Synthese (2013) 190:2141–2164
DOI 10.1007/s11229-011-9962-0

Mathematical forms and forms of mathematics: leaving the shores of extensional mathematics

Jean-Pierre Marquis

Received: 30 June 2010 / Accepted: 20 May 2011 / Published online: 10 June 2011

Abstract In this paper, I introduce the idea that some important parts of contemporary pure mathematics are moving away from what I call the extensional point of view. More specifically, these fields are based on criteria of identity that are not extensional. After presenting a few cases, I concentrate on homotopy theory where the situation is particularly clear. Moreover, homotopy types are arguably fundamental entities of geometry, thus of a large portion of mathematics, and potentially to all mathematics, at least according to some speculative research programs.

Keywords Philosophy of mathematics · Algebraic geometry · Category theory · Homotopy theory

1 The background

Mathematics in the twentieth century is often presented as being developed in a framework in which set-based structures are classified up to isomorphism. This picture follows from the standard set-based foundational framework. As is well-known, in that context an isomorphism is a *bijective* function that preserves the relevant structure. This is what I will call the *purely extensional* point of view. Let me briefly develop this perspective in more details, even though most of what I am about to say is considered as being nothing less than obvious and trivial by now.

The extensional point of view can also be thought as arising from the idea that mathematical objects are collections of elements on which a structure is added depending on the theory or the needs. It is rather obvious that, in this context, two objects are con-

J.-P. Marquis (✉)
Département de philosophie, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montreal, QC H3C 3J7, Canada
e-mail: Jean-Pierre.Marquis@umontreal.ca

sidered identical or as the same if there is a one-to-one and onto map from one to the other such that the structure is preserved by the map. Thus, a fundamental ingredient inherent to the criterion of identity is the cardinality of the underlying set of the object. I want to emphasize immediately one important fact: the idea makes perfect sense and is even indispensable. Indeed, for classifying order relations, algebraic structures, various geometric structures and in view of some applications, one *has* to differentiate structures *and* use cardinality as a differentiating property. From a more formal point of view, one *takes* as the interpretation of various predicates, relations, etc. in a structure, more precisely a *set* viewed as a collection of elements, thus as an extension. A formal language is therefore fully interpreted in this framework and the sense and the reference of predicates, relations, etc. have to be understood as extensions[1]. To *know* a mathematical object, one has to know its elements, its "points" and how they are related to one another by some given structural principle. Notice that it is really the latter that matters in the end, since any two isomorphic structures are considered to be the same, the exact and precise nature of the points themselves being irrelevant. This way of proceeding may yield some limitations, distortions, inadequacies and even counter-intuitive results. But as far as mathematics is concerned, the extensional point of view seems to be perfectly fine as it is.

Thus the extensional point of view has become what I will call a *form of mathematics* and it is clearly the form that underlies mainstream philosophy of mathematics and has occupied the forefront of the field at least over the last 50 years or so. The main alternative forms, namely intuitionism and its numerous constructive variants, are often seen as being based on deviant logical and mathematical ideas whose philosophical basis is associated either with what many consider to be dubious idealism or with a description of mathematical objects at odds with what most mathematicians would agree with.

My main claim in this paper is that there is, within "mainstream" mathematics a form of mathematical knowledge that has developed slowly but steadily from approximately 1950 onwards[2]. That form has ramified in various fields that are presently important research fields with applications in numerous and important domains but which are based on a non-extensional point of view. Furthermore, this alternative form of mathematical knowledge emerged, not from an a priori philosophical conception of mathematical knowledge, but from mathematical needs. We tend to forget that the extensional point of view emerged from a specific encoding of the notion of set and that, although its success was and still is spectacular, a different mathematical encoding arising from different mathematical constraints might be possible and as fruitful and powerful. As we will try to show, this form of mathematics arose in a circle of tightly related fields of mathematics: algebraic geometry, algebraic topology, in particular homotopy theory, and category theory. We will briefly survey how this form of

[1] There are, of course, ways of forcing an intensional point of view by using "variable sets" of what is usually called "possible worlds semantics", in other words variable extensions, hoping that in that way, one would model intensional aspects of a situation. This is not what we have in mind here, although there are connections.

[2] Needless to say, it has roots that go back in the nineteenth century, but we want to focus on mathematics that that has crystalized since the advent of first-order logic and set theory.

mathematics arises in algebraic geometry and category theory and focus on homotopy theory since it appears to contain a basic and fundamental notion of mathematical forms, namely *homotopy types*. Furthermore and unexpectedly perhaps, homotopy theory turns out to *have* connections with constructivism, although it is too early to see whether these connections are inescapable from a foundational viewpoint. We hasten to add that the same phenomenon most likely occurs in other areas of mathematics and only our lack of knowledge of these fields refrain us from mentioning them here. However, given that methods of algebraic geometry, categorical methods and homotopical methods are spreading slowly but surely in other mathematical areas and given that homotopy types, as we will argue, can be considered just as fundamental as the natural numbers, this form of mathematics is likely to become inescapable and can, perhaps, subsume in one way or another the extensional point of view.

We will proceed as follows. We will briefly show how algebraic geometry breaks from the extensional point of view first from the birational point of view and then in the contemporary setting, namely in the theory of schemes. Second, we will go over the situation in category theory, where the notion of equivalence replaces the notion of isomorphism. In fact, the latter turns out to be closely related to the notion of homotopy equivalence, our central concern in this paper. We will define homotopy types, show how they do not fall under the standard extensional point of view and sketch the basic ingredients of classical homotopy theory. We will explore the *epistemological* dimensions of this new form of mathematics in a companion paper.

2 Identity in algebraic geometry

Algebraic geometry has a long and tortuous history. In the twentieth century alone, it went through various phases, from the circumvoluted work of the Italian school, the first transcription into the modern algebraic dressing, Weil's foundational approach and, finally, Grothendieck's setting in terms of schemes, which is now considered standard (for an instructive historical survey of algebraic geometry, see Dieudonné 1985). I will entirely ignore the fascinating historical aspects of the story and concentrate on the main ingredients relevant to my main thesis.

From a naive point of view, algebraic geometry seems to sit squarely at the center of the extensional framework. Indeed, algebraic geometry is the study of algebraic varieties and the latter are informally defined as being the solutions of a system of polynomial equations over a field k. Thus, one can think of an algebraic variety V as a set of points which are the common zeros of the system of polynomials. Algebraic curves are special cases of algebraic varieties. For instance, the curve in $\mathbb{C}^2$ defined by $\{(x, y) : x^2y + xy^2 - x^4 - y^4 = 0\}$ is an algebraic variety. A different example of an algebraic variety is given by the quadratic cone in $\mathbb{C}^3$ defined by $\{(x, y, z) : x^2+y^2-z^2 = 0\}$. The graphs of transcendental functions are not algebraic varieties. Thus, very roughly speaking, algebraic geometry studies geometric objects defined by systems of polynomials, the latter being syntactical expressions involving basic arithmetical operations. This in itself immediately raises a series of interesting conceptual issues—for instance completely different equations might give rise to the same variety—and these conceptual issues receive satisfactory solutions by precise mathematical results, e.g. Hilbert's Nullstellensatz.

Classical algebraic geometry is usually done over the field of complex numbers $\mathbb{C}$, since in this context algebraic geometry becomes easier, by the so-called fundamental theorem of algebra. One of the key elements characterizing the passage from the classical setting to the contemporary setting consists in moving from $\mathbb{C}$ to an arbitrary algebraically closed field k and, even to any field k. Thus, in the contemporary setting, one can start with an arbitrary field k, and consider the polynomial ring $k[x_1, \ldots, x_n]$ over that field. But doing so requires rewriting the whole theory on new grounds.

Again, in the classical setting, three kinds of varieties are defined: *affine* varieties, *projective* varieties and *quasi-projective* varieties. A projective variety is a variety defined by *homogeneous* polynomials in the (complex) projective space $\mathbb{P}^n$. Recall that a polynomial $f \in \mathbb{C}[x_0, \ldots, x_n]$ is said to be *homogeneous* if all its terms have the same degree. The most general class of varieties is the class of quasi-projective varieties. Given a topological space X, a *locally closed set* of X is an intersection of an open set and a closed set of X. A quasi-projective variety is then a locally closed subset of $\mathbb{P}^n$, where the topology of $\mathbb{P}^n$ is the Zariski topology. We will get back to the Zariski topology in a short while. It can be seen that the class of quasi-projective varieties includes all projective varieties, all affine varieties, and all Zariski open subsets of these. When we want to talk about any one of them, we talk about an *algebraic* variety.

In each case, one can define an appropriate notion of a morphism of varieties in terms of polynomials between them. Two varieties V and W are then treated as being identical if they are isomorphic, that is if there is an isomorphism between them, where the latter notion simply means that the given morphism is bijective and its inverse is also a morphism of varieties. So far, we are squarely in the extensional point of view. Things start to change subtly when we move to what is called the birational point of view in algebraic geometry.

Birational geometry was more or less launched by Riemann, when he understood that topological invariants—for instance Riemann's genus of a curve—could be attached to a class of birationally equivalent irreducible algebraic curves, developed by the Italian school who obtained spectacular results in the problem of classification of algebraic surfaces and then became more or less lay dormant until the 1960s when Hironaka proved his fundamental desingularization theorem. A new impetus was given to the field in the 1980s when Mori proved his famous result and launched what is now called the minimal model program, which is still an active area of research (see Grassi 2009 for a survey of birational geometry).

As usual, when invariants can be attached to mathematical objects, one hopes to be able to use them in order to obtain classification results about these objects. This is more or less the motivation underlying the birational point of view: to obtain a classification of algebraic varieties. Thus, one defines an equivalence relation on algebraic varieties and in this case two algebraic varieties will be equivalent if, and this is the interesting point, they are the same *almost* everywhere. This idea is formalized by the concept of a rational map between algebraic varieties, as follows.

Let X be an algebraic variety and U and U' be dense open subsets of X[3]. Suppose, moreover, that the maps $U \xrightarrow{\phi} Y$ and $U' \xrightarrow{\phi'} Y$ are morphisms of algebraic varieties (really, of quasi-projective varieties). Then (U, ϕ) and (U', ϕ') are *equivalent* if the mappings ϕ and ϕ' coincide on $U \cap U'$. This defines an equivalence relation. Given this relation, a *rational map* $X \dashrightarrow Y$ is an equivalence class of morphisms defined on dense open subsets of X (see, for instance, Hartshorne 1977, pp. 24–27 for details).

It is obvious from the definition that a rational map is *not* a set-theoretic function. It is an equivalence class of maps defined only on dense open subsets of X. One can think of a rational map as a morphism defined on an *arbitrary* dense open subset, since it does not matter which dense open subset is chosen. It should be observed that on the domain of definition, a rational map *is* a morphism of varieties. It is in this sense that a rational map is defined almost everywhere. We finally get to our main definition:

Let X and Y be (irreducible) algebraic varieties. Then X and Y are said to be *birationally equivalent* if there are rational maps $f : X \dashrightarrow Y$ and $g : Y \dashrightarrow X$ such that the composites $f \circ g$ and $g \circ f$ agree with the identity maps on dense open sets where the composite makes sense as morphisms of algebraic varieties.

Informally, this simply says that two algebraic varieties are birationally equivalent if they are isomorphic on a dense open subset. But one should not confuse the notion of being birationally equivalent with the notion of being isomorphic as varieties. For instance, one can show that the projective plane $\mathbb{P}^2$ is birationally equivalent to the product of two projective lines $\mathbb{P}^1 \times \mathbb{P}^1$, but they are not isomorphic. As we have already mentioned, the main point underlying the notion is the fact that birational equivalence preserves many important invariants of a variety (see, for instance, Grassi 2009, pp. 103–107).

Our claim here is that we are, even in the classical framework, moving away, albeit slowly and imperceptibly, from the extensional point of view[4]. Birationally equivalent algebraic varieties are not identified via bijective structure preserving maps. One could retort that the notion of birational equivalence is unique to algebraic geometry and, as such, does not open the door to a new form of mathematics. Furthermore, the classification of algebraic varieties up to birational equivalence is equivalent to the classification of function fields up to *isomorphism*. Thus we are back to the extensional point of view after all.

However, we believe that this objection does not faithfully reflect the situation. First, it is a fundamental notion in algebraic geometry: the classification of algebraic curves, and more generally, of algebraic varieties of higher-dimension rests upon that notion. Second, the notion is entirely subsumed in the more recent and powerful framework developed by Grothendieck and his school in the 1960s[5]. Third, the

[3] Again, we haven't explained as yet where the topology comes from. This is, of course, an important point and we will give the formal definition later. For the time being, simply assume that a variety has a topology.

[4] I want to thank an anonymous referee who suggested that I introduce the case of algebraic geometry in order to show that my main point was in fact more general and widespread than what my paper suggested at first.

[5] In fact, Hartshorne, certainly one of the standard references in the field, devotes two whole chapters of his book to show how to rewrite the classical results on algebraic curves and algebraic surfaces in the contemporary language. See Hartshorne (1977, Chaps. IV, V).

contemporary framework has applications in much wider spectrum of mathematical disciplines, from number theory to analysis. It does so by generalizing the methods in various directions, e.g. for non-algebraically closed fields, and by introducing powerful new invariants via a host of cohomology theories. Fourth, the new language, the language of schemes, introduces a new way of thinking about *points* in the context of algebraic geometry. The latter are derived from the algebraic data and not given by a pre-existing set-theoretical framework. We will now briefly sketch this additional facet of our story.

One of the striking achievements of contemporary algebraic geometry is the systematic exploitation of abstract algebraic concepts and methods in order to solve geometric problems. Contemporary algebraic geometry rests heavily on commutative algebra, homological algebra and category theory. One usually starts with a (algebraically closed) field k and then consider the polynomial ring $A = k[x_1, \dots, x_n]$ in n variables over k. But let us simply start with an arbitrary commutative ring R with unit and define the *spectrum of* R, denoted by $\mathrm{Spec}(R)$.

First, recall that an *ideal* of a ring R is a subset I such that (i) $0 \in I$; (ii) if $a, b \in I$, then $a + b \in I$; and (iii) if $a \in I$ and $r \in R$, then $ar \in I$ and $ra \in I$. In words, an ideal I of R is an additive subgroup such that $RI \subset I$. An ideal P is said to be *prime* if (i) $1 \notin P$ and (ii) if $ab \in P$, then $a \in P$ or $b \in P$.

The underlying set of the spectrum of a ring R is the set of prime ideals of R. A *point* P of $\mathrm{Spec}(R)$ is simply a prime ideal P of R. But $\mathrm{Spec}(R)$ also has a topological structure, given by the so-called Zariski topology. The *closed* sets of the topology are defined thus: for each subset $S \subset R$, we define

$$V(S) = \{P \in \mathrm{Spec}(R) : S \subset P\}$$

where P is of course a prime ideal of R. It can be verified that this does indeed define a topology.

And this is not all. $\mathrm{Spec}(R)$ also has a sheaf structure[6]: for each open set U, one can associate $\mathcal{O}(U)$ a commutative ring with unit in a systematic fashion and in such a way that these rings are compatible with the topology in a strict fashion.

Thus, the spectrum of a ring R forms what is called a *ringed space*: that is, it is a topological space together with a sheaf of rings on the space. We thus get to the definition of a scheme: an (affine) *scheme* is a locally ringed space[7] isomorphic to the spectrum of some ring.

Although it is not direct, one can recover all the classical notions of algebraic geometry in this new setting. But we also obtain some surprising phenomenon creeping in that are directly relevant to our claim. For instance, in the standard set-theoretical setting, a one point topological space has a unique continuous function to itself, namely the identity mapping. It is easy to construct a spectrum $\mathrm{Spec}(R)$ with a unique point such that it has many endomorphisms different from the identity map: simply take R to

[6] We will not define the notion of a sheaf here. The interested reader is urged to consult, for instance, Mac Lane and Moerdijk (1994).

[7] We haven't given the technical definition of a *local* ringed space. The reader should consult (Hartshorne, 1977, Chap. II, Sect. 2) for a precise definition.

be a field k! It is easy to see that in that case, there is a unique prime ideal, thus a unique point and it is as easy to see that there are endomorphisms different from the identity map whenever k is not a prime field[8]. This is probably the simplest example we can give of a break from the extensional point of view. And there are other indications of the breakdown: mappings between schemes are *not* determined by their points. Again, as such, these fact do not, as such, show any short comings of the extensional point of view. But they do open the door to ways of thinking in which the notion of point is not pre-existing and where identities might rely on concepts different from bijective structure preserving maps. And this is our main point.

One of the key conceptual ingredient to Grothendieck's approach to algebraic geometry was his decision to take *all* commutative rings as the proper algebraic setting to work with. Given his definition of schemes, one can then show that the category of affine schemes is *categorically equivalent* to the (opposite) category of commutative rings with unit. Surprisingly, the algebraic side contains implicitly a whole lot of the geometric content involved in various situations. This allows one to move freely between the algebraic and the geometric sides via the categorical equivalence given by functors. One can in fact claim that, from a categorical point of view, these categories are, in some sense, identical. But we are now talking about the identity of categories, to which we now turn.

3 Identity in category theory

When Eilenberg and Mac Lane introduced categories in 1945, they were immersed in the extensional point of view and thus it seemed natural to stipulate that two categories **C** and **D** are identical if and only if they are isomorphic (see Eilenberg and Mac Lane 1945). Formally, this simply means that there are functors

$$F : \mathbf{C} \longrightarrow \mathbf{D} \quad \text{and} \quad G : \mathbf{D} \longrightarrow \mathbf{C}$$

such that $G \circ F = Id_{\mathbf{C}}$ and $F \circ G = Id_{\mathbf{D}}$. This forces the functors F and G to be bijective functions preserving the categorical structure, that is composition of morphisms.

Since categories, as mathematical objects, were not used as such by mathematicians before the late 1950s, no one noticed that a different notion of identity between categories was in fact required. It is only when Grothendieck successfully extended results of homological algebra to sheaves that he realized that a more relaxed notion was needed to account for a specific identity between two different presentations of sheaves[9]. I do want to underline the fact that the right notion of identity for categories arose from the mathematical practice and that, a priori, there was no reason to believe that the notion of isomorphism of categories would not be appropriate. The appropriate

[8] I want to thank an anonymous referee who has rightly pointed out to me that prime fields are precisely the fields whose spectra have no non-identity endomorphisms.

[9] As a matter of fact, Grothendieck did not give the correct definition in his original paper, probably the result of an innocent mistake. For more about the historical context and Grothendieck's motivation, see Krömer (2007).

notion is that of *equivalence* of categories, defined thus: two categories **C** and **D** are said to be *equivalent* whenever there are functors

$$F : \mathbf{C} \longrightarrow \mathbf{D} \quad \text{and} \quad G : \mathbf{D} \longrightarrow \mathbf{C}$$

together with natural isomorphisms

$$\alpha : Id_{\mathbf{C}} \longrightarrow G \circ F \quad \text{and} \quad \beta : Id_{\mathbf{D}} \longrightarrow F \circ G.$$

Thus, it seems that we have merely replaced equalities by isomorphisms. Thus, when we move from **C** to **D** and back to **C** via the functors F and G, we don't end up at our starting point, but with an object that is isomorphic to the original and the same happens when we start from **D** and end in **D**. The difference might seem innocuous, but from a conceptual point of view and from an extensional point of view, it is a remarkable difference. Indeed, in this case, the functors are not in general bijective. Thus, the cardinalities of **C** and **D** can be different although, as categories, **C** and **D** will be considered to be indistinguishable. To illustrate how different categories can be from the purely extensional point of view, take the category **C** to be the one object category with the identity morphism and take **D** to be a category with uncountably many singletons with the identity morphism on each object and, for each pair of objects, the unique morphism between them. These two categories are equivalent, although one is finite and the other uncountable[10].

In the case of categories, one might rebut that one can fall back on the notion of the skeleton of a category and recall that two categories are equivalent if and only if their skeletons are isomorphic. This would show, presumably, that the working underlying notion is that of isomorphism and that we still are in an extensional framework after all.

To better understand the objection and how one can respond to it, let us first fix the terminology. Informally, a category **C** is *skeletal* if isomorphic objects of **C** are in fact equal. In a more philosophical jargon, there is no redundant copies of isomorphic types in **C**. Given a category **C**, a *skeleton* of **C** is a skeletal category **S** such that **S** is a full subcategory of **C** and there is a functor $I : \mathbf{S} \longrightarrow \mathbf{C}$, the inclusion functor, such that I is an equivalence of categories[11] (see Mac Lane 1998, p. 93 for more on the notion). Mac Lane gives the example of the category of all finite sets: one of its skeleton is the (full) subcategory with objects all finite ordinal numbers, where 0 is the empty set and each $n = \{0, \ldots, n-1\}$ as usual.

Although this is fine and technically sound, it fails to constitute an objection. First, the foregoing results according to which every category has a skeleton and two

[10] In fact, any category equivalent to the category consisting of a single object and its identity arrow is called *categorically contractible* in the literature and the latter concept occupies an important conceptual position in certain contexts. See, for instance, Dwyer et al. (2004).

[11] Recall that a category **S** is a *subcategory* of a category **C** if the objects of **S** are objects of **C**, the arrows of **S** are arrows of **C**, for each arrow f of **S**, the domain and the codomain of f are in **S**, each object of **S** has its identity arrow and each pair of composable arrows in **S** has its composite in **S**. A subcategory **S** is *full* whenever the inclusion functor I is full, that is when to every pair of objects X, Y of **C** and to every arrow $g : I(X) \longrightarrow I(Y)$, there is an arrow $f : X \longrightarrow Y$ of **S** such that $g = I(f)$.

equivalent categories have isomorphic skeletons rest on the axiom of choice, thus on an underlying set theoretical foundation. In some settings, e.g. internal categories in some fixed topos, the claim that every category has a skeleton is simply false. Thus, it is only within the extensional point of view that these claims can be proven. But one can argue that: 1. the proper set theoretical setting for category theory has still to be provided and 2. one can develop category theory in an autonomous fashion, that is, independently of current set theoretical frameworks. Furthermore, as we have already emphasized, as far as the practice of category theory is concerned, it is really the notion of equivalence that is at work and *not* the notion of isomorphism of categories. Third, the result does not capture the full structure of categories and, in fact, the proper setting in which the identity of categories arises. When one looks at categories and functors between them, one inescapably has to consider natural transformations between functors and the structure arising from them, and the latter structure is at least a 2-category, even a bicategory.

Thus, if categories come to occupy a central role in the foundations of mathematics, one would expect that the extensional point of view might recede slowly to the background. I want to emphasize immediately that this would *not* mean that the notion of set would be evacuated from mathematics. A certain conception of sets might still be fundamental to the whole enterprise, although it would not rest on the extensional principle. Whether this is a feasible, conceptually and mathematically sound project remains to be seen (but see Makkai 1998 for a sketch of what such a theory would look like).

4 Homotopy types

I will not dwell here on the fascinating and important history of homotopy theory (see Dieudonné 1989; Marquis 2006). I will recall the basic definitions relevant to our purpose and briefly explain why homotopy theory is considered to be at the foundations of algebraic topology (for more thorough presentations, see Hatcher 2002; Aguilar et al. 2002 or Rotman 1988).

Let us start with the notion of a homotopy between continuous functions. Let X and Y be topological spaces and $f, g : X \to Y$ be continuous maps. A *homotopy* H *from* f *to* g is a continuous map $H : X \times [0, 1] \to Y$ such that

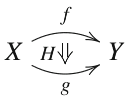
$$H(x, 0) = f(x), \quad \forall x \in X$$

and

$$H(x, 1) = g(x), \quad \forall x \in X.$$

This is the formal definition and it is probably not very enlightening. The underlying idea is extremely simple: a homotopy is a *continuous deformation* of f into g. This is a pervasive and general concept: for instance, whenever f and g are continuous functions and represent processes, then a homotopy is a deformation of f into g by "infinitesimal changes". I will get back to the conceptual significance of the idea once we will have a better grasp of its properties.

Whenever there is a homotopy from f to g, we say that the maps f and g are *homotopic* and it is denoted by $f \sim g$. Here is a very abstract but useful graphical representation of a homotopy:

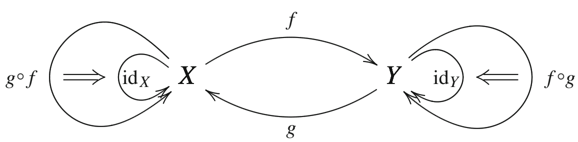


The homotopy H from f to g is here represented by a vertical double arrow.

The concept of a homotopy is so simple that one might wonder how it could be mathematically so useful. Informally, it captures the idea that a certain mathematical property is stable under continuous deformations and that idea turns out to be mathematically important and powerful since it is a property of many physical and formal systems.

This idea is captured by the following elementary fact: it can be shown that $f \sim f$, every map is homotopic to itself, that if $f \sim g$, then $g \sim f$, i.e. the relation is symmetric and that being homotopic is a transitive relation, that is if $f \sim g$ and $g \sim h$, then $f \sim h$. Thus, it is an equivalence relation between continuous maps and it is therefore possible, very often judicious, to consider *homotopy classes* of maps, namely $[f] = \{g : f \sim g\}$. This, in itself, might not seem to be very exciting, but in some specific mathematical contexts, the solution to a problem is provided not so much by a *specific* map, but rather by the *homotopy class* of such maps. In such cases, mathematicians say that they are working "up to homotopy", that is up to a continuous deformation.

We now move to the notion we are interested in: homotopy equivalent spaces. Two spaces X and Y are said to be *homotopy equivalent* if there are continuous maps $f : X \to Y$ and $g : Y \to X$ such that $f \circ g \sim \mathrm{id}_Y$ and $g \circ f \sim \mathrm{id}_X$, that is the composites $f \circ g$ and $g \circ f$ are *homotopic* to the identities id_Y and id_X respectively. Informally, this means that there are continuous transformations of the images of the spaces into the spaces themselves. Again, here is an abstract but useful representation of the situation.



Being homotopy equivalent is an equivalence relation and, thus, a particular space is a token of a homotopy *type*. We should point out immediately that there are in general many different homotopies between two given maps and, thus, in particular, there are many different homotopies between two given spaces. Since being identical in this context amounts to being homotopically equivalent, the identity between two spaces is not, in this context, an all or nothing affair. Two spaces can be homotopy equivalent in many different ways and each and everyone of these homotopy equivalence yields information about the spaces. This should be viewed as a strength of the theory and not a drawback, for whenever there is an identity between two spaces, this identity itself, that is the system of homotopies, contains a lot of information about the different ways the spaces can be deformed one onto the other. A first reaction might be to reject this form of identity as being incomprehensible, but on the contrary, I submit that the notion of identity at work here is philosophically fundamental: we are dealing with entities that can be continuously transformed into one another. Informally, this could certainly constitute a model of identity over time or of one and the same entity that changes or takes various forms but stays nonetheless the same. That which stays the same is modeled here by the homotopy type. Although we will concentrate here on the universe of mathematical objects, I believe that reflections on this kind of identity goes well beyond the realm of mathematics and touches fundamental problems of metaphysics.

Notice, and this is a crucial point, that nowhere have I said that the continuous maps f and g need to be bijections. This is the key difference between being *homeomorphic* and being homotopy equivalent. Recall that two spaces X and Y are *homeomorphic* if there are continuous maps $f : X \to Y$ and $g : Y \to X$ such that the composites are *equal* to the identities id_Y and id_X respectively, i.e. $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. Whenever this is the case, the maps are *necessarily* bijections. It trivially follows that two homeomorphic spaces are necessarily homotopy equivalent. But homotopy equivalent spaces need not be homeomorphic in general. A few simple examples should suffice to illustrate the situation and allow us to make our point.

Here are standard and simple examples that appear in any textbook that illustrate what the notion of homotopy equivalence amounts to. Let 1 denote the one-point space. Then it is easy to show that 1 and the real line $\mathbb{R}$ are homotopy equivalent! (the proof follows from the fact that all continuous maps $\mathbb{R} \to \mathbb{R}$ are homotopic). A space homotopy equivalent to 1 is said to be *contractible*. For instance, any open interval (a, b) is contractible and so is any closed interval $[a, b]$. A space with a single point and the real line $\mathbb{R}$ are tokens of the same homotopy *type*. And so is the real plane $\mathbb{R}^2$ and even the n-dimensional space $\mathbb{R}^n$. Since an open interval and a closed interval cannot be homeomorphic—one is compact while the other is not—this provides us with a simple illustration that two spaces can be homotopic without being homeomorphic.

A somewhat different example is provided by the so-called annulus, denoted by A, and defined by

$$A = \{(x, y) \in \mathbb{R}^2 : 1 \leq \sqrt{x^2 + y^2} \leq 2\}.$$

It can be shown that the annulus is homotopy equivalent to the circle S^1 and to the cylinder $S^1 \times [0, 1]$. In the same spirit, the punctured plane $\mathbb{R}^2 - \{0\}$ is also homotopy equivalent to the circle S^1. The annulus A, the circle S^1, the cylinder $S^1 \times [0, 1]$ and the punctured plane $\mathbb{R}^2 - \{0\}$ are tokens of the same homotopy type. And so are the open disk $S^1 \times \mathbb{R}$, the solid torus $S^1 \times \mathbb{D}^2$ and the Möbius band. Here again, e.g. the circle S^1 and the Möbius band, we have spaces that are homotopic but not homeomorphic.

Here are a few slightly different examples involving highly visual and fundamental objects: finite polyhedra. A finite polyhedron can be described informally as a topological space homeomorphic to a "regular" polygon, for instance what is called a *simplex* in the literature. A simplex is a polygon constructed from simple geometric blocks, like points, line segments, triangles, etc. For example, a tetrahedron is a finite simplex and thus a finite polyhedron. From a topological point of view, one wants to forget the specific geometric figures used to construct a polyhedron and thus a polyhedron is defined to be a topological space homeomorphic to a simplex. Homeomorphic types, which can be defined in the obvious manner, of one-dimensional polyhedra are *graphs*, whereas the homotopy types of these constructions correspond to *numbers*[12]. The classification of homotopy types of finite polyhedron is far from being a trivial task.

Homotopy types are, informally, *abstract spaces* or *abstract shapes*. This is certainly not precise enough, since the latter expression is already too vague to convey any interesting meaning. I am tempted to propose to call a homotopy type a *khôra*, to borrow Plato's expression. By this, I mean to suggest that a homotopy type is a basic space in which mathematical forms can be "embodied"[13]. But I will not use that terminology in this paper. The main point here is that there need not be a bijection between tokens of a given type. In other words, a deformation of one space onto another does not have to preserve the cardinality of the underlying sets. Of course, there are bijections between the annulus A, the circle S^1 and the punctured plane $\mathbb{R} - \{0\}$, but they are not *continuous* maps. Clearly, there is no bijection between the one point space 1 and the real line $\mathbb{R}$! They are nonetheless tokens of the same type. This is a clear indication that we have left the extensional point of view behind[14].

In a sense, it is not even clear that one needs or should refer to *elements* of the underlying sets in the case of homotopy types. In fact, it is hard to see how one could define a homotopy type as a set of elements with a certain structure. We understand how the equivalence relation over the collection—really the *category*—of topological spaces generates the corresponding equivalence classes, but I submit that this is a poor representation of homotopy types themselves. I should underline immediately that we do not want to represent these types as *sets* of homotopy equivalent spaces. For reasons that will become clear in subsequent sections, mathematicians would rather have a direct abstract description of these types in a language that is appropriate for them.

[12] More specifically, graphs are classified by the Euler characteristic and the latter is a complete invariant in this case.

[13] I do *not* want to say that homotopy types are mathematical *structures*. There are clear differences between the way mathematical structures are usually conceived and homotopy types. I will come back to the structure of homotopy types in a latter section. How this affects, if at all, various strands of mathematical structuralism will have to be discussed elsewhere.

[14] The situation is considerably different from what we find, for instance, when we examine the natural numbers. It is of course easy to describe natural numbers as types based on the standard equivalence relation of equinumerosity. Two crucial differences have to be underlined. First, in the case of numbers, one can provide a uniform description of numbers in terms of sets, e.g. Von Neumann ordinals, whereas it is hard to see how this could be done for homotopy types. Second, and this is probably even more important for our purpose, in the case of the natural numbers, the equivalence relation between tokens of numbers, i.e. sets, is *essentially* a bijection, thus sitting right at the core of the extensional point of view.

Why should philosophers of mathematics pay any attention to homotopy types? Are there good mathematical reasons to believe that they have a particular status? One of our goals for the remaining parts of this paper is to argue that not only are they indispensable in mathematics, but moreover they are *fundamental* to mathematics and the standard extensional point of view cannot reflect this adequately.

Let us first start with their fundamental nature. To wit:

> All this underlines the fundamental importance of homotopy types of polyhedra. There is no good intuition what they actually are [sic], but they appear to be entities as genuine and basic as numbers or knots. (Baues 1995, p. 5)

The fact that we do not have a good intuition of what homotopy types (even of polyhedra in this case) actually are does not mean that we do not possess a considerable amount of information about them or that we cannot evaluate their importance. Baues not only claims that they are "genuine entities" but also that they appear—notice the hesitation—to be as genuine as numbers or knots. One possible analogy—too crude, I am afraid—is that homotopy types are to geometrical forms what prime numbers are to numbers in general[15]. However, what matters here is not the analogy, but rather the reasons why we believe that there is a reasonable analogy at work. And the reasons are both empirical and theoretical.

On the theoretical side, it has become clear that all of algebraic topology is done "up to homotopy", that is all the various homology, cohomology theories and other similar constructions are done up to homotopy equivalence. This shows that homotopy types underlie all the tools developed to detect topological invariants. We will come back to this point later.

On the empirical side, work done over the last 40 years or so show that "homotopy types of polyhedra are archetypes underlying most geometric structures" (Baues 1995, p. 5). This last claim rests on specific mathematical theorems and specific relationships, namely forgetful functors, between particular categories of geometric entities. Figure 1 exhibits some of the categories involved and the forgetful functors between them (Baues 1995, p. 5).

The category of polyhedra (with *homotopy equivalences* between polyhedra) sits at the very bottom. Thus, in a very specific sense, all the geometric structures lie above homotopy types and some of their most important properties depend directly on the latter. Here are some of the mathematical results mentioned by Baues that illustrate this dependence.

> Some of the arrows in the table correspond to results in the literature. For example, every differentiable manifold is a polyhedron, see J.H.C. Whitehead or Munkres [Baues includes references]. Any (metrizable) topological manifold is proper homotopy equivalent to a locally finite polyhedron though a topological manifold needs not to be [sic] a polyhedron, see Kirby and Siebenmann. Any semi-analytic set is a polyhedron, see Lojasiewicz. There are also connections

[15] It might be possible to develop an argument to show that homotopy types are required in the same way that natural numbers are required even for the most basic geometric representations. Natural numbers and (finite) homotopy types would go hand in hand and a system of *written* symbolic representations would require both numerical components and geometric components.

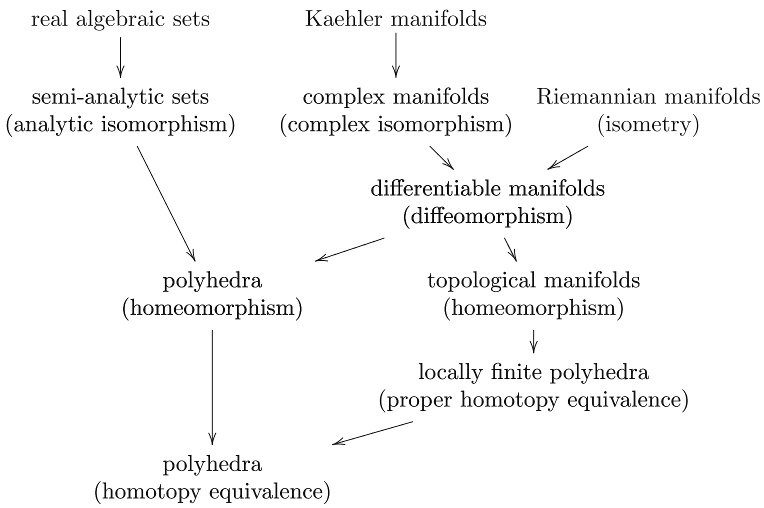


Fig. 1 Forgetful functors between categories of geometric systems

> between the objects in the table in terms of realizability. For example, each differentiable manifold admits the structure of a Riemannian manifold, or each closed differential manifold has the structure of an irreducible real algebraic set (in fact, infinitely many birationally non isomorphic structures), see Bochnak and Kucharz.
> The famous Poincaré conjecture states that the homotopy type of a 3-sphere contains only one homeomorphism type of a topological manifold. Clearly not every finite polyhedron is homotopy equivalent to a closed topological manifold. (…) By the result of M.H. Freedman all simply connected 4-dimensional Poincaré complexes have the homotopy type of closed topological manifolds, they do not in general have the structure of a differentiable manifold by the work of Donaldson. (Baues 1995, p. 6)

And the list could be expanded.

Two classification problems sit at the core of algebraic topology: the classification of topological spaces up to homeomorphism and the classification of homotopy types. It should be clear that it is the latter that is more fundamental, since the former depends on the latter. From what we have said so far, it is hard to see how the problem of classification of homotopy types could be handled: what kind of properties characterizes homotopy types? It turns out that the (homotopy) dimension and the connectivity are two basic properties of homotopy types. Furthermore, homotopy types even possess a structure[16].

[16] Thus, we *are* introducing structures after all! But our original point remains: homotopy types are not defined as being sets equipped with a structure. The story is considerably more complicated than this characterization of mathematical structuralism suggests. For some of the problems this form of structuralism faces, see for instance Carter (2005, 2008).

Recently, Baues introduced an analogy between finite homotopy types and the atomic nature of molecules (see Baues 2002 for details). The analogy runs as follows: in the same way that molecules can be decomposed into atoms by applying certain forces and that the latter atoms are indecomposable with respect to these forces, finite spaces can be decomposed into atoms by topological means and the latter cannot be split any further by these means. Furthermore, in the same way that physical atoms are ordered by the number of protons inside their nucleus, topological atoms (of low-dimensions) are ordered by the number of cells inside them. In this analogy, the hydrogen atom corresponds to the 1-sphere S^1. But Baues gives a general definition of an atom of topology based on two components. First, since an atom, by definition, cannot be decomposed, an appropriate notion of decomposability for topological spaces has to be given. Second, the definition rests on a technical condition linking the dimension of a homotopy type and its connectivity. Thus, we first need to understand these notions.

Informally, a space should be decomposable if it is composed of parts, that is subspaces that are joined together by some topological means, more specifically by some homotopical means. Let A and B be two spaces and $a_0 \in A$ and $b_0 \in B$ be two points. The *one point union* $A \vee B$ of A and B is obtained by gluing the two points together[17]. A space X is said to *decomposable* if it is homotopy equivalent to a one point union $A \vee B$ where A and B are non-contractible; otherwise the space X is said to be *indecomposable*. A *finite* space X is a space homotopy equivalent to a finite polyhedron[18]. A topological atom should be an indecomposable finite space. However, this is not enough. For instance, the one point space 1—and thus any contractible space—cannot be an homotopical atom, since homotopy types are simply not "built up" from it. This is a further indication that we simply cannot think of homotopy types as made up of points, or in other words, as sets of points.

This is linked to the connectivity of a space. A space X is $(n-1)$-*connected* if each continuous map $S^i \to X$ with $i \leq n-1$ admits a continuous extension $B^{i+1} \to X$, where B^{i+1} is the $(i+1)$-ball. The *connectivity* of a space X, written $\mathrm{conn}(X)$, is $n-1$ if X is $(n-1)$-connected but not n-connected[19]. For instance, since any contractible space X is trivially n-connected for all n, $\mathrm{conn}(X) = \infty$. The torus $S^1 \times S^1$ is 0-connected but not 1-connected and so $\mathrm{conn}(S^1 \times S^1) = 0$. Finally, the *homotopy dimension*, $\dim(X)$ of a finite space X is the minimal dimension of a polyhedron homotopy equivalent to X[20]. We are now ready for the definition:

An *atom* is a finite space X which is indecomposable and which satisfies

$$\dim(X) = 2\mathrm{conn}(X) + 1.$$

17 More precisely, it is a pushout but the idea is clear enough.

18 We haven't defined the notion of a finite polyhedron. The latter is usually given via the notion of a finite simplicial complex. The informal idea should be clear enough. See Rotman (1988, Chap. 7).

19 See Hatcher (2002, p. 346) for equivalent definitions.

20 Again, this relies on the definition of the dimension of a finite polyhedron. A precise definition can be given but since it captures the usual geometric idea of the dimension of a polyhedron, we will simply skip it. See Rotman (1988, Chap. 7) for details.

Notice that by definition the dimension of an homotopical atom is always an odd number. The homotopy type of the 1-sphere, i.e. the circle, S^1 is an atom and in fact the only atom of dimension 1. The homotopy type of the 2-sphere, i.e. the sphere, S^2 is not an atom, for $\dim(S^2) = 2$ and $\operatorname{conn}(S^2) = 1$[21]. Consider now the torus $S^1 \times S^1$. It is certainly indecomposable. But its homotopy type is not an homotopical atom since it has $\dim(S^1 \times S^1) = 2$ and $\operatorname{conn}(S^1 \times S^1) = 0$ and thus, it does not satisfy the equation of the definition (not to mention the fact that its dimension is even).

It can be shown that any finite space X with $\dim(X) \leq 2\operatorname{conn}(X) + 1$ is homotopy equivalent to a finite one point union $X_1 \vee \cdots \vee X_r$ of (suspended) atoms X_i and that each finite space X can be split into atoms[22]. Last but not least, there is a complete list of atoms of dimension ≤ 11 (see Baues 1995, 1996, 2002).

One methodological remark has to be brought forward: as should be clear from the foregoing remarks, homotopy types are never studied directly, so to speak. One has to pick tokens of a type and work with them, making sure that all the constructions and calculations are indeed done up to homotopy. As we will see, this situation lifts up directly to the general and abstract case.

Let us pause for a moment to reflect on the possibility that homotopical atoms could be sets with a structure. After all, the circle S^1 is the generic token of the one-dimensional homotopical atom and it certainly can be defined as a set of points satisfying a simple equation. But this will not do, at least for two reasons. First, by pulverizing a circle into its points, we are applying "forces" that are neither relevant nor legitimate. It is certainly not legitimate, since we have chosen one particular token of a homotopical atom and considering *its* set of points as being constitutive is entirely arbitrary. Second, and this is certainly the most important reason, the only relevant properties of a homotopy type are those that are preserved by homotopy equivalences and, as we have seen, the cardinality of a set of points of a token of a homotopy type is not preserved by homotopy equivalences.

Again, we seem to be forced to acknowledge the fact that sets of points are irrelevant to the nature of homotopy types. To get a better grip on the structure of homotopy types in general—and not only of finite polyhedra—we have to go into some homotopy theory, both classical and contemporary.

5 Searching properties of homotopy types: classical homotopy theory

It is extremely hard to prove that two spaces are *not* homotopy equivalent. In fact, one can say that homotopy *theory* was more or less built up just for that purpose: to provide invariants of spaces (and nowadays various kinds of mathematical structures) such that whenever two spaces have different invariants, then they are not homotopy equivalent, that is they are tokens of different homotopy types. Ideally, a homotopy type would be completely characterized by some invariants associated to it. Today, the basic invariants of homotopy theory are provided by the so-called *homotopy groups*.

21 But it is what Baues calls a *suspended* atoms since it is the suspension of the 1-sphere.

22 This is not quite precise but we leave the exact technical statement aside since it has no direct bearing on our argument.

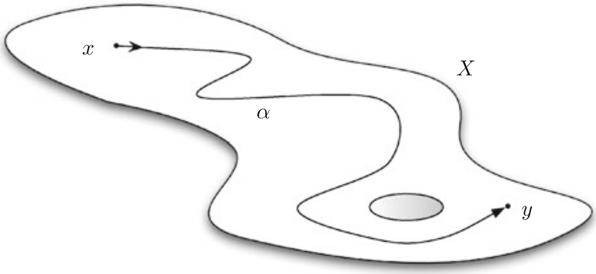


Fig. 2 A path α from x to y in X

The first homotopy invariant is in fact a *set*[23]! It is the set of path components of a space. To understand how it arises, one has to define the path components of a space. Given a topological space X, a *path in* X is a continuous map $\alpha : [0, 1] \to X$ (from now on, we will denote the unit interval [0, 1] with the canonical topology by I). A *path from* x *to* y *in* X is a path α such that $\alpha(0) = x$ and $\alpha(1) = y$. The point x is called the *origin* of the path and the point y, the *end point* of the path (Fig. 2). Whenever the origin and the end coincide, that is when $x = y$, a path is called a *loop based at* x.

Paths are used to define an equivalence relation between points of a space X: $x \simeq y$ in X if there exists a path α such that $\alpha(0) = x$ and $\alpha(1) = y$. Whenever this is the case, x is said to be *connected* with y by a path α. It easy to convince oneself that it is indeed an equivalence relation. A space X is said to be *path connected*[24], if $x \simeq y$ for all points x and y of X. An equivalence class $[x]$ of this equivalence relation is called a *path component* of X. The *set* of path components of a space X is denoted by $\pi_0(X)$. Hence a space X is path connected if and only if it has only one path component. Informally, the set $\pi_0(X)$ can be thought as measuring the number of continuous parts of X, if a part is thought of as constituting a whole whose unity is provided by the fact that any two of its points can be joined to one another by a continuous path.

The set of path components of a space X is an invariant in the following sense. First, any continuous function $f : X \to Y$ induces a function $f_* : \pi_0(X) \to \pi_0(Y)$. Second, the identity function $\mathrm{id}_X : X \to X$ induces the identity function $(\mathrm{id}_X)_* : \pi_0(X) \to \pi_0(X)$. Third, given two continuous maps $f : X \to Y$ and $g : Y \to Z$, $(g \circ f)_* = g_* \circ f_* : \pi_0(X) \to \pi_0(Z)$. This simply means that moving to path components of spaces is a functorial construction. In particular, this implies that if $f : X \to Y$ is a homeomorphism, then $f_* : \pi_0(X) \to \pi_0(Y)$ is an isomorphism. Clearly, homotopy types differ by the number of path components they have: a homotopy type with one path component, that is a path connected space, is different from a homotopy type with two path components or any number of path components different from 1. Thus, more generally, if a space X has n path components and a space Y has m path components and $m \neq n$, then X and Y are tokens of different homotopy types.

[23] The fact that sets naturally show up in this picture does not indicate that we need a set *theory* in order to develop the theory. Sets are derived and play a minor theoretical role.

[24] Given the definition of the previous section, being path connected is the same as being 0-connected.

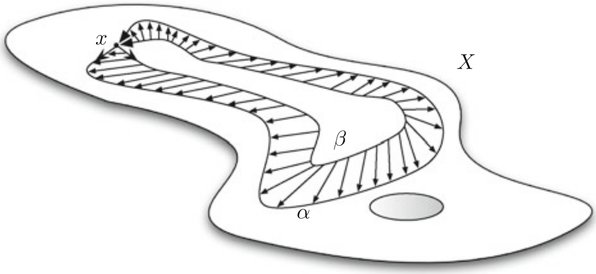


Fig. 3 A homotopy from loop β to loop α in X

The next homotopy invariant is the first *algebraic* invariant and it played a key role in the development of algebraic topology. It is the *fundamental group* of a space X. To understand how it is defined, we have to go back to paths in a space X. The following facts are needed to define the fundamental group:

1. Given a point $x \in X$, the *constant path* or *constant loop* at x is defined by $c_x(t) = x, \forall t \in I$;
2. Paths compose: given paths $\alpha : I \to X$ and $\beta : I \to X$ such that $\alpha(1) = \beta(0)$, then the *product* $(\alpha\beta)$ can be defined in the obvious manner;
3. Given a path $\alpha : I \to X$, the *inverse path* $\alpha^{-1} : I \to X$ can be defined;
4. The concept of a homotopy between two paths $\alpha, \beta : I \to X$ with the same origin and end point can be defined (Fig. 3);
5. A loop homotopic to a constant loop is said to be *nullhomotopic* or *contractible*;
6. A space X is said to be *contractible* to x_0 if the identity map $\text{id}_X : X \to X$ is nullhomotopic. In other words, there exists what is called a contraction: a map $D : X \times I \to X$ defined by $D(x, 0) = x$, $D(x, 1) = x_0$. As we have already seen, a contractible space has the homotopy type of a point.

Thus, once again, being homotopic defines an equivalence relation $\alpha \simeq \beta$ between paths with the same endpoints and, in particular, between loops based at the same point. Given a space X with a point x_0, sometimes called a based point, one considers the homotopy classes of loops based at x_0. It can be shown from the foregoing facts that the set of *homotopy classes of loops based at* x_0 thus defined form a *group*, the fundamental group of X based at the point x_0. A space X with a selected based point x_0 is called a *pointed space* and is denoted by (X, x_0). More formally, the fundamental group of X based at the point x_0 is defined by

$$\pi_1(X, x_0) = \{[\alpha] \mid \alpha \text{ is a loop based at } x_0\}.$$

It can be shown that $\pi_1(\mathbb{R}, 0) \cong 0$, the trivial group with one element. In fact, the fundamental group of any contractible space is trivial. Although it requires more calculations to prove, it can be established that $\pi_1(S^1, p) \cong \mathbb{Z}$, where the point p is taken to be the north pole by convention. Thus, although the real line and the circle are both path-connected spaces, their fundamental groups differ. The latter detects the fact, so to speak, that $\mathbb{R}$ is contractible whereas S^1 is not.

π_1, just like π_0, is a functorial construction and, thus, it is homotopy invariant. More formally, a *pointed* map, that is a map $f : (X, x_0) \to (Y, y_0)$ between pointed spaces such that $f(x_0) = y_0$, induces a group homomorphism $f_* : \pi_1(X, x_0) \to \pi_1(Y, y_0)$. Given two pointed maps $f : (X, x_0) \to (Y, y_0)$ and $g : (Y, y_0) \to (Z, z_0)$, then $(\mathrm{id}_X)_* = \mathrm{id}_{\pi_1(X,x_0)} : \pi_1(X, x_0) \to \pi_1(X, x_0)$ and $(g \circ f)_* = g_* \circ f_* : \pi_1(X, x_0) \to \pi_1(Z, z_0)$. Finally, and this is the precise formulation that the fundamental group is homotopy invariant, if $f : X \to Y$ is a homotopy equivalence, then the induced homomorphism $f_* : \pi_1(X, x_0) \to \pi_1(Y, f(x_0))$ is an isomorphism for every point $x_0 \in X$. Thus, since there cannot be an isomorphism between $\pi_1(\mathbb{R}, 0)$ and $\pi_1(S^1, p)$ and π_1 is homotopy invariant, we can conclude that there cannot be an homotopy equivalence between $(\mathbb{R}, 0)$ and (S^1, p). From there, it follows easily that the spaces $\mathbb{R}$ and S^1 cannot be homotopy equivalent.

A space X is said to be *simply-connected* if it is path-connected and the fundamental group is trivial, that is $\pi_1(X, x_0) = 0$, for some base point x_0. Equivalently, a space is simply-connected if every loop is nullhomotopic. Being simply-connected is an important concept of topology and it is inherently homotopical[25].

Homotopy groups do not stop at π_1. Historically, it took some time before mathematicians realized that the groups π_n for $n > 1$ were of mathematical value. Indeed, Čeck defined them at the World Congress of Mathematics in Zurch in 1932, but was apparently convinced by Hopf and Alexandroff to set them aside, probably because they noticed that the π_n's are all abelian for $n > 1$ and thought that they would be useless since higher homology groups are abelian in general (see Alexandroff 1962). Hurewicz reintroduced them in full generality in a series of four papers published in 1935–1936, proved that they were systematically related to homology groups in an interesting and fundamental fashion and went on to show novel and useful theorems with their help (incidentally, Hurewicz also introduced the concept of homotopy types in these papers). From that point on and to this day, mathematicians tried and are still trying to find ways to compute them, a task that turned out to be extraordinarily difficult.

The general definition of higher homotopy groups is essentially geometric. One starts with the collection of continuous pointed maps from the pointed n-sphere (S^n, p) into a pointed space (X, x_0), which is usually denoted by

$$\mathrm{Hom}((S^n, p), (X, x_0)).$$

Pictorially, one looks at images of n-spheres that are nailed at x_0. But as with the fundamental group, we have to consider *homotopy classes* of continuous pointed maps from (S^n, p) into (X, x_0), which is denoted by $[(S^n, p), (X, x_0)]_*$. This means that we consider images of n-spheres up to homotopy, that is whenever there is a homotopy between two such images, they are considered as being essentially the same. The latter structure has a natural group structure and is the n-th homotopy group π_n of (X, x_0) (for the proof that it has a group structure, see Hatcher 2002, Chap. 4 or Rotman 1988).

25 In the terminology of the second section, a simply-connected space is 1-connected. The 2-sphere S^2 is simply-connected but the 1-sphere, the circle, S^1 is not.

Formally,

$$\pi_n(X, x_0) = [(S^n, p), (X, x_0)]_*.$$

From now on, we will simply write $\pi_n(X)$ and $[S^n, X]$.

Considering the case when $n = 1$ in the foregoing definition, we get that

$$\pi_1(X) = [S^1, X].$$

Clearly, homotopy classes of maps of circles with a base point into a base point x_0 in X is the same as considering homotopy classes of loops based at x_0 in X and thus we get a formally equivalent definition of the fundamental group.

In the best of possible worlds, homotopy groups would characterize homotopy types completely. In that world, a space X would be homotopy equivalent to a space Y, i.e. they would be tokens of the same homotopy type, if and only if $\pi_n(X) \cong \pi_n(Y)$ for all n. However, we do not live in such a world and our world is more subtle and more interesting. In our world, we can nonetheless introduce a weaker definition: a map $f : X \to Y$ is a *weak homotopy equivalence* if for each base point $x \in X$, the induced map $f_* : (X, x) \to (Y, f(y))$ is a bijection of sets for $n = 0$ and an isomorphism of groups for $n \geq 1$[26]. In fact, this definition introduces a weaker form of homotopy equivalence: two spaces X and Y are *weakly homotopy equivalent* if there is a weak homotopy equivalence between them. Notice, however, that the latter definition depends directly on the existence of a continuous map $f : X \to Y$ and although the definition is reflexive and transitive, it is clearly not symmetric. Thus, as such, it does not generate an equivalence relation between spaces. The continuous map $f : X \to Y$ does not necessarily have an inverse. It remains to be seen whether it is possible to "force" the existence of an inverse so that this relation becomes a genuine equivalence relation between spaces. It turns out that is it and thus we have two kinds of homotopy types. In the case of a weak homotopy equivalence, there is a map from space X into Y such that it sparks off the machinery of homotopy groups and the latter yield the same readings. Thus, we cannot detect a difference between X and Y with our machinery. In the case of a homotopy equivalence, we are in a situation to actually deform continuously the space X into the space Y and vice-versa. I believe that even the notion of weak equivalence—which is now prevalent in contemporary mathematics—is philosophically important. It is a case where there is a link between two different systems or two different states of the same system and, although the link is not reversible so that it is possible to go back to the identity, the two systems or the two states of the system can nonetheless be identified, at least as far as the specific properties are concerned.

Restricting the notion of weak homotopy equivalence to "nice" spaces, we get Whitehead's theorem: if X and Y are connected CW-complexes and if $f : X \to Y$ is a continuous map such that $f_* : \pi_n(X, x_0) \to \pi_n(Y, y_0)$ is an isomorphism for all n, then f is a homotopy equivalence, i.e. X and Y are tokens of the same homotopy

[26] It is of course possible to give examples of weak homotopy equivalences between spaces that are not homotopy equivalent. See Hatcher (2002, Chap. 4).

type[27]. Thus, for CW-complexes, being weakly homotopy equivalent and being homotopy equivalent coincide (see Hatcher 2002, Chap. 4 for a proof of Whitehead's theorem).

In the case of *simply*-connected CW-complexes, Whitehead's theorem takes the following form: if X and Y are simply-connected CW-complexes and $f : X \to Y$ a continuous map, then f is a homotopy equivalence if and only if one of the following two equivalent conditions hold:

1. the map f induces an isomorphism of homotopy groups $f_* : \pi_n(X) \to \pi_n(Y)$ for $n \geq 2$;
2. the map f induces an isomorphism of homology groups $f_* : H_n(X) \to H_n(Y)$ for $n \geq 2$.

Thus, homotopy groups and homology groups are sufficiently powerful and fine-grained to characterize simply-connected homotopy types. Not surprisingly, homotopy groups and homology groups encode 'connectedness' and 'dimension' and as we have indicated in Sect. 2, they are the basic invariants of homotopy types.

> The main numerical invariants of a homotopy type are 'dimension' and 'degree of connectedness'. [...]
> The dimension is related to homology since all homology groups above the dimension are trivial, whereas the degree of connectedness is related to homotopy since below this degree all homotopy groups vanish. It took a long time in the development of algebraic topology to establish homology and homotopy groups as the main invariants of a homotopy type. (Baues 1995, pp. 12–13)

We are simplifying considerably here, leaving aside the question of a geometric realization of a map (see Baues 1996). In fact, the situation is slightly more complicated: it has been shown that two simply-connected CW-complexes have the same homotopy types if and only if they have the same homotopy groups, the same homology groups and the same Postnikov invariants—in the last case, cohomology groups and operations creeps in. Computing these invariants is extraordinarily hard and a host of intricate devices like loop spaces, suspensions, exact sequences, fibrations, spectral sequences, spectra, various categories and functors, etc. are needed, even for reasonably "nice" spaces (see Baez and Shulman 2006; Baues 1996; Hatcher 2002).

Thus, the structure of homotopy types is captured by, basically, homotopy groups, homology groups and other algebraic invariants, some of which still have to be discovered. We have to pause and reflect on what we have so far. Homotopy types were originally given by an underlying equivalence relation on topological spaces and continuous maps between them. However, this equivalence relation does *not* yield a criterion of identity for homotopy types themselves. For that, as in all the other similar cases, one has to have a theory of these entities from which the appropriate criterion of

[27] In fact, a CW-complex is always homotopy equivalent to a polyhedron and is, in that sense, a "nice" space. One has to be careful here: as Hatcher rightly emphasizes, Whitehead's theorem does *not* say that two CW complexes with isomorphic homotopy groups are homotopy equivalent. It says that if there is a map that induces isomorphisms between homotopy groups, then that map is a homotopy equivalence. Hatcher (2002, p. 348) gives examples of spaces that have isomorphic homotopy groups but that are *not* homotopy equivalent.

identity is extracted. As we have said, in an ideal world, homotopy groups would have provided the criterion of identity for homotopy types. But such is not our mathematical world and even if it had been, we would still be facing a rather different theoretical context than the ones we have been accustomed to work with.

We have to compare and contrast the case of homotopy types with other cases of mathematical entities, for instance topological spaces themselves and groups. Let us first come back to the way we have been trained to think about these systems. A topological space X is a set and a family of subsets of X satisfying the usual closure conditions. Then, one works with the (finite) intersections and the (arbitrary) unions, together with other set-theoretical operations, to define significant subspaces and superspaces. A group G is a set together with a binary operation and a constant satisfying the usual axioms for groups. Again, one uses this operation and its properties, together with set-theoretical properties compatible with the operation, to find properties of groups and define significant subgroups and supergroups. This is the general pattern underlying the standard mathematical structures expressed in a first-order language and interpreted in a universe of sets. In the case of topological spaces and groups, it seems perfectly fine to say that the structure stipulated by the axioms of the theory restricts the underlying structure of subsets implicit in the definition of these entities. The underlying set of a given group G has a lattice of subsets within which the lattice of subgroups of G sits.

Let us now go back to homotopy types. I have never seen a mathematical paper on homotopy theory starting with "Let X be a homotopy type…". We are not starting with a first-order theory interpreted in the universe of sets. No one would say, as far as I know, that a homotopy type is a set together with the following operations satisfying such and such property or a set with a family of subsets such that it closed under such and such operations or any variation on these themes. We could and might start with a first-order theory, but its interpretation will have to be in a different context altogether. One obvious possibility is to interpret the primitives of the theory directly in a universe of homotopy types or of some other universe in which the latter can be comprehended. But for that to be possible, we have to have a better understanding of that universe. Homotopy types have a structure and homotopy types have properties but these properties are intrinsically associated with intricate algebraic constructions, e.g. homotopy groups, homology and cohomology groups, Postnikov towers, etc. There are general constructions at work, e.g. suspensions, loop spaces, coproducts, (homotopy) limits and colimits. The same remarks apply to weak homotopy types. In fact, so far we are not even sure that we can treat them as genuine types. Both of these entities are intrinsically geometric and, in a sense, over and above any set theoretical property and machinery. They live and thrive in a different environment altogether. That environment is the world of categories and it was precisely in that world that Quillen, following the work of Kan, that homotopy theory found an axiomatic setting. But this is another story altogether, a story in which the extensional point of view is replaced by a purely abstract approach. We will discuss the epistemological aspects inherent to this approach in another paper.

6 Conclusion

We hope to have shown that there are important parts of mathematics in which the notion of identity at work is not based on the purely extensional point of view. We do not want to conclude that the purely extensional point of view has to be abandoned or that it is wrong. However, we do want to indicate that there might be alternatives to it, even from a foundational perspective. Although it is too early to tell, it is conceivable that one might be able to propose a foundational framework in which identities are governed by different principles than the ones we have been assuming progressively in the twentieth century and underly most of recent philosophy of mathematics. There are sketches of such foundational frameworks on the table, for instance Awodey and Warren (2009) or Makkai (1998) and only more research will allow us to see more clearly what lies ahead.

References

Aguilar, M., Gitler, S., & Prieto, C. (2002). *Algebraic topology from a homotopical viewpoint*. New York: Springer.

Alexandroff, P. (1962). *Commemoration of Eduard Čech* (pp. 29–30). Czech Republic: Academia Publishing House of Czechoslovak Academy of Sciences.

Awodey, S., & Warren, M. A. (2009). Homotopy theoretic models of identity type. *Math. Proc. Cambrdige Philos. Soc., 146*(1), 45–55.

Baez, J. C., Shulman, M. (2006). Lectures on n-categories and cohomology. arXiv. math.CT.

Baues, H.-J. (1995). Homotopy types. In I. M. James (Ed.), *Handbook of algebraic topology* (pp. 1–72). New York: Elsevier.

Baues, H.-J. (1996). *Homotopy type and homology*. Oxford: Clarendon Press. Oxford Science Publications.

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

Carter, J. (2005). Individuation of objects—a problem for structuralism?. *Synthese, 143*(3), 291–307.

Carter, J. (2008). Structuralism as a philosophy of mathematical practice. *Synthese, 163*(2), 119–131.

Dieudonné, J. (1985). *History of algebraic geometry: An outline of the history and development of algebraic geometry* Wadsworth mathematics series. Belmont, CA: Wadsworth International Group (J. D. Sally, Trans. from French).

Dieudonné, J. (1989). *A history of algebraic and differential topology, 1900–1960*. Boston: Birkhäuser.

Dwyer, W. G., Hirschhorn, P. S., Kan, D. M., & Smith, J. H. (2004). *Homotopy limit functors on model categories and homotopical categories. Mathematical surveys and monographs* (Vol. 113). Providence, Rhodes Island: American Mathematical Society.

Eilenberg, S., & Mac Lane, S. (1945). A general theory of natural equivalences. *Transactions of the American Mathematical Society, 58*, 231–294.

Grassi, A. (2009). Birational geometry old and new. *Bulletin of the American Mathematical Society (N.S.), 46*(1), 99–123.

Hartshorne, R. (1977). *Algebraic geometry. Graduate texts in mathematics, No. 52*. New York: Springer.

Hatcher, A. (2002). *Algebraic topology*. Cambridge: Cambridge University Press.

Krömer, R. (2007). *Tool and object: A history and philosophy of category theory. Science networks. Historical studies* (Vol. 32). Basel: Birkhäuser.

Mac Lane, S. (1998). *Categories for the working mathematician. Graduate texts in mathematics* (2nd ed., Vol. 5). New York: Springer.

Mac Lane, S., & Moerdijk, I. (1994). *Sheaves in geometry and logic: A first introduction to topos theory*. Universitext. New York: Springer.

Makkai, M. (1998). Towards a categorical foundation of mathematics. In: *Logic colloquium '95 (Haifa). Lecture notes logic* (Vol. 11, pp. 153–190). Berlin: Springer.

Marquis, J.-P. (2006). A path to the epistemology of mathematics: Homotopy theory. In J. Ferreirós & J. J. Gray (Eds.), *The architecture of modern mathematics.* (pp. 239–260). Oxford: Oxford University Press.

Rotman, J. J. (1988). *An introduction to algebraic topology. Graduate texts in mathematics* (Vol. 119). New York: Springer.