The abstract method is at the core of what came to be known as ‘modern mathematics’. As a by-product, the abstract method opened the door to the idea that there are levels of abstraction in mathematics: there are parts of mathematics that are more abstract than others. Mathematicians themselves speak that way spontaneously. Here are three representative passages found in textbooks:

However, as we began to think about the task at hand, ..., we decided to organize the material in a manner quite different from that of our earlier books: a separation into two levels of abstraction... [12, p. xii]

*

This text attempts a different approach, letting the abstract concepts emerge gradually from less abstract problems about geometry, polynomials, numbers, etc. This is how the subject evolved historically. This is how all good mathematics evolves — abstraction and generalization is forced upon us as we attempt to understand the “concrete” and the particular. [26, p. 3]

This book is intended as an introduction to that part of mathematics that today goes by the name of abstract algebra. The term “abstract” is a highly subjective one; what is abstract to one person is very often concrete and down-to-earth to another, and vice versa. In relation to the current research activity in algebra, it could be described as “not too abstract”; from the point of view of someone schooled in the calculus and who is seeing the present material for the first time, it may very well be described as “quite abstract”. ([11, pp. 1 - 2])

The quotes state without any further explanation that there are levels of abstraction. The second quote goes somewhat further and asserts that the levels of abstraction arose historically and that abstraction was ‘forced’ upon mathematicians. It seems that for these quoted mathematicians, the proposal that there are ideas, theories, concepts that are more abstract than others is obvious and does not need any further clarification or discussion. It is as if the claim is based on a common, clear and widespread mathematical experience.

The last quote has a different flavor. The author suggests that “being abstract” is a subjective term. As I hope to show in this paper, I think that this claim is plain wrong. I submit that the author is confusing familiarity with a subject with its abstract character.

My goal in this paper is to clarify what mathematicians might mean by levels of abstraction. I believe that the claim can be taken literally and that it is not a subjective matter. It is, as I will argue, a consequence of the abstract method itself. The paper will therefore proceed as follows. In the next section, I will provide an overview of what I take to be the main components and properties of the abstract method in mathematics. I will then propose a definition of levels of abstraction based on some of the elements of that method. Finally, in the last section, I will explore some of the philosophical consequences of this approach¹.

2 The abstract method: basic components

I have elsewhere presented some of the historical roots and components of the abstract method in mathematics. (See [20].) It is nonetheless necessary to recapitulate and develop from a different perspective the conceptual ingredients of my analysis.

¹I urge readers to consult the paper by Hourya Benis Sinaceur for a slightly different point of view on the matter. See [2].

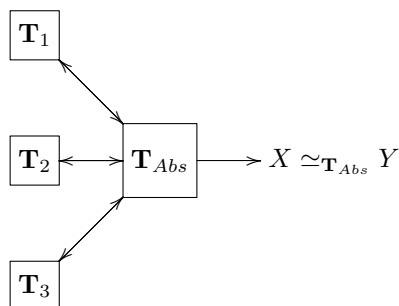
Informally, the abstract method can be described as follows. It is made up of three different components². The first component is the identification of invariant properties in theories which seem to be essentially incompatible at first sight; 2. The second component is the systematic ignorance of some of the specific properties of the objects of each theory (this is the step that justifies the name “abstraction” since one has to “subtract” key properties of the objects of the given theories) and the presentation in an appropriate language of the invariant properties; 3. The last component is the identification of a criterion of identity on the basis of the properties chosen and, thus, the introduction of a new type of entities, the abstract entities, and the exploration of the properties of these new entities.

Let me reformulate these components, using a somewhat more formal set-up. The abstract method comprises three basic parts that are related systematically and thus make a whole. The first part is what I call a *domain of significant variation*. It is made up of at least three essentially different theories, denoted by $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$. The second part is a *method of presentation* together with a *method of development*, which was historically identified with the axiomatic method and which yields a *new* theory, denoted by \mathbf{T}_{Abs} . The third part is a *new criterion of identity* for the emerging entities, $X \simeq_{\mathbf{T}_{Abs}} Y$.

The links between the parts are as follows. Between the domain of significant variation and the method of presentation and development, we have what I will call the *formalist stance* and it consists in identifying the invariant properties within the domain of variation and forgetting what the symbols used actually refer to. Between the method of presentation and the criterion of identity, we have the *extraction of a new criterion of identity* together with the understanding of how it applies, that is using the new criterion in a way that the following form of Leibniz’s principle applies:

$$\text{If } P(X) \text{ and } X \cong_{\mathbf{T}_{Abs}} Y, \text{ then } P(Y).$$

Here is a diagram capturing the components and their relations.



One simple historical example might be useful before I explain the diagram

²The temporal order of these components can vary. Indeed, historically one finds various combinations, although most cases seem to follow the order of our presentation, for reasons that are not very hard to understand. It is, however, interesting to note that there are cases where the development follows a different path.

and thus the method in more details. One of the very first cases of an extraordinarily successful abstraction in the history of modern mathematics is certainly that of metric spaces, introduced by Fréchet around 1906 in the context of functional analysis³. At the turn of the century, mathematicians thought of manifolds as subspaces of spaces of real or complex points. Fréchet, for instance, was dealing with the usual manifolds, namely $\mathbb{R}, \mathbb{R}^2, \dots, \mathbb{R}^n, \dots, \mathbb{C}, \mathbb{C}^2, \dots, \mathbb{C}^n$ together with functions between them on the one hand, and infinite-dimensional functional spaces together with operators between them on the other hand. In his thesis, Fréchet gives the following four examples of functional spaces. (See [8] or [27].)

1. Let J be a closed interval of the real line \mathbb{R} and consider the space \mathbb{R}^J of continuous functions $f : J \rightarrow \mathbb{R}$. A metric on \mathbb{R}^J is defined by

$$d(f, g) = \max(|f(x) - g(x)|) \quad \forall x \in J.$$

2. Consider the space $E_\infty = \mathbb{R}^\mathbb{N}$ of infinite sequences $x = (x_1, x_2, \dots)$ of real numbers. A metric on E_∞ is given by

$$d(x, y) = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{|x_n - y_n|}{1 + |x_n - y_n|}.$$

3. A space of parametrized curves in \mathbb{R}^3 with the standard Euclidean metric between points. Using the latter, Fréchet defines a metric between the curves.
4. Finally, let A be a complex plane region whose boundary consists of one or more contours. Let $\{A_n\}$ be a sequence of bounded regions such that $A_n \subset \text{int}(A_{n+1})$ and $A_n \subset \text{int}(A)$ and such that any given bounded region in the interior of A is in the interior of some A_n for n sufficiently large. Consider the space $\{f : \text{int}(A) \rightarrow \mathbb{C} \mid f \text{ is holomorphic}\}$ and let

$$M_n(f, g) = \max(|f(z) - g(z)|) \quad \text{when } z \text{ is in the closure of } A_n.$$

The metric between two such functions is then defined by

$$d(f, g) = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{M_n(f, g)}{1 + M_n(f, g)}.$$

Although I haven't described the third example in details, I hope that it is nonetheless clear that these examples are radically different from one another and, perhaps even more so, from the spaces of points \mathbb{R}^n and \mathbb{C}^n . I submit that if we did not know about the metric involved in each case, we might not see the invariant features involved. Indeed, we are accustomed to attribute certain

³It is known that Fréchet knew about the case of groups and that it provided at least guidelines and a model of what could be achieved by moving up the ladder of abstraction.

properties to real functions: continuity, differentiability, roots, maximum, minimum, etc., we represent the graph of a function as a one-dimensional path in the codomain, thus as something that necessarily has a length, we think of a real function as a systematic relation of dependence between two or more properties, as a quantity that varies according to a certain pattern or whose variation depends on another variation. A function is essentially thought of as being dynamic. The four examples given by Fréchet are of this kind. A (real) point is, well, a point. It has none of the properties of a function. Thus, the properties of the elements of \mathbb{R} and even \mathbb{R}^n are incommensurable with the properties of the elements of a functional space. I want to insist on the fact that given the properties of functions and given that we think of functions with their properties, it is hard to conceive of a *space* of functions, that is treating the latter as being points. It is as if we were trying to think of the properties of functions and forget about them at the same time. Of course, as soon as we have succeeded in thinking of them as spaces, we stumble upon what is certainly seen as being the main difference between these spaces and the usual spaces of points: the examples given above are *infinite* dimensional. Thus, we also have two different types: finite dimensional spaces on one side and infinite dimensional spaces on the other.

I want to emphasize the fact that in order to see the invariant features between the different theories, one has to *forget* or *ignore* essential aspects of the objects and their properties involved. One has to ignore key properties of functions, of series, of the complex numbers, etc. One of the ways to succeed this operation is to concentrate on the formalism, the symbols and the operations on these symbols. This posture is what I call the *formalist stance*. Historically, it is easy to find mathematicians emphasizing the necessity of this stance. Here is one clear illustration:

In the following an attempt is made to present Galois theory of algebraic equations in a way which include equally well all cases in which this theory might be used. Thus we present it here as a direct consequence of the group concept illuminated by the field concept, as a *formal structure completely without reference* to any numerical interpretation of the elements used. (Weber, 1893, p. 521 quoted by [4, p. 36])[my emphasis]

Weber says clearly that in this case, one has to forget what the symbols stand for and concentrate on the formal structure, on the relations and operations that exist between the objects. The same posture applies to the case of metric spaces: one has to forget the idiosyncrasies of the elements of the various spaces, what the symbols refer to and focus on the invariant structural features that these domains share. Notice that if one does not move to the next step, namely the extraction of an appropriate criterion of identity for new objects, then the only coherent philosophical position left is a variant of formalism.

Thus, the next step consists in finding the proper language and the right properties to present the invariant features involved in the domain of significant variation. Historically, the axiomatic method seemed to be perfect for the

purpose. So much so that mathematicians very often *identify* the axiomatic method with the abstract method. I will come back to this point later.

Once the invariant properties have been identified, it is only natural to investigate the logical consequences of these properties and verify that it is possible to develop relevant and useful mathematics from them. Then and only then a criterion of identity for the new, abstract entities can be extracted. In other words, the criterion of identity can not be given *a priori* but is derived from the theory⁴.

Let us now look at these components and their relations one by one.

2.1 Domain of significant variation

Thus, owing to its abstract foundation, modern algebra winds a unifying band of method *around essentially different things* and in this way contributes to the required organic and systemic unification of mathematics. [10, pp. 18-19][my emphasis]

In the foregoing diagram, I have put *three, essentially different* theories. I claim that these two elements, that is the number of theories and the fact that they differ in an essential manner, are necessary and sufficient for the abstract method to be used in a particular case.

The number three is not accidental. It seems to play a key role in the method itself. I have to emphasize that I present this claim as an empirical hypothesis. Although with hindsight, there are good reasons to believe that the number of theories is indeed necessary, it came to me *a posteriori*, by examining the historical record. Let me start with a long quote from an historian of mathematics who happened to come to the same conclusion for a specific important case.

It is obviously true that the concept of a permutation group derived from the development of the theory of algebraic equations and from Galois theory. But this development, associated with the names of A. Vandermonde, J.-L. Lagrange, P. Ruffini, N.H. Abel, E. Galois, J.-A. Serret and C. Jordan, is just *one* of the historical roots of group theory. The mathematical literature of the nineteenth century, and especially the work of decisive importance for the evolution of the abstract group concept written at the century's end, make it abundantly clear that that development had *three* equally important historical roots, namely, the theory of algebraic equations, number theory, and geometry. Abstract group theory was the result

⁴I have to come back to my earlier remark concerning the temporal order involved here. It is tempting to read the diagram from left to right and to think that there is a chronological order in that reading. Although it is in most cases correct, there are counterexamples to that reading, e.g. Boolean algebras or lattices. What seems to be correct is the step from the presentation of the theory to the extraction of the criterion of identity. As far as I know, that step is always taken in that order, although once the method has been well-understood, the delay between the presentation of the theory and the extraction of the criterion of identity can be very short.

of a gradual process of abstraction from implicit and explicit group-theoretic methods and concepts involving the interaction of its three historical roots. I stress that my inclusion of number theory and geometry among the sources of causal tendencies for the development of abstract group theory is grounded in the historical record and is not the result of a backward projection of modern group theoretic thought. ([30, p. 16])

This is but one case. What is striking is that when one looks at the development of ring theory and field theory, one finds a similar situation. (See, for the case of ring theory, [14, 5] and, for the case of field theory, [15].) As we have seen, Fréchet took the time to present four different function spaces, as well as the usual manifolds. What is even more striking is that the abstract method, conceived as genuine method in algebra, arose only after there were three different cases where the method turned out to be fruitful: groups, rings and fields.

We can also find *theoretical* reasons that makes the claim *a priori* plausible. Indeed, when there are two theories, it seems reasonable to think that one will either consider a simple generalization or an analogy between them. When there are at least three essentially different theories, then the essential differences make it difficult to believe that a simple generalization is possible and the fact that there are at least three brings us immediately outside the realm of analogies.

The idea of a *significant* variation can be expressed more formally as follows. Each theory \mathbf{T}_i has its own criterion of identity $a \simeq_{\mathbf{T}_i} b$. Therefore, we are dealing with (at least) three distinct criteria of identity satisfying Leibniz's principle:

1. In \mathbf{T}_1 , if $P(\mathbf{a})$ and $\mathbf{a} \simeq_{\mathbf{T}_1} \mathbf{b}$, then $P(\mathbf{b})$.
2. In \mathbf{T}_2 , if $Q(\mathbf{f})$ and $\mathbf{f} \simeq_{\mathbf{T}_2} \mathbf{g}$, then $Q(\mathbf{g})$.
3. In \mathbf{T}_3 , if $R(\alpha)$ and $\alpha \simeq_{\mathbf{T}_3} \beta$, then $R(\beta)$.

These are incompatible in the sense that one would not, for instance, necessarily attribute certain properties, say P_1, \dots, P_n to the objects α_i of \mathbf{T}_3 and conversely, one would not apply the properties R_1, \dots, R_k to the objects \mathbf{a}_i of \mathbf{T}_1 . In fact, in some cases, even if one would try to apply, let us say P_i to α_j , one would in some cases consider $\neg P_i$ instead. Thus, in order to abstract properly, it is necessary to be able to ignore, forget or subtract some of the properties of the objects in the given theories. Furthermore, the properties that have to be subtracted are not necessarily the same in each case. I submit that this is a real cognitive stumbling block to abstraction in mathematics. A clear empirical hypothesis could be formulated and be tested on students learning vector spaces, group theory, topology or any other similar theory.

In other words, the epistemic attention has to shift from certain pregnant features of the objects under study to the invariant elements involved across the various theories. The latter are usually operations or relations between these objects. Again, I want to emphasize that at this stage, one could stop and either work in a purely formal fashion or consider that one is doing algebra

in the classical sense of that word, that is working on generalized arithmetic operations. As we have indicated elsewhere (see [20]), it is easy to find this attitude in the 19th century, especially when one looks at the development of algebra during that period. The situation changes radically once it is clear that it is possible to consider a new type of entities supporting these properties and relations. Once the invariant features have been identified, recognized as such and as being the same in all cases, then a criterion of identity can be formulated. The invariant component is from then on circumscribed clearly and independently of the original entities. These are seen to be instances of these new types and are studied as such, that is, there is a shift of attention from the old criterion of identity and its associated properties to the new criterion and its associated properties.

Very often, it is then possible to discover and construct new, unforeseen instances of these new abstract entities. Thus, the domain of variation can expand and is never fixed once and for all. In more philosophical terms, once the new types have been fixed, known examples become tokens of the type and new, unforeseen tokens can be constructed or discovered.

2.2 The axiomatic method and the abstract method

From an historical perspective, the abstract character of contemporary mathematics can be attributed, at least in part, to the axiomatic method. However, many mathematicians *identify* the abstract method with the axiomatic method. To wit:

The *abstract or postulational* development of these systems must then be supplemented by an investigation of their “structure.” ([17, p. 18], quoted also by [4, p. 258]) [My emphasis]

Mac Lane’s formulation is typical of the period: the abstract and the axiomatic seem to be interchangeable. This attitude still prevails today:

It is abstraction — more than anything else — that characterizes the mathematics of the twentieth century. There is both power and elegance in the axiomatic method, attributes that can and should be appreciated by students early in their mathematical careers and even if they happen to be confronting contemporary abstract mathematics in a serious way for the very first time. ([29, p. ix])

Again, the author writes as if the axiomatic method and the abstract method were the same thing. The axiomatic method is but one method of presentation and development of the invariant content of theories in the domain of variation. The axiomatic method played a key role in the construction of the abstract method, but the latter is not the same thing as the axiomatic method.

It is extremely easy to convince oneself that the axiomatic method has nothing to do, by itself, with the abstract. Indeed, one can use the axiomatic method to present a theory about a unique set of entities, as it was probably conceived to

present Euclidean geometry. One can even use the axiomatic method to present a physical theory, e.g. a theory about the concrete world. Thus, as such, it is hard to see how one can identify the abstract with the axiomatic.

The fact is, that axiomatic method has a long history and therefore acquired various, even incompatible, roles in mathematics and in foundational studies. Let us briefly recapitulate some of the functions of the axiomatic method.

Let us start with Hilbert:

..., Hilbert's own use of the axiomatic method involved, by definition, an acknowledgment of the conceptual priority of the concrete entities of classical mathematics, and a desire to improve our understanding of them, rather than a drive to encourage the study of mathematical entities defined by abstract axioms devoid of immediate, intuitive significance. ... All the concepts introduced in algebra derive their meaning, their justification and their properties from those of the systems of complex and real numbers, rather than the other way around. ([4, p. 170])

Here the axiomatic method is clearly separated from "abstract axioms". Corry claims that, for Hilbert, the axiomatic method had little to do with the abstract method and, what it had, it had only in a derivative manner. The key element, for Hilbert, is the desire to improve our understanding of concrete entities of classical mathematics and the axiomatic method becomes a tool for that very purpose.

For him [Hilbert], the real focus of interest lay in the interrelation among the various groups of axioms, rather than among the individual axioms across groups. For him, the groups corresponded to the isolable basis of our spatial intuition and the main task of his axiomatic approach was to show the way in which they logically interacted to create the body of geometric knowledge. ([6, p. 5])

Often, when a domain of inquiry is axiomatized, the goal is to lay bare the underlying logical structure of the system of definitions and proofs, because the field has become a complete conceptual mess. Clarity and order are brought by a clean and lean axiomatic development. The structure of logical dependence between 1) primitive notions and the defined ones are exhibited and 2) the axioms and the deduced propositions is displayed openly. Thus, a structure of justification between the notions as well as between the propositions and based solely on logical relationships becomes transparent. As such, the axioms can be presented as constituting the foundations of the given domain. Hilbert's work in the foundations of geometry belongs to that category.

Some philosophers and mathematicians attribute an additional epistemic function to axioms in this context: the axioms should provide basic, self-evident truths upon which the whole given domain rests. Frege, for one, certainly

believed that axioms had to have this epistemic feature⁵. This desiderata is clearly not an ingredient of the abstract method. It is hard to convince oneself that the axioms of group theory, for instance, are self-evident.

After Hilbert and using some of Hilbert's tools, logicians and mathematicians turned their attention to axioms themselves. This became known as postulational analysis and was energetically developed in the USA by E.H. Moore and his students.

But in the case of E.H. Moore, his students at Chicago, and some other contemporary USA mathematicians, their study of the *Grundlagen* led to development of a point of view that diverged from Hilbert's in this significant yet subtle matter: they turned the analysis of systems of axioms into a field of intrinsic mathematical interest in which the requirements introduced by Hilbert oriented the research questions and afforded the main technical tools to deal with them. ([6, p. 5])

The function of the axiomatic method is now different. One wants to reduce the number of axioms, investigate their independence, the categoricity of the system, their (syntactic) coherence and, later the semantic coherence, completeness, etc. Clearly, the abstract method has, in itself, nothing to do with this usage of axioms.

Of course, axioms and the axiomatic method did play a key role in the rise of the abstract method. The axioms capture the invariant features of the theories under investigation. Once these have been identified, the axiomatic method allows for the systematic and rigorous development of the consequences of these features. One *could* use a different method of presentation of the invariant features. It depends on the linguistic means available. For instance, nowadays, it would be possible to use a graphical language to present a new theory by using what are called *sketches*. (See [1] for an introduction to the language of sketches as a method of presentation of theories.) Be that as it may, I think that the identification of the axiomatic method with the abstract method

⁵Needless to say, this epistemic standpoint was not and is certainly not adopted by all mathematicians and logicians using the axiomatic method this way. In fact, the whole spectrum of positions is likely to be found within the community of mathematicians and logicians. Contrast, for instance, the following claim made by Hölder

[the axioms of arithmetic] which are based on the fact that we take it as evident that there are certain processes, which, as we say, proceed according to determinate rules that are always performable in a definite way and in certain cases can be carried out repeatedly without end. (Hölder, quoted by Scanlan [24, 988].)

where Hölder mentions that we take certain things or processes as being evident, with the following claim made by his contemporary Peano

Logical questions thus become completely independent of *empirical* or *psychological*, questions (and, in particular, or the *problem of knowledge*) and every question concerning the *simplicity of ideas* and the *obviousness of facts* disappears. (Peano, quoted by Scanlan [24, 988].)

The distance between these two positions could hardly be greater.

introduces some confusion. Indeed, some mathematicians started to attribute to the axiomatic method virtues that belong to the abstract method. I will get back to this important point in the section on philosophical consequences.

2.3 Extracting a criterion of identity

I claim that the extraction of a criterion of identity has been the blind spot in the mathematicians' journey through the abstract method. Contemporary mathematicians nowadays suppose that the proper criterion of identity comes almost automatically with the right axiomatic theory. For, this is how we are taught modern mathematics: one learns the axioms, some of their most important consequences and immediately after, the proper notion of homomorphism for the structure given and the resulting notion of isomorphism for that structure. We now take for granted the last notion and its importance. Historically, there is a time lag between the identification of a certain mathematical structure and its proper criterion of identity. This is not surprising as such. What is somewhat surprising is the length of the delay between the identification of the properties and the proper criterion of identity. Two historical cases can be mentioned: the case of homeomorphism for topological spaces and the case of equivalence for categories.

Topology slowly emerged as a field at the beginning of the 20th century and the theory was axiomatized for the first time by Hausdorff in 1914⁶. However, the definition of homeomorphism is nowhere to be found in the original edition of the book, although the concept is clearly there, probably under the influence of Fréchet's work in which it is clearly identified. It is, however, identified as such by Hausdorff in his 1927 edition. The clear conception of the notion of homeomorphism and its importance for topology was written clearly and unambiguously by Kuratowski in 1921, thus almost ten years after Hausdorff's axiomatization. If one consider the roots of topology and looks at *analysis situs*, then one finds that it took far more than a decade to set the record straight, in fact, almost fifty years. According to G. H. Moore, "at different times and by different authors, at least four distinct concepts were identified in Euclidean spaces with those mappings under which topological properties were invariant." ([22, p. 333]) These four concepts are: the concept of injective continuous mapping, that of homeomorphism, the notion of deformation and, finally, the concept of diffeomorphism. What will certainly surprise a contemporary mathematician is that "it took decades for mathematicians to learn to distinguish clearly between them." [22, p. 334] These four different concepts were clearly seen as being different by the community at large in the 1930s only.

Another significant example is provided by category theory. (See [16] and [18] for more on the history of category theory.) In Eilenberg and Mac Lane's original paper published in 1945 ([7]), one does find the notion of isomorphism

⁶The axiomatization of topology went through various phases from 1914 until the late 1950's. This is an interesting case where the different axiomatic frameworks are proposed mainly for practical reasons, that is, the modifications suited the specific needs of various mathematicians at different times.

of categories and it was certainly assumed by Eilenberg and Mac Lane that the latter was the proper criterion of identity for categories. However, for reasons that would take us too far from our present concern, the notion of isomorphism of categories turned out to be inappropriate as a criterion of identity in the practice of category theory, in particular when one considers functor categories, e.g. categories of sheaves in the context of algebraic geometry. I hasten to add that although Eilenberg and Mac Lane considered functor categories in their original article, the latter played a purely auxiliary role in the original paper. This fact changed completely when the theory came in the brains of Grothendieck and Kan in the mid 1950s. Indeed, functor categories moved to the center stage from that moment on. Be that as it may, the proper criterion of identity for categories, namely the notion of equivalence of categories, was introduced explicitly as such by Grothendieck in his famous *Tohoku* paper published in 1957 ([9]), twelve years after Eilenberg and Mac Lane's original paper.

Our main point is simple: it is not until the proper criterion of identity has been identified and applied systematically that the theory acquires an autonomy, both epistemological and ontological. Notice also that it is the presence of a new criterion of identity that allows to say that we are indeed in the presence of a new type of abstraction, for as we have seen, the usage of the axiomatic method in itself does not entail the need of a new criterion of identity. In the case of a given classical domain with an given criterion of identity, the role of the axiomatic method is radically different than the one it plays in the context of the abstract method. The identification of the proper criterion of identity is of fundamental importance, since it allows to sift the properties of the resulting theory from the properties of the previous theories. In other words, it captures the process of abstraction itself.

I claim that we now have all the pieces laying in front of us. I have presented the bare bones of the abstract method that characterized modern mathematics. We are now ready to see how the abstract method leads to levels of abstraction.

3 The abstract method and levels of abstraction

Let us first examine the very idea of levels of abstraction in mathematics and start with a few paradigmatic cases.

I claim that the notion of topological space is more abstract than the concept of metric space. Informally, the idea is that there is more significant variation among topological spaces than there is among metric spaces. This informal claim rests in part on a series of precise mathematical results that are well-known.

First, every metric space give rises to a topological space in a canonical way. Second, every map between metric spaces, i.e. every continuous map that do not increase any pairwise distance⁷ is trivially a continuous map between the associated topological spaces. In other words, there is a canonical functor from the category of metric spaces to the category of topological spaces. However,

⁷The definition of maps between metric spaces is always a delicate matter. We have chosen a simple class of morphisms. The main point remains, no matter how we choose them.

it is also a well-known result that not all topological spaces are metrizable. Only topological spaces satisfying certain properties are metrizable. Finally, two different metric spaces can give rise to the same topological space.

Are these facts enough to conclude that the concept of topological space is more abstract than the concept of metric space? Notice that both categories are “just as big”, that is, it seems hopeless to exploit some cardinality condition in general. Being less abstract *cannot* merely be captured, for instance, by the fact that the extension of the concept D, e.g. being a metric space, is strictly included in the concept C, e.g. being a topological space. The structures involved, their properties and how they are related have to be used to make sense of the notion of levels of abstraction. Notice also that it seems inappropriate to simply say that the notion of topological space is more general than the notion of metric space. It is indeed, but from an epistemological point of view, to focus on generality is to miss the point of the conceptual difference between the two notions.

Let us go up the stairs of abstraction. I claim that the notion of locale is more abstract than the notion of topological space. (See [13, 23] for more on the theory of locales.) The reasons are essentially the same as those of the foregoing example. To see this, let us briefly recall the definition of a locale and certain basic facts about them.

We first have to start with the algebraic notion of *frame*. A *frame* is a partially ordered set with finite meets and arbitrary joins such that meets distribute over joins. A morphism of frames is a map of sets that preserves finite meets and arbitrary joins. Notice that a frame is automatically a Heyting algebra, but that frame homomorphisms need not preserve the Heyting implication. In fact, frames are sometimes identified with complete Heyting algebras. We can therefore consider the category of frames and frame homomorphisms. The category of *locales* is the opposite of the category of frames, that is it has the same objects but the morphisms go in the opposite direction.

To understand the link between locales and topological spaces, one has to consider the algebra of open sets of a topological space. It is easy to see that it is a frame. Thus, every topological space gives rise in a canonical way to a locale and every continuous map gives rise to a map of locales. However, the notion of locale, although closely related to the notion of topological space, yields a different mathematical theory of space. The crucial difference is that, in a very specific sense, the theory of locales is the theory of pointless topology. Whereas a topological space always has an underlying set of points, the basic notion in locale theory is the notion of open subspace, the notion of point being a special case of the latter. For instance, there are non-empty locales that are without points⁸. The key is of course the notion of point. In a category \mathbb{C} , a point of an object X of \mathbb{C} is given by a morphism from the terminal object 1 , assuming the latter exists, into X . In the category **Set** of sets and functions, there is a bijection between the elements of a set X and its points in this sense, for the

⁸Thus, it seems rather obvious that the notion of locale could be useful for mereology. As far as I know, no one has looked at the links between traditional mereological notions and the theory of locales.

terminal object of the category of sets is any singleton set. In the category **Top**, a point of a space X is a continuous morphism from the one point space 1 into X . The locale corresponding to the one point space 1 is of course the two element poset, in fact Boolean algebra $\mathbf{2}$ and, therefore, a point of a locale $\mathcal{O}(X)$ is a morphism $\mathbf{2} \rightarrow \mathcal{O}(X)$. Given that the category of locales is the opposite of the category of frames, points of locales correspond to frame homomorphisms $\mathcal{O}(X) \rightarrow \mathbf{2}$. As is well-known, there are non-atomic Boolean algebras, thus pointless locales.

As in the previous case, there are different (non-sober) topological spaces, that is non homeomorphic spaces that give rise to isomorphic locales⁹. Again, since there are pointless locales, the notion of locale is indeed also more general than the notion of topological space and not all locales are spatial¹⁰. I claim that in this case also, it would be insufficient to merely focus on the difference in generality between the concepts. There is something else, something epistemologically more important at work.

Here is an example of two notions that I believe are at the same level of abstraction: the notion of group and the notion of abelian group. I claim that they are just as abstract and that the property that separates them does not introduce a difference in their level of abstraction.

It would be easy to multiply the examples and explore various cases. (I encourage the reader to do so!) But our goal here is more modest. We only presented these examples to convince the reader that it seems plausible to say that a given mathematical notion is more abstract than another and I hope that these examples are plausible.

Let us sum up. We are now in the following position. First, there is circumstantial empirical evidence that mathematicians accept the idea that there are levels of abstraction in mathematics. Second, the abstract method, as it was developed in the 20th century, clearly opens the door to a procedure that can yield different levels of abstraction, provided that the latter notion makes sense at all. Third, there are cases of mathematical notions about which it seems reasonable to say that they differ in their levels of abstraction. It is another matter to say exactly what differing in levels of abstraction might mean. Assuming that we have sufficient evidence to conclude that there is a need for some sort of philosophical clarification of the notion of levels of abstraction, I present the following proposal.

Let \mathbb{T}_1 and \mathbb{T}_2 be two mathematical theories such that there is a canonical way to transform every model of \mathbb{T}_2 into a model of \mathbb{T}_1 . Given a model M of \mathbb{T}_2 , we denote the canonical model of \mathbb{T}_1 it gives rise to by $F(M)$. We say that \mathbb{T}_1 is *more abstract* than a theory \mathbb{T}_2 if

1. $(\forall M_1)(\forall M_2)(M_1 \cong_{\mathbb{T}_2} M_2 \Rightarrow F(M_1) \cong_{\mathbb{T}_1} F(M_2))$
2. $(\exists M_1)(\exists M_2)(F(M_2) \cong_{\mathbb{T}_1} F(M_1) \wedge M_1 \not\cong_{\mathbb{T}_2} M_2)$.

⁹This is not an entirely trivial result. It requires some fiddling around. One way to go to build the appropriate spaces can be found in [3].

¹⁰Informally, a locale is spatial if it arises from the lattice of open sets of a topological space. It is of course possible to give an intrinsic characterization.

3. The objects $F(M_2)$ and $F(M_1)$ are non trivial.

The first clause simply says that the theories are connected in a systematic fashion, i.e. the criteria of identity of \mathbb{T}_2 is preserved by the canonical transformation F . Notice that any functor between two categories of models automatically preserves isomorphisms. Hence, we have a specific framework in which this condition is easily satisfied. More importantly, the functor F has to be *canonical*. This is obviously a tricky expression and although it is used by mathematicians in numerous circumstances and occasions, it does not have a precise technical meaning. Nor will we try to give one to it¹¹. The second condition captures the idea that we are dealing with an abstraction process. As we have indicated in section 2.3, a criteria of identity allows one to determine the relevant properties of given objects. Thus, the condition stipulates that there is at least one property that allows one to distinguish between M_1 and M_2 but that when one moves to the canonical objects $F(M_1)$ and $F(M_2)$ arising from M_1 and M_2 respectively, it is no longer possible to distinguish them. In other words, the discriminating properties were abstracted from M_1 and M_2 . The third condition is simply there to block trivial cases where one would erroneously conclude that there is a new level of abstraction when, in fact, one has in some sense abstracted everything.

We can now assert that topology is more abstract than the theory of metric spaces and that the theory of locales is more abstract than topology. Our informal exposition above contains all the required ingredients to verify that the formal characterizations applies. I will leave to the reader the pleasure of exploring other cases and come to its conclusions.

4 Some philosophical consequences

This section will be sketchy and brief. I will barely touch upon some of the consequences of the foregoing proposal here.

4.1 Epistemological consequences of the abstract method

The axiomatic method is thus first a method of economy of thought, in that it allows to condense many different reasonings in one. (...) But the axiomatic method is also a method of discovery and of progress¹². (Weil, pp. 19-20) [My translation]

¹¹In practice, mathematicians know what it means to say that something is given canonically or that there is canonical construction or presentation of something. Very roughly, it means that the construction does not depend on a choice from a mathematician in the process. Whether this can be made more precise is an open question. See, however, [19] on the subject.

¹²“La méthode axiomatique est donc d’abord méthode d’économie de pensée, en ceci qu’elle permet de condenser en un seul plusieurs raisonnements différents. (...) Mais la méthode axiomatique est aussi méthode de découverte et de progrès.”

I believe this is another revealing case of a mathematician confusing the axiomatic method with the abstract method. I submit that it is really the abstract method that has the virtues expounded by Weil and that the axiomatic method is, as I have already indicated, a useful tool in the application and development of the abstract method. Thus, what characterizes the development of mathematics in the 20th century is the systematic usage of the abstract method and, as a consequence, the fact that mathematics has become more abstract in the foregoing precise sense. As Weil sees so clearly, the epistemological benefits of the abstract method are immense. First, many different results can now be proved in a uniform manner and from a common core. There is, to use a contemporary language, a form of “compression” of the mathematical information. One and the same proof now applies to significantly different mathematical domains and the proof exhibits the bare essentials involved in the result. This was also mentioned explicitly by Banach in the introduction of his book on functional analysis.

The aim of the present work is to establish certain theorems valid in different functional domains, which I specify in what follows. Nevertheless, in order not to have to prove them for each particular domain, I have chosen to take a different route...; I consider sets of elements about which I postulate certain properties; I deduce from them certain theorems, and I then prove for each particular functional domain that the postulates adopted are true for it. (Banach, quoted by Moore [21, 280])

That is a crystal clear exposition of the abstract method and its advantages.

Second, the abstract method is a method of discovery. Again, as such, the axiomatic method is not traditionally thought in those terms¹³. Certainly, philosophers never thought of the axiomatic method as a method of discovery. In the context of the abstract method, however, it makes perfect sense. Indeed, given (at least) three domains where there is significant variation, if indeed one can discover common traits that can be used to prove important features of the situations, the method can be seen as a method of discovery and, as the quote by Weil shows, it has been considered as such.

Last, but not least, Weil associates the method with the idea of progress. I cannot venture into the conception of progress underlying Weil’s claim, apart from the trivial claim that it is seen in a positive light. I will nonetheless speculate that progress in this particular context is linked to understanding. It can certainly be claimed that by identifying features that are abstracted from significantly different situations that nonetheless allow to prove important facts of these situations, one is led to conclude that *essential* properties and relations have been brought to the fore and that these essential features, these new concepts, allow us to understand why these results hold. For now, we have the bare essentials involved in these significantly different cases. Notice that, as another benefit of the method, one obtains a form of unity of mathematics, a

¹³See, however, the recent paper [25].

unity that is not so much the fact that all mathematical entities can be defined as sets, but rather as a working unity, as concepts that play important roles in different domains. Certain concepts arise in domains that are significantly different, sometimes in unexpected ways. It is of course their abstract character that makes this even possible.

4.2 Being abstract

Traditionally, the property of being abstract is opposed to the property of being concrete. Thus, an entity is either abstract or concrete, period. As such, the distinction is presented as being an ontological distinction. The usual strategy is to provide a way to characterize what it is to be concrete, e.g. being causally efficacious or having spatio-temporal coordinates, and then say that something is abstract if it is not concrete. It is well-known that these criteria all have shortcomings. No matter how you describe them, it is clear that mathematical entities are *not concrete* in the ontological sense. But I would rather oppose concrete to non-concrete than to abstract. I imagine that it is possible for something to be not-concrete without being abstract, that is without being the result of a process of abstraction. Be that as it may, it should be clear by now that the problem I meant to explore is purely epistemological. It is rather a question of how to introduce differences *within* the realm of abstractions.

It is a consequence of my approach that the distinction between levels of abstraction is *relative to a given background*. I have not given a absolute characterization of levels of abstraction. In a sense, it is purely “local”. For instance, the concept of topological space is more abstract than the concept of metric space but it is less abstract than the concept of locale. Thus, although the distinction is not subjective, it is relative. Someone *can* say that something is abstract while somebody else will say that the same piece of mathematics is not that abstract and it won’t constitute a contradiction. In these cases, one has to unearth the underlying context, the various domains of variation that are involved, the relationships between the concepts, if there is any, and the canonical connections between them, if there is any. It should also be obvious that there are concepts that are simply incomparable with respect to their level of abstraction. I believe that this is as it should be.

4.3 Going up the ladder of abstraction

I claim that mathematics in the 21st century *is*, as a matter of fact, becoming more abstract and it is a inevitable result of the application of the abstract method in various contexts.

Sometimes, mathematicians do not even see that they are introducing a domain of variation. This phenomenon happened at least twice in the 20th century and they turn out to be conceptually closely related: category theory and homotopy theory. I will ignore homotopy theory in this paper, although it certainly deserves a whole book. (See, however, [1].) When they were first introduced by Eilenberg and Mac Lane, categories were not even considered as

being genuine mathematical entities. They were merely conceived as a convenient tool. The language of category theory was thought to be merely useful in that it brought a certain order in fields that were at that time rather messy, e.g. algebraic topology and the emerging homological algebra. But the introduction of a language introduces with it the possibility of expressing in a new way properties, expressing new properties and detecting invariant patterns that were hidden before or simply nonexistent. If, moreover, that language has the resources to express in a new fashion criteria of identity, as is the case of category theory, then everything is in place for the process of abstraction to start and this is precisely what happened with category theory. First, categories include in an intrinsic fashion a criterion of identity for the objects of a given category, namely the notion of isomorphism for that category. It naturally comes with the notion and the context. Once you have identified a category, you have a notion of isomorphism, a criterion of identity, for its objects. However, as I have already mentioned, the proper criterion of identity for categories themselves is not the notion of isomorphism of categories. It is the notion of equivalence of categories. It is easy to convince oneself that by adopting the latter criterion we are going up the ladder of abstraction. And it does not stop there. In fact, there is a whole hierarchy of levels, called higher-dimensional categories, that each new level being more abstract than the preceding. Presenting the theory of weak n -categories would require too much space and would lead us into unnecessary technical complexities for our purposes. Suffice it to say that our framework is coherent with the claim that contemporary mathematics is becoming more abstract. The foregoing framework allows us to say what it could mean.

4.4 Abstracting as a way of simplifying

The abstract method constituted a clear cognitive gain. As I have indicated, going in the direction of abstraction is a way of unifying different mathematical domains. The introduction of a level of abstraction is seen as a way of clarifying and distilling what, in some cases, has become a complex domain or, in other cases, exhibits similarities, parallels indicating the possibility of an underlying common framework. The previous disjunction is clearly not exclusive. The new abstract level not only simplifies the situation but it also yields a better control and understanding of the concepts involved. The axiomatic method is an essential part of that process. Axiomatization should be seen, in this light, as a form of design. For instance, in the case of analogies or similarities, axioms capture either a common structure or common properties again leading to a better control and understanding of the features at work. The abstract method is thus used as a seam, a filter in these processes. It brings to the fore the Archimedean points upon which solutions to given problems work. What was previously immersed in a mountain of irrelevant details is unearthed and shown to constitute the mechanisms making concepts work together. This is precisely why we feel justified in speaking of abstraction. As I have said, the process leads to new mathematics, conceptually systematic and organized according to clear principles.

In recent times, the abstract method has even been presented as being inevitable. Here is a wonderful quote illustrating this claim:

The first part of the paper, on which everything else depends, may perhaps look a little frightening because of the abstract language that it uses throughout. This is unfortunate, but there is no way out. It is not the purpose of the abstract language to strive for great generality. The purpose is rather to simplify proofs, and indeed to make some proofs understandable at all. The reader is invited to run the following test: take theorem 2.2.1 (this is about the worst case), translate the complete proof into not using the abstract language, and then try to communicate it to somebody else. [28, 318]

It is not possible to go into Waldhausen’s remarkable and fundamental paper here, which provides an abstract and flexible setting for algebraic K-theory. We have to take his claims as a reliable testimony in the present context. But this brief paragraph contains important and bold claims that would deserve to be unpacked properly. Waldhausen insists that his approach is *not* motivated by the search of the most general solution. He was aware that his framework was more general than Quillen’s framework for algebraic K-theory introduced in the previous decade. Generality is merely a byproduct of the search for the right context. The fundamental purpose is “to make some proofs understandable at all”. The latter is achieved by a process of simplification which, in turn, is itself a byproduct of the abstract method.

Understanding how there can be different levels of abstraction also helps us to see how there can be different mathematical styles. For instance, using categories right from the start and trying to solve given mathematical problems by using categorical concepts and methods and finding a solution up to categorical equivalence is a more abstract way of doing mathematics than by using sets with structures and the resulting notion of isomorphism involved. The reader who knows even just a little bit about the history of algebraic geometry in the second half of the 20th century can recognize Grothendieck’s style in the former and why it was seen as being so outlandish by many. Grothendieck thought that by going up the ladder, one would see better the overall organization of the field, would understand better the general topography of the landscape and that mathematicians would therefore find their way more easily. In his mind, the path to the proof would be trivial and obvious. Abstraction is seen as a radical way to simplify mathematics. Of course, this is possible only after one has understood how to go up the ladder and that it is a ladder. Others finds these heights dizzying or consider that going that far up is not worth the effort.

References

- [1] Michael Barr and Charles Wells. *Category theory for computing science. Repr. Theory Appl. Categ.*, (22):xviii+538, 2012.

- [2] Hourya Benis Sinaceur. Facts and levels of mathematical abstraction. *Philosophia Scientiae*, 18(1):1–32, 2014.
- [3] Olivia Caramello. A topos-theoretic approach to stone-type dualities. 03 2011.
- [4] Leo Corry. *Modern algebra and the rise of mathematical structures*, volume 17 of *Science Networks. Historical Studies*. Birkhäuser Verlag, Basel, 1996.
- [5] Leo Corry. The origins of the definition of abstract rings. *Gaz. Math.*, (83):29–47, 2000.
- [6] Leo Corry. Axiomatics between hilbert and the new math: Diverging views on mathematical research and their consequences on education. *International Journal for the History of Mathematics Education*, 2(2):21–37, 2007.
- [7] Samuel Eilenberg and Saunders Mac Lane. A general theory of natural equivalences. *Trans. Amer. Math. Soc.*, 58:231–294, 1945.
- [8] Maurice Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti Circ. Mat. Palermo*, 22:1–74, 1906.
- [9] Alexander Grothendieck. Sur quelques points d’algèbre homologique. *Tôhoku Math. J. (2)*, 9:119–221, 1957.
- [10] Helmut Hasse. The modern algebraic method. *Math. Intelligencer*, 8(2):18–25, 1986. Reprinting of Jahresber. Deutsch. Math.-Verein. **30** (1930), 22–34, Translated from the German by Abe Shenitzer, With comments by Bruce Chandler and Wilhelm Magnus.
- [11] I. N. Herstein. *Topics in algebra*. Xerox College Publishing, Lexington, Mass., second edition, 1975.
- [12] Nathan Jacobson. *Basic algebra. I*. W. H. Freeman and Co., San Francisco, Calif., 1974.
- [13] Peter T. Johnstone. *Stone spaces*, volume 3 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1986. Reprint of the 1982 edition.
- [14] Israel Kleiner. The genesis of the abstract ring concept. *Amer. Math. Monthly*, 103(5):417–424, 1996.
- [15] Israel Kleiner. *A history of abstract algebra*. Birkhäuser Boston Inc., Boston, MA, 2007.
- [16] Ralf Krömer. *Tool and object*, volume 32 of *Science Networks. Historical Studies*. Birkhäuser Verlag, Basel, 2007. A history and philosophy of category theory.

- [17] Saunders MacLane. Some Recent Advances in Algebra. *Amer. Math. Monthly*, 46(1):3–19, 1939.
- [18] Jean-Pierre Marquis. *From a Geometric Point of View: a study in the history and philosophy of category theory*, volume 14 of *Logic, Epistemology, and the Unity of Science*. Springer, 2009.
- [19] Jean-Pierre Marquis. Canonical maps and category theory: or why categorical foundations are necessary. Forthcoming.
- [20] Jean-Pierre Marquis. Mathematical abstraction, conceptual variation and identity. In P-E Bour; G. Heinzmann; W. Hodges; P. Schroeder-Heister, editor, *Logic, Methodology and Philosophy of Science, proceedings of the fourteen international congress*, Forthcoming.
- [21] Gregory H. Moore. The axiomatization of linear algebra: 1875–1940. *Historia Math.*, 22(3):262–303, 1995.
- [22] Gregory H. Moore. The evolution of the concept of homeomorphism. *Historia Math.*, 34(3):333–343, 2007.
- [23] Jorge Picado and Aleš Pultr. *Frames and locales*. Frontiers in Mathematics. Birkhäuser/Springer Basel AG, Basel, 2012. Topology without points.
- [24] Michael Scanlan. Who were the American postulate theorists? *J. Symbolic Logic*, 56(3):981–1002, 1991.
- [25] Dirk Schlimm. Axioms in mathematical practice. *Philos. Math. (3)*, 21(1):37–92, 2013.
- [26] R. Solomon. *Abstract Algebra*. Pure and applied undergraduate texts. American Mathematical Society, 2003.
- [27] Angus E. Taylor. A study of Maurice Fréchet. I. His early work on point set theory and the theory of functionals. *Arch. Hist. Exact Sci.*, 27(3):233–295, 1982.
- [28] Friedhelm Waldhausen. Algebraic K -theory of spaces. In *Algebraic and geometric topology (New Brunswick, N.J., 1983)*, volume 1126 of *Lecture Notes in Math.*, pages 318–419. Springer, Berlin, 1985.
- [29] John J. Watkins. *Topics in commutative ring theory*. Princeton University Press, Princeton, NJ, 2007.
- [30] Hans Wussing. *The genesis of the abstract group concept*. MIT Press, Cambridge, MA, 1984. A contribution to the history of the origin of abstract group theory, Translated from the German by Abe Shenitzer and Hardy Grant.