Brit. J. Phil. Sci. 32 (1981), 349–358 *Printed in Great Britain*

Category Theory and the Foundations of Mathematics*

by J. L. BELL

In recent years the mathematical discipline of *category theory* has emerged to play a growing role in the perennial debate about the foundations of mathematics. Some of its proponents have suggested that it should replace set theory as the 'official' foundation for mathematics. My purpose here is to discuss the background to this claim and to examine its tenability.

1 BACKGROUND TO THE EMERGENCE OF CATEGORY THEORY

With the creation and refinement of *set theory* in the hands of Cantor, Zermelo, von Neumann and others, the problem of providing an infrastructure for the elaborate developments in areas of mathematics such as abstract algebra, analysis and topology appeared to be solved. The domain of *sets* (structured by *membership*) came to be regarded by most mathematicians (those of a constructive tendency such as the intuitionists being the most notable exceptions) as the source of 'raw material' for building the structures required in all branches of mathematics. Thus, although the notions of *structure* and *operation* on structures had come to play a fundamental role in most mathematical disciplines, these notions were *not* taken as primitive, but were themselves explicated by *reduction* to the more fundamental notions of set and membership. On the other hand, the perspicuity and apparent reliability of the set-theoretical framework enabled mathematicians not primarily interested in set theory (*i.e.*, the vast majority) simply to take this reduction for granted and instead concentrate on isolating and axiomatizing the basic mathematical features of these structures. With the rise of abstract algebra in the 1930s it came to be recognised that these features and the laws governing them (I am thinking particularly of the notions of isomorphism, homomorphism, substructure, *etc.*) had a kind of universality and even inevitability that was apparently *independent* of their set-theoretical origin. Moreover, it was observed that many of the basic notions of abstract algebra could be derived from the

* Paper delivered at the Conference of the British Society for the Philosophy of Science, Chelsea College, September 1980. I am indebted to Solomon Feferman and Mike Hallett for valuable criticisms and suggestions.

single idea of *structure-preserving function*, or, as it is now customarily known, *morphism*. Thus the attitude gradually emerged that the crucial characteristic of mathematical structures is not their internal constitution as set-theoretical entities but rather the relationships among them as embodied in the network of morphisms.[1] This attitude, strikingly reminiscent of the operational structuralism associated with linguistics and psychology, was exemplified most strongly by the Bourbaki school in France, which had proposed a 'structuralist' account of mathematics as far back as the 1930s. However, although the account of mathematics they gave in their 'Eléménts' was manifestly structuralist in *intention*, in *actuality* they still defined structures as *sets* of a certain kind, thereby failing to make them truly independent of their 'internal constitution'. In fact, it was not until the early 1940s that an axiomatic framework emerged which gave better expression to the idea of operational structuralism in mathematics: I am referring to Eilenberg and MacLane's theory of *categories* and *functors*. Here for the first time we have an axiomatic theory which takes the notions of structure and morphism as *primitive* and is indifferent to any particular set-theoretic constitution that structures may have.

It will be helpful to give the formal definitions of category and functor.[2] A *category* E consists of two classes, the members of the first of which—denoted by the letters X, Y, ... —are called *objects*[3] (structures) and the members of the second of which—denoted by the letters f, g, ... —are called *arrows*[4] (morphisms). Each arrow f is assigned an object X as *domain* and an object Y as *codomain*, indicated by writing $f: X \to Y$. If g is any arrow $g: Y \to Z$ with domain Y, the codomain of f, there is an arrow $fg: X \to Z$ called the *composition* of f and g. For each object Y there is an arrow id_Y : $Y \to Y$ called the *identity arrow* of Y. These notions are assumed to satisfy the following identity and associativity axioms:

$$f \cdot id_Y = f, \quad id_Y \cdot g = g, \quad f(gh) = (fg)h$$

for any arrows $f: X \to Y, g: Y \to Z, h: Z \to W$.

Note that these axioms are readily formulated within an appropriate first-order language: the resulting theory is *first-order category theory*.

Given two categories D and E, a *functor* F from D to E consists of a pair of functions (both denoted by F), one from the class of objects of D to that of E, and the other from the class of arrows of D to that of E, such that

$$\text{if } \ f: X \to Y \text{ in } D, \quad \text{then} \quad F(f): F(X) \to F(Y) \text{ in } E;$$

$$F(id_X) = id_{F(X)} \quad \text{for all objects } X \text{ of } E,$$

[1] In particular, it came to be seen that the notion of *identity* appropriate for structures is not set-theoretic equality but *isomorphism*, an idea going back in essence to Dedekind and Klein.

[2] For an account of category theory, see, *e.g.* MacLane [1971].

[3] The use of the term 'object' here in place of 'structure' is intended to reinforce the idea that the theory is not about *particular* structures, but mathematical structures *in general*.

[4] The term 'arrow' here is a metonym and stems from the habit of depicting functions as arrows.

and

$F(fg) = F(f)F(g)$ for any composable arrows f, g of D.

A functor may be thought of as a *morphism of categories*.

Categories and functors abound in mathematics; as examples of the former one may consider:

Set: objects, all sets; arrows, all (set) functions.
Grp: objects, all groups; arrows, all group homomorphisms.
Top: objects, all topological spaces; arrows, all continuous maps.

Prominent functors include:

the 'forgetful' functor from *Grp* or *Top* to *Set* which assigns to each group or topological space its underlying set (this functor has the effect of 'forgetting' the structure!);
the 'free group' functor from *Set* to *Grp* which assigns to each set the free group it generates;
the 'homology group' functors from *Top* to *Grp* which, for each natural number n, assign to each topological space X its nth homology group $H_n(X)$.

This last example played a key role in the creation of category theory. In fact, it was the problem of providing a smooth theory of homology groups, and of the limiting processes involved in their construction, which first suggested the idea of a functor to its creators. It is clear that, in any case, Eilenberg and MacLane regard the idea of *functor* as being in some sense more fundamental than that of *category*. As they remark in their original paper:[1]

> It should be observed that the whole concept of a category is essentially an auxiliary one. Our basic concepts are essentially those of a *functor* and of a natural transformation.... The idea of a category is required only by the precept that every functor should have a definite class as domain and a definite class as range, for the categories are provided as the domains and ranges of functors.

Thus, as far as its creators were concerned, the notion of category was only introduced in order to furnish the more basic notion of functor with set-theoretic legitimacy. Indeed the essentially operational spirit of the enterprise is revealed when they go on to say:

> Thus one could drop the category concept altogether and adopt an even more intuitive standpoint, in which a functor such as 'Hom' [*i.e.* the functor which assigns to each pair of objects of a category the collection of arrows between them] is not defined over the category of 'all' groups but for each particular pair of groups which may be given. This standpoint would suffice for applications, inasmuch as none of our developments will involve elaborate constructions on the categories themselves.

However, the observation that categories were well-nigh ubiquitous in

[1] Eilenberg and MacLane [1945].

mathematics soon convinced the mathematicians concerned of the fundamental importance and autonomy of the notion of category. From a philosophical point of view, a category came to be thought of as an embodiment of the 'abstract structure' that all its constituent objects exemplify, or even literally to be that 'abstract structure'. (Thus for example the category *Grp* is 'group structure'.) So category theory itself came to be viewed as a theory of (mathematical) Structure.

Another consequence of the ubiquity of categories in mathematics was that they came to be regarded as mathematical objects in their own right and thereby subject—as are all mathematical objects—to a whole range of mathematical *constructions*. One of the most important of these is the construction of *functor categories*. Given two categories D and E, the functor category E^D has as objects all functors from D to E and as arrows the so-called *natural transformations* (the definition of which we omit) between such functors. But here the *set-theoretic* definition of category (involving, as it does, the notion of *class*) causes a snag. For if the collection of objects or arrows of D is a *proper class* in the sense of Gödel-Bernays set theory (such a category is called *large*; if both collections are sets the category is called *small*), then so is each functor from D to E and consequently these functors cannot be collected into a class. Thus the current set-theoretical framework does not allow the formation of functor categories E^D when D is large (*e.g.* when D is any of the examples given above). And, for similar reasons, the same is true of the putative category of all categories (small or large).

Now the restrictions on the formation of classes (Zermelo) or on the operations to which these classes can be subjected (von Neumann) had of course been imposed originally in order to resolve the set-theoretic antinomies. These restrictions were not judged by the majority of mathematicians to be of intolerable severity because in most areas of mathematical practice the 'proscribed' collections such as the class of all sets simply did not play any role. (In fact, much early work in the foundations of set theory was devoted to eliminating references in mathematical arguments to these proscribed classes.) But the natural tendency of category theory to form categories consisting of *all* objects of a certain kind once again thrust these proscribed classes, with their attendant difficulties, into the foreground; and the restrictions on the formation and manipulation of these classes imposed by the official set-theoretic framework came to be regarded by some category theorists as an irksome and possibly even unnecessary curtailment of their mathematical activity. This feeling, reinforced by the steady incursion of category-theoretic techniques and ideas into many branches of mathematics (regarded as 'subversion' by some of the more conservatively minded) led certain category-theorists first to question the *adequacy* of the current set-theoretic foundation, then its *necessity*, and finally to propose category theory *itself* as a possible foundation for mathematics. I will argue here that, although there are grounds for accepting that current set-theoretical foundations are inadequate for 'full' category theory, and although category theory unquestionably has foundational *significance*, its very nature makes it unsuitable as an exclusive *foundation* for mathematics.

2 CATEGORY THEORY: A FOUNDATION FOR MATHEMATICS?

In what sense could category theory serve as a foundation for mathematics? There seem to be (at least) two possible senses: first, a strong sense, in which *all* mathematical concepts, *including* those of the current logico-metatheoretical framework for mathematics, are explicable in category-theoretic terms. And secondly, a weaker sense in which one only requires category theory to serve as a (possibly superior) substitute for axiomatic set theory in its present foundational role.

Now it seems to me implausible that category theory is, or could be, foundationally adequate in the stronger sense. For consider the meta-theoretical framework in which category theory (or any other first-order theory) is embedded. This framework has two basic aspects: the *combinatorial*, which is concerned with the formal, finitely presented properties of the inscriptions of the ambient formal language, and the *semantical*, which is concerned with the interpretation and truth of the expressions of that language. Neither one of these aspects is—at present—reducible to the other. The former deals with *intensional* objects such as proofs and constructions whose actual *presentation* is crucial, while the latter employs extensional objects such as classes whose identity is determined independently of how they may be presented or defined. So if category theory is to furnish a foundation for mathematics in the stronger sense, it must provide convincing accounts of *both* of these aspects. But a category is defined to be a class of a certain kind, and classes are *extensional*, while combinatorial objects are generally *not*. Since there is no reason to suppose that a satisfactory account of intensional (*i.e.* non-extensional) objects can be given solely in terms of extensional ones, it seems to me that category theory as currently formulated in terms of classes must fail to provide a faithful account of the combinatorial aspect at least. (Of course for similar reasons this 'weakness' is shared by set theory.)

As far as the *semantical* aspect is concerned, we recall that the interpretation of an expression of a classical first-order language involves a reference to *classes* or *pluralities* in an essential way (as the 'range' of the variables in the expression). In particular, grasping the concept of *logical truth* for sentences of classical first-order languages requires that one has already grasped the concept of class. To put it another way, the concept of class is *epistemically prior* to the concept of (classical) logical truth. So if category theory is to serve as an autonomous basis for classical semantics, and in particular give a satisfactory independent account of logical truth, it must be possible to give an explication of the notion of class (at least in so far as it is involved in deriving the concept of logical truth) *solely in terms of the notion of category*, and without already having defined the latter notion in terms of classes. But this seems to me highly dubious, for it is surely the case that the unstructured notion of class is epistemically prior to any more highly structured notion such as category: in order to understand what a category is, you first have to know what a class is.[1] This also applies, *mutatis mutandis*, to the notion of *functor* whose explication involves grasping the epistemically prior notion of *operation*.

It seems to me that these considerations show that category theory as currently conceived is not capable of serving as a foundation for mathematics in the strong sense. Of course, this is hardly surprising since it is widely recognised that *no* single foundational scheme is at present capable of providing a convincing explication of *both* combinatorial and set-theoretical objects. What we actually possess is an informal system of 'multiple' foundations, with distinct combinatorial and set-theoretical components.

Let us turn now to the weaker sense in which category theory could serve as a foundation for mathematics, namely as a substitute for axiomatic set theory in its current foundational role. One possible means of achieving this would be to construct a formal *interpretation* of some 'foundationally adequate' first-order version of set theory (*e.g.* Zermelo–Fraenkel set theory with choice, *ZFC*) in a suitable consistent extension *T* of first-order category theory in such a way that the interpretation of any theorem of *ZFC* is provable in *T*. Now this has in fact already been achieved:[2] *T* may be taken to be the theory of *elementary toposes* (a finite extension of the first-order theory of categories), augmented by certain other axioms, notably category-theoretic versions of the axiom of choice and the axiom scheme of replacement.

The interpretation of *ZFC* in *T* is performed as follows. First, one observes that, by the well-known Mostowski collapsing lemma, the notion *transitive set* essentially corresponds to the notion *extensional well-founded relation*. Now a relation may be regarded as a function from its field to the power set of its field, *a fortiori* as an arrow within the category of sets. Furthermore, extensionality and well-foundedness of a relation can be translated into purely categorical properties of the corresponding arrow. These properties may be 'lifted' to any category which is a model of *T* (in fact any topos). An arrow possessing these properties then provides a category-theoretic formulation of the notion of transitive set and is accordingly called a *transitive set-arrow*. In view of the fact that every set is a subset of a transitive set, one defines a *set-arrow* to be a subarrow of a transitive set-arrow, *i.e.* a pair (f, r) consisting of a transitive set-arrow r and a monic arrow f such that the codomain of f coincides with the domain of r. One can then define the relation of *membership* between set-arrows and with some difficulty show that, for any model E of T, the collection of set-arrows

[1] For a similar conclusion, see Feferman [1977]. My argument here owes much to this article.
[2] See Osius [1974].

in E, structured by this new relation of membership, is a model of ZFC. Thus, if we interpret the notion of *set* as *set-arrow* we obtain a translation[1] of ZFC into T.

We may conclude from all this that it would be *technically* possible to give a purely category-theoretic account of all mathematical notions expressible within axiomatic set theory, and so formally possible for category theory to serve as a foundation for mathematics insofar as axiomatic set theory does. On the other hand, as we have seen above, the actual translation of set theory into category theory is awkward and has (unlike the basic category-theoretic notions themselves) a factitious character which renders it unsuitable as a means of formalizing those mathematical notions which are normally expressed set-theoretically.

What this translation of set theory really amounts to is the replacement of the notion 'mathematical object as *set*' by the notion 'mathematical object as *pair of arrows* (of a certain kind) in a category'. It would seem, however, that a more convincing and natural formalization of mathematics within category theory would be obtained if mathematical objects could be construed as categories *tout court*. (This would also be more in keeping with the structuralist view that mathematical objects are given as structures and that categories provide an embodiment of the idea of structure.) To this end Lawvere has formulated a first-order theory[2] Σ in the language of categories which purports to capture the characteristic features of the category of all small categories. Furthermore, he claims that all mathematics currently formulable within axiomatic set theory can be expounded within Σ. Unfortunately, certain technical defects in Σ have come to light which make it difficult to assess whether Lawvere's bold program can be carried out in its original form. And even if we grant that these technical difficulties will be overcome, it seems doubtful whether a system like Σ will ever be accepted as an autonomous 'foundation' for mathematics in the sense that, say, ZFC is. For, in developing the notions of workaday mathematics within Σ it seems to be necessary to bring in the notion of set 'through the back door', so to speak, in the form of 'trivial' or 'discrete' categories. Of course, this is not technically an appeal to the notion of set because the definition of discrete category is given entirely within Σ. But the question automatically arises as to exactly *why*, in introducing ordinary mathematical notions into the theory, one must make a detour through the somewhat opaque notion of discrete category. It is difficult to see how this can be explained except by appeal to the notion of 'unstructured' category, *i.e. set*. This, it seems to me, will inevitably make a system like Σ appear artificial as a 'foundation' for mathematics, despite the beauty and naturalness of the category-theoretic notions themselves.

[1] There is also an obvious translation of T into ZFC: simply interpret the notion *object* as *set*, and *arrow* as *mapping*. Thus ZFC and T are actually formally *equivalent*.

[2] Lawvere [1966].

3 THE NATURE AND SIGNIFICANCE OF CATEGORY THEORY

In expressing these negative views about category theory as a potential foundation for mathematics, I do not mean to imply that current set-theoretical foundations should be taken as adequate for everything the category-theorists want to do, nor do I wish to claim that category theory is just another branch of mathematics, devoid of foundational content. On the contrary, with regard to the former, I have already pointed out that the operations on large categories which appear so natural to category-theorists are *not* justified by current set-theoretical foundations and so appear to demand an extension or reformulation of the set-theoretical framework to accommodate them. In this connection, however, it should be noted that the failure of set theory to justify the unlimited application of category-theoretic operations is a consequence of its success in eschewing the overcomprehensive collections which were originally deemed responsible for the paradoxes. (This fact was clear to Eilenberg and MacLane, for they point out in their original paper that in category theory 'no essentially new paradoxes [over and above those of intuitive set theory] are apparently involved'.) In fact, set theory's failure to embrace the notion of arbitrary *category* (or structure) is really just another way of expressing its failure to capture completely the notion of arbitrary *property*. This suggests the possibility that a suitable framework for 'full' category theory could reasonably be sought within a theory of such arbitrary properties. Although such a theory would certainly have to transcend current set theory, *it would however be under no obligation to appear category-theoretic in nature*. The notion of a category would in all probability continue to be a *derived* notion and not a primitive one.[1]

Nevertheless, category theory is more than just another abstract mathematical theory. Like set theory, it provides a general framework for dealing with mathematical structures, and—again like set theory—it achieves this by transcending the *particularity* of structures. But set theory and category theory go about doing this in entirely different ways. Set theory strips away structure from the ontology of mathematics leaving pluralities of structureless individuals open to the imposition of new structure. Category theory, on the other hand, transcends particular structure, not by doing away with it, but by *generalising* it, that is, by producing an *axiomatic general theory of structure*. The success of category theory, and its significance for foundations is due to the *ubiquity of structure* in mathematics.

It may be said that category theory, while still dependent on set theory as the ultimate source of mathematical entities, nonetheless frees mathematics from the particular *form* imposed on it by having to regard these entities as pluralities of elements. The power and fertility of the 'element-free' formulation of mathematics provided by category theory is most strikingly realised by applying it to set theory *itself*. This gives rise to a startling new

[1] See Feferman [1977] for an 'intensional' theory of partial operations and properties in which much of 'full' category theory can be formulated.

kind of 'model' of set theory called an (elementary) *topos*.[1] A topos is a category E which has the following features in common with the category *Set* of sets.

(1) E has a terminal object and all finite products.

(2) There is in E a 'truth-value' object Ω which plays the same role in E as the truth-value set $2 = \{0, 1\}$ plays in *Set*, *i.e.* for each object X there is a natural correspondence between subobjects of X and arrows from X to Ω ('characteristic functions' on X).

(3) For each object X of E there is a 'power object' PX in E which plays the formal role of a power set of X in E.

These conditions can all be formulated in the first-order language of category theory: hence the use of the term 'elementary'. In addition to the category of sets, examples of toposes include the category of sheaves of sets on a topological space (a sheaf may be thought of as a set varying through space) and the category of all diagrams $X_0 \to X_1 \to X_2 \to \ldots$ of sets X_i (such a diagram may be regarded as a set varying through discrete *time*). A topos may thus be thought of as a generalized model of set theory in which the 'sets' are varying in some manner and are generally not determined by their 'points'.

Topos theory has striking connections with *logic*. For example, in any topos one can give natural definitions of arrows (to be thought of as 'logical operations' in E), $\neg : \Omega \to \Omega$; $\wedge, \vee, \Rightarrow : \Omega \times \Omega \to \Omega$ in such a way that, if we regard these arrows as algebraic operations on Ω, the resulting algebra is a Heyting algebra, *i.e.* satisfies the laws of *intuitionistic* propositional logic. In this sense intuitionistic logic is 'internalised' in a topos (although of course there will be toposes in which *classical* logic is internalised, *e.g.* the category of sets). With some justice, then, we may regard a topos as an instrument for reducing *logic* to *mathematics*, the remarkable thing being that the logic obtained is not (in general) classical, but intuitionistic.

Because the internal logic of a topos is intuitionistic, one would intuitively expect mathematics done 'in a topos' to be in some sense constructive. However, at present it is not clear just how this intuitive expectation is to be explicated. Certainly there seem to be strong connections between *intuitionistic* mathematics and the forms that mathematical arguments take within a (general) topos. This applies both to arithmetic and to the analysis of the continuum, which can be made to resemble the intuitionistic continuum in that it will not be built up from 'points'. In fact one can construct toposes in which the unit interval $[0, 1]$ fails to be compact and in which every function from reals to reals is continuous. Topos theory thus provides an entirely new framework for modelling intuitionistic mathematics.

It is generally agreed that the most significant recent contribution to mathematical logic was Paul Cohen's construction of models of set theory in

[1] For an introductory account of topos theory, see Goldblatt [1979].

which the continuum hypothesis and the axiom of choice fail. The techniques he invented have led to an enormous proliferation of essentially different models of set theory and the rise of a 'relativistic' attitude toward the set-theoretical foundations of mathematics. This attitude involves abandoning (or, at least, reserving judgment about) the idea that mathematical constructions should be viewed as taking place within an 'absolute' universe of sets with fixed and predetermined properties. Instead, one works in suitably chosen *models* of set theory having the properties required to carry out the constructions in question. Now topos theory carries this idea[1] to its conclusion: it provides models of set theory in which even the *logic* can be tailored to suit the construction. This fact may provide further evidence for logical *pluralism* within mathematics, as already exemplified by the existence of classical and constructive logics. Be this as it may, the paramount achievement of topos theory is to have identified the basic core of set theory in such a way that the set concept becomes manifest in contexts (such as algebraic geometry or constructive mathematics) where before its presence was at most tacit. Thus category theory, far from being in opposition to set theory, ultimately enables the set concept to achieve a new universality.

London School of Economics

REFERENCES

EILENBERG, S. and MACLANE, S. [1945]: 'General Theory of Natural Equivalences', *Transactions of the American Mathematical Society*, **58**, pp. 239–94.

FEFERMAN, S. [1977]: 'Categorical Foundations and Foundations of Category Theory' in Butts and Hintikka (*eds.*), *Logic, Foundations of Mathematics and Computability Theory*. Dordrecht: D. Reidel, pp. 149–69.

GOLDBLATT, R. [1979]: *Topoi, the Categorial Analysis of Logic*. Amsterdam: North-Holland Publishing Company.

LAWVERE, F. W. [1966]: 'The Category of Categories as a Foundation for Mathematics', in *Proceedings of the La Jolla Conference in Categorical Algebra*, Springer-Verlag, pp. 1–20.

MACLANE, S. [1971]: *Categories for the Working Mathematician*. Berlin: Springer-Verlag.

OSIUS, G. [1974]: 'Categorical Set Theory: A Characterization of the Category of Sets', *Journal of Pure and Applied Algebra* **4**, pp. 79–119.

[1] It is interesting to note that Lawvere and Tierney's formulation of the notion of topos was partly motivated by the idea of analysing Cohen's proof of the independence of the continuum hypothesis in category-theoretic terms.