Category Theory: an abstract setting for analogy and comparison

R. Brown and T.Porter

Abstract

'Comparison' and 'Analogy' are fundamental aspects of knowledge acquisition. We argue that one of the reasons for the usefulness and importance of Category Theory is that it gives an abstract mathematical setting for analogy and comparison, allowing an analysis of the process of abstracting and relating new concepts. This setting is one of the most important routes for the application of Mathematics to scientific problems. We explore the consequences of this through some examples and thought experiments.

Let us start with a quote from Shakespeare's 'A Midsummer Night's Dream'.

The Poet's eye in a fine frenzy rolling
Doth glance from heaven to earth and from earth to heaven
And as imagination bodies forth the form of things unknown
The Poet's pen turns them to shapes and gives to airy nothing
A local habitation and a name.

We feel many mathematicians would liken this description of the role of the Poet to their own attitudes to Mathematics. Undoubtedly, mathematics has coined many names for the forms of things previously unknown. The notion of 'name', and its role in thought, is very subtle.

In trying to address the question 'What is Category Theory?', we felt it necessary to reflect on 'What is Mathematics?' and 'How does Category Theory fit into its overall structure?' In fact, as Category Theory is being seen as having potential uses in other parts of Science, perhaps its position within Science as a whole might also be useful to consider.

Our initial view of Category Theory, both within Mathematics and more widely, suggested various 'themes', in no particular order:

- mixed algebraic and geometric/combinatorial structures,
- enables comparison between objects,
- a formalisation and abstraction of 'analogy, and from 'analogy' to 'abstraction' itself,

- a structural context for structural mathematics,
- combines local structure with locational information,
- gives meaning to the idea of structure,
- comparisons between concepts,

and the list goes on. In our discussion here, we will try to address the meaning of some of these themes and the reasons they seem to us to give insights into the question of what Category Theory is.

Comparison and Analogy in Mathematics and Science.

Central classical themes in Science include the classification of objects within a particular context. As an example, take the classification of minerals. Various elementary attributes can be noted: this one is green, that one is reddish brown. The crystals of the first have this shape, which is not the same as that of the second, and so on. We then classify the two minerals into different classes with different names, initially just as a useful verbal label, later as our knowledge evolves to include some of the *key* attributes for the context, either in quite ordinary language (green, hexagonal, ...) or encoded, for instance, by letters, numbers and symbols as with a chemical formula. Of course, the question of why they look different is then central to the next depth of the study, but first the possible hierarchical class structure has to be determined by *comparison.* We note, as another example, that the Linnean hierarchical classification of plants, and their naming, was a great scientific advance.

Comparison allows us to build a specification of a concept or class. It is an essential feature of developing an ontology[1], within a subject area or amongst a group of interacting individuals or agents.

Mathematics has been defined as 'the science of pattern' or (on one website) as 'the science (or group of sciences) dealing with the logic of quantity and shape and arrangement'. 'Pattern' can mean a lot of different things - but it is clear that determining pattern involves comparison once again. It may be that a pattern is repeated, a fact observed by identifying that the given object is 'the same as' a transformed version of it. 'The same as' may be 'partial' in as much as not all the observable properties are the same, some attributes being themselves transformed. The observation, attributable to Klein, that it was the *allowable* transformations for a geometric context that determined the type of pattern being studied is just one example of this. If you are comparing triangles in the plane, you may be interested in the lengths of corresponding sides or merely the angles between the sides. Of course, the use of 'corresponding' again implies comparison. The beauty of geometry is partially finding that attributes are linked: if the angles of two triangles are the same, then the corresponding sides are linked

[1] We are using 'ontology' here in the sense often used in Artificial Intelligence, as precisely a 'specification of a conceptualisation', [8, 9]. In [5], 'Ontology is the theory of objects and their ties'.

by a more subtle relationship than 'equality', namely 'common ratio'. ('Equality' itself implies a use of comparison, as does 'ratio' and most other simple binary relations[2].) In the first of the two cases the transformations must preserve the attribute of distance between points, in the second they do not, so we get more 'allowable transformations' and a different geometry.

The second 'definition' we gave of Mathematics talks of quantity, shape and arrangement. 'Quantity' clearly involves comparison; 'shape' is closely related to 'pattern', whilst 'arrangement' again involves comparison in various ways as a glance at any book on the theory of graphs and its application to combinatorial problems will show. We thus are left to conclude that 'comparison' is an important aspect of Mathematics, of Science, and, in fact, of general knowledge acquisition.

What about analogy? This is the 'flip side' of comparison and is essential in the 'inductive' side of knowledge acquisition. Suppose we are looking at two objects, or concepts, A and B, say, and have some partial specifications of them. (By a *specification* in this sense we will mean a list of attributes with, perhaps, some known linkages between the attributes as in the geometric case (angles and lengths) mentioned earlier.) Suppose further that some of the specified attributes of A are the same as those of B. To understand the logical relationships between (sets of) attributes, it is natural to test if others of A's attributes are also valid for B. (We may not yet have knowledge as to whether B satisfies some particular attribute or not.) This is nearly a 'that reminds me of' situation. The partial matching, via a comparison, of the properties of A and B leads to an analogy, a test, experiment or an attempt at a proof and perhaps an extension of the comparison, or perhaps the beginning of an abstraction process.

To illustrate this, we refer to a well known mathematical analogy between ways of combining knots and ways of combining numbers. Given a knot K tied in a piece of rope, we can tie another one, L, say, to the right of it, to get a new knot $K + L$, (note the notation suggesting an analogy to addition). If the second knot was the unknot, (so could be manipulated back to being completely unknotted), then $K + L$ could be reduced by similar manipulations back to K itself (without letting go of the ends)! The unknot is behaving rather like the number 0 in addition of numbers. This sum operation on knots is behaving like addition of numbers. It has some of the same attributes: it is a binary operation (it takes two 'things' and produces a new 'thing'). There is a zero 'thing' in both contexts, (the unknot and zero respectively). There seems some analogy between the two situations. This analogy suggests further comparisons of the two situations. For instance with numbers $m + n = n + m$ always; is it always true that for knots $K+L = L+K$? Yes, but you have to take care what $=$ means! (As we mentioned earlier, this is a common occurrence and we will return to it later.) What about another property: the existence of negatives or additive cancellation: if n is a number, there is another number m such that $n + m = m + n = 0$. Here the

[2]The role of 'equality', which is fundamental to 'comparison' will be looked at later on when higher dimensional category theory is examined. It is often the case that 'equality' is not quite what it seems to be naively.

corresponding attribute fails for knots. The basis for the analogy was a partial matching between the attributes of the two objects or concepts. The analogy failed to extend, revealing a new observation on ways of combining knots.

We can attempt to adjust the analogy. Whole numbers under multiplication do not allow division. You cannot solve the equation $2x = 1$ with positive whole numbers. This might give a better analogy, so now we think of $K + L$ as being analogous to $m.n$ rather than to $m + n$. Now 1 corresponds to the unknot and our comparison looks better. In fact we have that if $K + L = 0$, the unknot, then $K = L = 0$, which is analogous to: if $m.n = 1$ for positive whole numbers then $m = n = 1$. Better still we might try some more complicated property of numbers such as 'every whole number other than 1 can be written as a product of prime numbers'. The analogy works. There are knots that are 'prime' or 'irreducible': K is prime if given any equation $L + M = K$ then either L or M must be the unknot. Every knot has a decomposition as a sum of prime knots[3]. Moreover, the proof can be written in an abstract form that works for both numbers and knots and much more generally. Stripping away the context specific details gives a proof that shows what is 'really' happening. This example is not particularly categorical, but it is typical of the type of insight into structure that often occurs with a category theoretic approach based on comparison and analogy.

The important point is that we are not saying that knots and numbers are somehow comparable, but that what *is* comparable are the *relations* between knots and the *relations* between numbers. This situation arises widely in category theory, in comparing categories of particular structures.

This hurried and superficial look at the place of comparison and analogy within Science and Mathematics thus suggests that comparison, in many different guises, is essential for determining the basic classifications of objects, whilst analogy allows the exploitation of partial matching of patterns, obtained by comparisons, to suggest new questions and possible logical linkages between 'contexts' or between objects. In fact, so essential is it, that the previous sentence is a bit strange since it is difficult to imagine any classification without comparison! Analogy and comparison also play key roles in the formation of new concepts and in the process of abstraction and unification, as we will see later.

And what about categories?

For the convenience of the reader, and to set up our notational conventions, we will recall the definition of a category.

A category, $\mathcal{C}$, consists of a family of objects $Ob(\mathcal{C})$ and for each pair of objects, A, B in $\mathcal{C}$, a set $\mathcal{C}(A, B)$ of 'arrows' from A to B, together with a way of 'composing' arrows that match

$$\circ : \mathcal{C}(A, B) \times \mathcal{C}(B, C) \to \mathcal{C}(A, C).$$

[3]For a gentle introduction to the arithmetic of knots, see [3], either in its form as a brochure or in the web-based / CDRom version. A more detailed treatment can be found in several books on Knot Theory including [7]

(If $f \in \mathcal{C}(A, B)$ and $g \in \mathcal{C}(B, C)$, we write $g \circ f$ or simply gf for the 'composite' arrow which is in $\mathcal{C}(A, C)$.)

This data is to satisfy:
(i) composition is associative, so if $h \circ (g \circ f)$ is defined, $(h \circ g) \circ f$ is as well (and conversely) and they are equal;
(ii) there are identities for each object, so there are 'identity' arrows I_A in $\mathcal{C}(A, A)$ for A in $Ob(\mathcal{C})$, such that if $f \in \mathcal{C}(A, B)$.

$$I_B \circ f = f = f \circ I_A.$$

(We will sometimes write $f : A \to B$ if $f \in \mathcal{C}(A, B)$.)

We can view a category as giving a fairly general abstract context for comparison. The objects of study are the objects of the category. Two objects, A and B, can be compared if the set $\mathcal{C}(A, B)$ is non-empty and the various arrows $A \to B$ are 'ways of comparing them'. The composition corresponds to: *if we can compare A with B and B with C, we should be able to compare A with C.* Comparison is also seen as being directional. We may be able to compare A with B, but there need not be a way of comparing B with A. The composition constrains our use of 'compares', but it is still remarkably useful.

Various comments need to be made here. We first note that we have used 'arrow' rather than 'morphism' which is often used in this context. The reason is that 'arrow' has, intuitively, attributes such as 'direction', 'start' (or 'source') and 'finish' (or target'), but 'morphism', as it includes 'morph' has suggestions of 'shape', or 'structure' being preserved as in the more classical algebraic forms 'homomorphism', 'isomorphism', etc. It is usual, when discussing elementary category theory, to start by giving examples of categories that consist of 'structured sets' and 'structure preserving' functions between them. Such categories form an important source of examples on which to test and illustrate notions of category theory, and in such examples 'morphism' seems a highly appropriate term. But in introducing category theory to graduate students, these 'sets with structure' categories can raise a problem if over used as examples. There sometimes grows up in the user the false idea that objects in categories are *always* sets with structure, so they try to force all categories to be of that form. Here we adopt a viewpoint that is nearer to our 'comparison-analogy' themes and which is distantly linked to the Klein *Erlangen Programme* view of geometry. From this viewpoint, in a category $\mathcal{C}$, the 'structure' of an object A in the category only has meaning in relationship to the other objects and hence by reason of the various sets, $\mathcal{C}(A, B)$ and $\mathcal{C}(B, A)$, and their interrelationships via composition[4]. Because of this, 'arrow' is better for us here than 'morphism'.

As a slightly silly example, we take a category $\mathcal{C}$ consisting of, as objects, (mathematical) groups and, as arrows, functions between the corresponding underlying sets. These arrows do not preserve any of the ordinarily defined algebraic

[4]The famous Yoneda lemma and the theory of functor categories allows one to formalise this very neatly and also show more exactly the extent to which the 'sets with structure' view is a valid one.

structure of the objects. The result is that, loosely speaking, using the category $\mathcal{C}$, no one can distinguish two groups if they have the same cardinality. Thus the 'structure' of an object of $\mathcal{C}$ is really just that of a set. This raises the philosophical issue of whether two copies of the same object 'are the same'. For instance, is a group G, considered as an object of this $\mathcal{C}$, 'the same as' G considered as an object of the category, Grps, of groups and homomorphisms? Although the example is a contrived one and is consequently slightly silly, this is not just an 'academic' question as it relates to questions of inheritance of attributes among data types in Symbolic Computation. Mathematicians often abuse notation as a useful tool to link related contexts. As an example, the infinite cyclic group, C_∞, consists of all (formal) powers a^n, both positive and negative, of some single generator a, say. Here C_∞ is a group under multiplication. The ring of positive integers, denoted $\mathbb{Z}$ (for *zahlen*) consists of integers considered with *addition and multiplication*. If we forget the multiplication on $\mathbb{Z}$, we get the integers under addition (as the only binary operation specified) and they form a group isomorphic to C_∞, essentially by pairing each integer n with the corresponding power, a^n of a. Is $\mathbb{Z}$ 'the same as' C_∞? Probably not: it has different attributes; but confusing the two is innocent of dire consequences most of the time and is a useful recognition that $(\mathbb{Z}, +)$ and $(C_\infty, \cdot)$ are isomorphic. Of course, it is sometimes important to make the distinction between $\mathbb{Z}$ considered as a ring and considered as an additive group. From a category theoretic view there are two categories: that, Rngs, of rings and homomorphisms (preserving both addition and multiplication); and that, Grps, of groups and their homomorphisms (which preserve the composition). There is a functor, that is a way of comparing categories, from Rngs to Grps obtained by forgetting the multiplicative structure and considering the operation of addition as being the composition in a group. This comparison and its properties tells one a lot about the relationships between these algebraic structures. Each category serves as a structural context and gives meaning to a specific idea of structure relative to that context.

Another less contrived example comes from observing that any partially ordered set $(P, \leqslant)$ gives us a category, $\mathcal{P}$, with $Ob(\mathcal{P}) = P$, the set of elements of P, and for $x, y \in P$, $\mathcal{P}(x, y)$ is empty unless $x \leqslant y$, in which case it is a singleton set. To take a specific example, let $(P, \leqslant)$ be the partially ordered set, which we will call Div_{36}, of divisors of 36, with $m \leqslant n$ if and only if m divides n. Then $\mathcal{P}$ has objects 1, 2, 3, 4, 6, 9, 12, 18, and 36, and, for instance, $\mathcal{P}(2, 12)$ is a singleton, whilst $\mathcal{P}(12, 2)$ is empty. The composition is the only one possible and the identity, I_m, is the single element of $\mathcal{P}(m, m)$. Note that the objects do not look like sets with structure nor do the arrows look like functions, however they do provide a relevant means of comparison between the objects of $\mathcal{P}$.

In this example what is the structure of, for instance, the object 12. Clearly $\mathcal{P}(m, 12)$ is non-empty for $m =$ 1, 2, 3, 4, 6, and 12 only, and $\mathcal{P}(12, m)$ is non-empty for $m =$12 and 36 only, so $\mathcal{P}$ does allow one to identify the divisor structure of 12 and the multiples of it as well. Category theory looks at this in various ways depending on the aim of the investigation. With the small amount of theory we have introduced, the obvious point to make is that the information in the different sets $\mathcal{P}(m, n)$ is interrelated via the compositions. The various $\mathcal{P}(m, 12)$ with the various links between them give the partially ordered set / category Div_{12} of divisors of 12, whilst the $\mathcal{P}(12, m)$ give Div_3 as $3 = 36/12$. We actually have here an example of a construction of a slice or comma category, and again we will revisit it later.

Although the objects of $\mathcal{P}$ were just numbers, we might ask if we could still defend a 'objects are sets with structure' viewpoint in this situation? For instance, we might find some weird way of looking at $\mathcal{P}$ so that each object gave us a structured set and each arrow a 'morphism' and the result would faithfully reflect the structure of each object. To examine this point, look at the top element of the partially ordered set $P = \mathsf{Div}_{36}$, i.e., 36. This is the terminal object in the category $\mathcal{P}$. In general, an object T in a category $\mathcal{C}$ is *terminal* if each $\mathcal{C}(A, T)$ has exactly one object. (Be warned: many categories will not have terminal objects.) In the category, Sets, of sets and functions between them, any singleton set T is terminal and then for a set X, $\mathsf{Sets}(T, X)$ is essentially the same as X itself: so one possible way to attempt to extract some 'underlying' set of an object X in an arbitrary category $\mathcal{C}$ (if it happens to have a terminal object T) is to look at $\mathcal{C}(T, X)$. (Analogy at work!) Trying to apply this in our category $\mathcal{P}$, we would get for an object m, the set $\mathcal{P}(36, m)$, but unless m is 36, this is empty, so has not much structure! (We are not trying here to be conclusive, just to make the point that the obvious approach does not give one a set with structure.) The point is that we need all the $\mathcal{P}(n, m)$, and thus a family of sets to make things work.

This example shows other features that sometimes come as a surprise to the debutant. It gives a category $\mathcal{P}$ for which $\mathcal{P}(m, n)$ may be empty even when $\mathcal{P}(n, m)$ is not. Clearly these arrows can not be 'invertible'. (The order relation is not symmetric!) The idea of invertibility can be used to illustrate the extent that observations about mathematical objects can often be 'internalised' within category theory. A group homomorphism $\theta : G \to H$ is an *isomorphism* if it has an 'inverse'. This means, more explicitly, that there is a homomorphism $\phi : H \to G$, so that the two composite morphisms $\phi\theta$ and $\theta\phi$ are the respective identity homomorphisms. A group homomorphism is an arrow in the category of groups and homomorphisms, so why not abstract this to get versions of 'isomorphism' in other categories? An arrow $x \in \mathcal{C}(A, B)$ is *invertible* (or *is an isomorphism*) if there is a $y \in \mathcal{C}(B, A)$ such that $x \circ y = I_B$ and $y \circ x = I_A$. Note we do not know anything about the category concerned, just its name! Yet the notion makes sense abstractly given such a 'context'.

In some categories, all the arrows are invertible. Such categories are called *groupoids*. In any category $\mathcal{C}$ and for any object A in it, the self isomorphisms of A form a group in the usual sense of algebra. Of course, group theory grew out of the work of Galois, Lie, Klein and many others. A group of transformations of an object was a natural way of examining that object's structure, of comparing it with itself. Thus in a category each object has a group of transformations, a local algebraic measure of its symmetry structure. 'Local' because it relates only to that one object. Can all groups be realised in this way? Yes, and the reason is really sneaky! (Here again, the debutant finds the construction hard to digest, yet it is fundamental to how parts of category theory is evolving, see John Baez's numerous writings on 'categorification', for instance, [1].) Here is how it goes.

You have a group G and form up a category, $\mathcal{G}$. This category has exactly one object, which will be denoted $*$. We next need to specify $\mathcal{G}(*, *)$ as this will be the only set of arrows around. We set this equal to the set of elements of G. Finally we take composition $\circ$ to be multiplication within G and I_* to be the identity element of G. The single object $*$ is just an object. It is not some element of the group and is there solely as a means of having the elements of G as arrows from it to itself. Of course, within the category $\mathcal{G}$, the group of transformations of $*$ is, guess what, G! This leads one to say that a group is a one object groupoid and a groupoid is a many object group. Group theory is by no way diminished by this, in fact several classical proofs on group theory benefit from combining certain aspects of a groupoid approach, which spreads things out combinatorially making for a neater, shorter proof. What, then, is a one object category? It corresponds to a monoid, that is a set with a single associative binary operation which has an identity, but no inverse is demanded within the structure. Both categories and groupoids combine geometry and algebra, the locational and the local! They have a graph theoretic aspect but also a monoid or group theoretic one.

Examples of groupoids include the fundamental groupoid, ΠX, of a topological space, X. In this the objects are the points of X and if x, y are two such points, $\Pi X(x, y)$ is the set of homotopy[5] classes of paths starting at x and ending at y. The detailed construction can be found, for instance, in Brown, [2]. Exploring that example would give meaning to the 'local structure and locational information' theme, but in fact there is another example that can be used and which is more immediately central to Category Theory as such, namely the free category on a graph and the related free groupoid.

By a graph, Γ, (or more strictly speaking directed graph) we mean what is sometimes called a network: it has vertices and (directed) edges between some of them. (We will label the edges with lower case letters such as $a, b, \ldots$, with suffices if needed.) Now we form a category $FreeCat(\Gamma)$ having the vertices of Γ as its objects and between two such, v and w, $FreeCat(\Gamma)(v, w)$ is the set of paths in Γ starting at v and ending at w. A path from v to w in this context can be represented by a sequence of edge labels, $(a_1, a_2, \ldots, a_n)$ such that a_1 starts at v, a_n ends at w and each a_{i-1} ends at the start of the next edge, a_i. At a vertex v, there is also the trivial empty path that starts and ends at v. Composition in $FreeCat(\Gamma)$ is by concatenation of paths, so the trivial paths act as the identities. A similar construction can be made to obtain the free groupoid

[5]Two paths are homotopic if one can be continuously deformed into the other without moving the endpoints. This construction originally due to Poincaré, related, historically, to an analysis of a space for the purposes of integration along paths. This has lead very recently to attempts to extend the construction to handle situations relating to multiple integrals and integrals over higher dimensional spaces, and thus to 'multiple groupoids' and a higher dimensional algebra.

on Γ; we just add in, for each edge a from v to w, a reverse edge labelled a^{-1} from w to v and when forming path sequences we 'reduce' them by removing any adjacent pairs of the form a, a^{-1} or a^{-1}, a from the sequence. The result will be a groupoid, $FreeGpd(\Gamma)$. Both the free category and the free groupoid contain local information at each vertex/object and the locational information as to how to get between the various objects. They combine an algebraic compositional structure with a geometric or combinatorial one, the best of both worlds!

In passing we will mention that any equivalence relation gives a groupoid, so groupoids generalise both groups and equivalence relations.

Comparing categories

Yes, it had to come! Categories are mathematical structured objects, so we can strive to compare different categories.

The notion of 'morphism' between categories is 'the functor' as was already mentioned above. It assigns objects to objects and arrows to arrows, respecting composition and identities. We have already seen an example of a functor. In our partially ordered set, $\mathcal{P}$, we had $\mathcal{P}(m, 12)$ for varying m. Now given any category we can form another with the same objects but with the opposite direction on all arrows. This is called the *opposite* or *dual category*. For our partially ordered set example, $\mathcal{P}^{op}$ is the category corresponding to the partially ordered set with the reverse order so, in this, $m \leqslant n$ means n divides m, not the other way around. The various $\mathcal{P}(m, 12)$ define a functor

$$\mathcal{P}(-, 12) : \mathcal{P}^{op} \to \mathsf{Sets}.$$

The notation is designed to be self explanatory: the functor $\mathcal{P}(-, 12)$ assigns $\mathcal{P}(m, 12)$ to m. We leave the reader to work out why if $m \leqslant n$, the natural corresponding function is in the other direction from $\mathcal{P}(n, 12)$ to $\mathcal{P}(m, 12)$, and not the other way around.

This is a general construction; if $\mathcal{C}$ is a category and A any object, $\mathcal{C}(-, A) : \mathcal{C}^{op} \to \mathsf{Sets}$ is defined in the same sort of way. By this means we can embed any (small) category into a functor category (modulo a bit of difficulty with set theory and the size of the collections of objects, see the footnote below). In fact this suggests that any category can be realised as a category of families of sets with structure and structure preserving families of functions between (Yoneda's lemma), so although we warned against the simplistic view of categories as always consisting of 'sets with structure' and the corresponding structure preserving 'morphisms', it was not that far from the truth. It is just not a useful view of categories to have to the exclusion of others!

Another functor can be constructed from the free path examples. This goes from a category of graphs to the category of small[6] categories or small groupoids. Functors often have a lot of structural properties, for instance, these free category

[6]Small categories are ones in which the collection of objects is a set in the sense of set theory. This smallness condition is often avoided by using various tricks. It is not that annoying in practice, but at the same time is occasionally vital!

and free groupoid functors have a freeness property, that we will not explore here. One of the points of category theory is to seek common attributes for various classes of functor. That again suggests comparison, this time of functors. This is done by *natural transformations*, which we will not discuss further but will refer to standard texts on category theory.

Comparison and Analogy lead to abstraction and thus to fresh concepts, but also to unified treatments of existing concepts. For instance, in geometric contexts, the notion of a group action is often important. To specify an action of a group, G on a set X, we give for each $g \in G$, a permutation σ_g of X, so that for $g, h \in G$, $\sigma_g \sigma_h = \sigma_{gh}$. We say that X together with the action of G is a G-set. Of course there is a notion of morphism of G-sets and the G-sets and the corresponding morphisms together form a category, G–Sets.

If X has more structure than a mere set, for instance if it is a topological space or a vector space, then in the above definition we would replace 'permutation' by the appropriate notion of isomorphism in the category, Spaces or Vect, in which X is being considered. There will then be a corresponding category of G–Spaces or G–Vect. If we had any category $\mathcal{C}$, we could define a notion of G-action on an object of $\mathcal{C}$ and would get a corresponding category G–$\mathcal{C}$. (Again here we have abstraction of concepts 'by analogy', and the categorical language helps us in this process.)

The first unification is to note that a G-set corresponds to a functor

$$\mathbf{X} : \mathcal{G} \to \mathsf{Sets}.$$

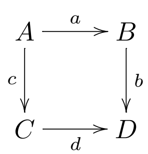
Such a functor will be specified by assigning a set to each object of $\mathcal{G}$, the associated category of the group, G, and so, as there is only one object, $*$, in $\mathcal{G}$, it picks out a single set $X = \mathbf{X}(*)$, that is the set on which G will act, but where are the permutations? For each $g \in G$, there is an arrow $g : * \to *$ in $\mathcal{G}$. Applying $\mathbf{X}$ gives a function $\mathbf{X}(g) : \mathbf{X}(*) \to \mathbf{X}(*)$. Relabelling, that is a permutation of X, i.e., the σ_g is this $\mathbf{X}(g)$. The details are easy to check. We could, if we had given the details, have checked that a G-morphism of G-sets corresponds exactly to a natural transformation of the corresponding functors. This category G–Sets is thus a functor category in which guise it would be denoted $\mathsf{Sets}^{\mathcal{G}}$ or similar. We could equally well define a G-object in a category $\mathcal{C}$ as being a functor from $\mathcal{G}$ to $\mathcal{C}$.

This does not stop there. General categorical constructions (products, coproducts, limits and colimits) can be defined in a category $\mathcal{C}$ and then a single unified proof can be given to say when they exist in the corresponding category G–$\mathcal{C}$ of G-objects in $\mathcal{C}$ using this unification of treatment of the concepts. Category theory gives good abstractions, often 'the right' abstractions[7]. The advantage of

[7] Some mathematicians object quite rightly to saying 'the right' abstraction, however the most natural constructions within a subject area are often of a categorical nature. The problem then is not with the category theorists being prescriptive as to what is 'right', rather in the success rate of categorically based concepts.

unified treatments is that they show why things are happening. Subject specific proofs do not always show this and so may not be so widely applicable.

The above discussion was in part based on an observation about categories of the form $\mathcal{C}^{\mathcal{G}}$ or similar. Structure from $\mathcal{C}$ in some way extends or lifts to structure on $\mathcal{C}^{\mathcal{G}}$. Another example of a very similar nature is that of commutative squares in a category $\mathcal{C}$. A square $\mathbf{X}$, being a diagram



of objects and arrows in a category $\mathcal{C}$, is *commutative* if $b \circ a = d \circ c$. This is, of course, just one shape of commutative diagram, but will illustrate our point. We can form a category, $\square$, having four objects, labelled, 1, 2, 3, and 4, say, and arrows $a_{12} : 1 \to 2$, $a_{13} : 1 \to 3$, $a_{24} : 2 \to 4$, $a_{34} : 3 \to 4$, and $a_{14} : 1 \to 4$ plus identities on each object. Composition is given by $a_{24}a_{12} = a_{14} = a_{34}a_{13}$. Then a commutative square in $\mathcal{C}$ is just a functor from $\square$ to $\mathcal{C}$. We get, for free, a notion of morphism between commutative squares, and a category $\mathcal{C}^{\square}$ of commutative squares in $\mathcal{C}$. Again properties of $\mathcal{C}$ give properties of $\mathcal{C}^{\square}$. The category $\square$ acts as a 'template' for all structures of a particular form.

It is not that far away from this example to get to an abstract notion of algebraic structure (of a given form) in a category $\mathcal{C}$, as being a particular type of functor from a 'signature' category to $\mathcal{C}$. This notion of algebraic structure again allows wide ranging and influential generalisations with applications in topology, geometry and recently in computer science and logic.

Category Theory, by taking analogy seriously, has thus provided a 'structural context for structural mathematics' and a 'meaning to the idea of structure'.

A case study: comma categories.

The way in which comparison leads to structural information can be approached in various ways. We have mentioned functor categories and the Yoneda embedding above, and also comma categories. These latter constructions give a clear illustration of how comparison with other objects does give useful information. We will only give the details of one example.

We earlier met the construction of a free groupoid on a graph. A particular case of that construction gives that of the free group, $F(X)$, on a set, X. (We leave the reader to make the connection exact by building a graph from the set X in such a way that the free groupoid on that graph has a single object, hence is a group. The change from a group to a groupoid is another aspect of 'categorification' à là Baez, [1].) The idea behind this 'free' construction is that one builds a group from X by adding in new elements to represent the result of 'multiplying' or inverting elements of X, doing this in such a way that no equations are satisfied in the result except those given by the axioms for a group.

Given a group G, the elements of G usually satisfy equations in addition to those required by the fact that G is a group. For instance, it may be that for any $x, y \in G$, $xy = yx$. This would not be true in an arbitrary group and so gives important structural information about G. That was an example of an equation assumed to hold for all elements of the group, but in, say, the symmetry group of the triangle, the rotation through 120 degrees is such that its cube is the identity element, again a piece of structural information but this time about a specific element of the group.

Our aim is to see if we can find some of that structural information about a group G by comparing free groups with it. By this we mean that we will look at group homomorphisms $F(X) \to G$, and by varying X get some information. (This is a bit like studying 12 in Div_{36} by looking at all $\mathcal{P}(m, 12)$.) We first note that this free group construction gives a functor $F : \mathsf{Sets} \to \mathsf{Grps}$. We want to look at 'approximations' to G by free groups. We could do this by examining the composite functor $\mathsf{Grps}(F(-), G) : \mathsf{Sets}^{op} \to \mathsf{Sets}$, but we will approach this from a slightly different direction using a construction that owes some of its structure to a *geometric intuition* about categories rather than an algebraic one. This construction is a particular case of the general comma category[8] construction. We form a category, $(F \downarrow G)$, with the approximations $f : F(X) \to G$ as objects, so the objects are arrows in another category, in this case that of groups and homomorphisms. If $f_1 : F(X_1) \to G$ and $f_2 : F(X_2) \to G$ are two such objects, then an arrow from f_1 to f_2 is an arrow $a : X_1 \to X_2$ in Sets, such that $f_2 F(a) = f_1$. This seems a bit complicated the first time it is encountered, but makes a lot of sense. The objects compare the free groups on various sets with G and a comparison between such varies the sets being used by applying a function.

The beautiful thing about this situation is that there is a very special object in $(F \downarrow G)$ as it has a terminal object. This means that there is a set X and a homomorphism $\varepsilon_G : F(X) \to G$ such that given any set Y and any homomorphism $f : F(Y) \to G$, there is a unique function $a : Y \to X$ such that $\varepsilon_G F(a) = f$.

There is no real mystery here. The set X is a copy of the set $U(G)$ of elements of G, so is obtained by 'forgetting' the multiplication structure of G, and ε_G is the homomorphism that takes a string of elements of G, hence an element in $FU(G)$, and composes them to get a single element. It is thus clear that ε_G is really the multiplication in disguise. With a bit of care this construction can be read off just from the fact that there is a terminal object in $(F \downarrow G)$. Conversely, we can obtain a from f explicitly: the free group $F(Y)$ contains the singleton strings each consisting of a single element of Y. If $\{y\}$ is such a string, then $f(\{y\})$ is an element of G. The function $a : Y \to U(G)$ sends y to $f(\{y\})$.

This is a recurring type of situation in algebraic contexts (and wider). There are 'forgetful functors' from some category to another, obtained by forgetting some of the structure of the objects. In the above example $U : Grps \to Sets$

[8] The terminology comes from the use of a comma in the notation used by Lawvere in early applications of the idea.

forgot the multiplicative structure. The above analysis set up a one-to-one correspondence between the arrows $F(Y) \to G$ in *Grps* and those $Y \to U(G)$. (In fact there is a second comma category $(Y \downarrow U)$ and this has an *initial object*, i.e., the dual situation!) The key concept here is that of adjoint functors: F is *left adjoint* to U. This in turn gives rise to an abstract approach to algebraic structure which has proved very important in analysing structure in Computer Science as well as in Algebra itself.

To return to our question of the structure of individual elements of a group, the special arrow $\varepsilon_G : FU(G) \to G$ compares G and the free group on G. If, for instance, a particular element, g, of G satisfies $g^{59} = 1$, then the string consisting of the element g repeated 59 times gives an element in $FU(G)$ that is sent to 1_G by ε_G, hence is in $\ker \varepsilon_G$. Thus the internal structure of G can be investigated by examining the terminal object. That investigation would take us away from category theory as such, as it usually will involve context specific methods. An analogous situation does, however, keep us in context of category theory and will provide a bridge to our next discussion.

Instead of looking at the categories of groups and sets and the free-forget pair between them, we could have substituted small categories and directed graphs. There is a forgetful functor from the category of (small) categories to that of directed graphs, which forgets the composition. (Categories are directed graphs, a combinatorial or geometric concept, which in addition have a composition, an algebraic notion.) The free category construction we looked at earlier gives a left adjoint to that forgetful functor. In the free category on a graph, there are no non-trivial commutative diagrams, not unsurprisingly as the commutativity of a diagram corresponds to an 'equation' in the arrows. We could apply the ideas we developed for groups to this context. If we do have, say, a commutative square, in a (small) category $\mathcal{C}$, then we can find two elements in $FU(\mathcal{C})$, which are representing the two paths around the square, so will be sent by $\varepsilon_{\mathcal{C}}$ to the same arrow in $\mathcal{C}$. We can think of these two paths as being 'equivalent'.

A question of identities.

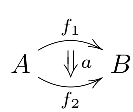
Earlier we asked about the meaning of 'the same as'. This was initially in the context of the Knots example. If a knot L 'was the unknot', then we could manipulate it to unknot it! In other words we had to produce some set of moves that would result in it becoming clearly unknotted. This sort of situation has also occurred, without comment, in other situations we have looked at. When forming the composition within the free groupoid on a graph we had to reduce the result of concatenation by cancelling cancellable pairs.

Such changes to a formula or expression are a standard technique, yet the language for this is not always clear. Thus we say $2+3 = 5$, but this can and even does confuse. The left hand side and the right hand sides of the 'equals' sign are not the same! The left hand side can be considered as an instruction, 'add 2 to 3', and the right hand side as giving the answer. Thus for this rewriting we could better say '$2 + 3 \to 5$ '[9] and think of it as an operation or process. There is a reverse process $5 \to 2+3$, but it is of a different character. (It is cutting up rather than putting together; it goes from one thing to two; it is a co-operation, which can be made precise within a categorical setting.) This analysis of a statement usually considered very simple is not a case of simple mathematics made difficult. It is essential for computer implementation, since the computer cannot guess and has to be told precisely what type of thing each symbol stands for.

Another example of an 'identity crisis' comes with the $\mathbb{Z}$ and C_∞ example. There is a temptation to say that these are 'the same', but does that mean they are 'equal'? We might identify them via the obvious isomorphism, but why choose that isomorphism and not, for instance, the one pairing n with b^n, where we have written b for a^{-1}, which just picks a different generating element for C_∞. Unless we pay attention to such queries, we are in danger of killing the very structures that we set out to study, in particular the internal symmetry structure of objects. The point is: *isomorphism is not identity.* In fact it is often useful to replace 'equal' by 'equivalent' and, of course, this is really what we have done on several occasions above. To prove two things are 'equal' is often impossible, partially because 'equality' is such a difficult concept. To prove that two things are 'equivalent' requires one to analyse the situation and to find the suitable notion of 'equivalence' and then to provide a proof that the two things *are* equivalent, which is much more 'healthy'.

A good example of this is equality of functions $f, g : X \to Y$. It is very easy to say that $f = g$ if and only if $f(x) = g(x)$ for all $x \in X$. This can be verified, perhaps on a computer, if X is finite and not too large. However, if X is the set of natural numbers, it is not possible to program a computer to decide this equality problem, even if f and g are always to be given by elementary formulae in arithmetic. This is the famous undecidability result. What we can ask is how f and g are specified as functions, and what are the ways of moving from one specification to another. This movement, hopefully reversible, is a kind of *equivalence.*

When, as increasingly is seen to happen, equivalences do need to be accounted for in detail, then category theory *per se* is not really adequate and higher dimensional categories are needed. In these, in addition to objects with (1-)arrows between them, there are 2-arrows between the 1-arrows (so $\mathcal{C}(A, B)$ is no longer just a set, but is a category). The picture



suggests two arrows from A to B with $a : f_1 \Longrightarrow f_2$ being a 2-arrow, that is, a way of comparing the two comparisons! Even that does not seem to be adequate

[9] The use of the arrow notation in category theory has often proved very suggestive for applications. Here we used it to indicate a 'process' with a definite direction.

for some situations and it has proved necessary to investigate 3-arrows between 2-arrows, etc. This theory is still in full development and new ideas are coming to the surface from various other areas of mathematics, computer science and physics. It is motivated and strongly influenced by difficult problems in algebraic topology, algebraic geometry and, as mentioned above, mathematical physics, so is not 'abstract nonsense' in any way. The sense of unification of large parts of modern mathematics comes from comparison and analogy as well as from abstraction. It is because of this that category theory has come to play such a fundamental role in much of this development.

Categories as mixed algebraic / geometric structures for local-to-global problems.

We have not said that much about the theme 'local structure with locational information'. Its importance is best exemplified by work in geometric contexts and we will briefly look at two geometers and how they interpreted this. They both have had a lasting influence on the development of category theory, whilst not being themselves category theorists.

What category theory first did in Eilenberg and Mac Lane's seminal 1945 paper was to give a simple algebraic definition for a general notion in mathematics, and to show how it applied in a number of situations, in particular, allowing a formal expression for the term 'natural transformation'. In fact similar algebraic axioms had appeared earlier in Brandt's 1926 paper giving the notion of groupoid. This notion was quite familiar to the algebraists at Chicago in the 1940s, since it is used in Albert's book on the structure of algebras. When one of us asked Eilenberg about this in 1985, he insisted that the notion of groupoid did not influence them, since had they had been aware of it they would have included it as an example! The Eilenberg-Mac Lane approach centred on categories as ways of managing large collections of mathematical objects, but the Brandt tradition considered it more as an algebraic object encoding some sort of symmetry.

Another early example of this can be seen in the work of Charles Ehresmann, [6], and in particular with his notion of structured category, so that he used topological and differentiable categories (and groupoids), and also introduced double and higher categories[10]. In Differential Geometry, a natural object to study is the tangent bundle of a smooth space, that is, the set of all tangent vectors at all points of the space. One of Ehresmann's ideas was that comparison of the points of the space could be done by looking at local transformations between the corresponding tangent spaces. The resulting groupoid would have extra structure coming from the differential structure on the space and the linear structure on each tangent space. The local behaviour of a geometric structure at a point or between two points in a geometric space, clearly is reminiscent of the structure of a groupoid in which at each object (point) one has a group of local symmetries. In the case mentioned above, the groupoid itself was 'smooth'.

[10]Without being too specific, we refer the reader to his Collected Works, [6], particularly Volume III-2.

The overall structure gave one interpretation of the *global* structure of the geometry. Investigating this led him to think about categories with various types of extra structure. His work was found to be difficult to follow, mostly because his viewpoint was that of a geometer not of the algebraist as at that time, algebra was the main user of categorical language, yet many of his constructions have profound implications in Algebra.

One of the reasons for the difference in style between Ehresmann's work and that of other early writers on category theory is that he was a student of Elie Cartan, in analysis and geometry. A great theme in his work is the relations between *local and global* properties, and the use of notions of categories and groupoids to express this theme. We have stressed earlier that the objects in a category have structure due to the morphisms between them and this connects that view into the much earlier ideas of Klein. Ehresmann tried to emphasis this view by naming his categories after the morphisms in them rather than the objects. The usual category of sets would thus be called the category of functions, and this allowed the category of sets with relations as morphisms to be called the category of relations. It was thus the allowable transformations in many object contexts that Ehresmann stressed.

The other major geometric contributor to categorical ideas has been Alexander Grothendieck. It is notable that his striking initial work was in functional analysis, and he could be said to have carried over the local-to-global themes there into the arena of number theory, where 'local' meant 'at a given prime'. He also developed vast tools of category theory to express these local-to-global problems. Grothendieck realised the power of categorical methods to express geometric analogies. He was able to adapt results and ideas from other areas and for instance, made clear the links between the categories of G-sets and various very geometrically defined categories. In fact, he saw this unification as being a generalisation of the Galois theory of fields in algebraic number theory. These links were most clearly described by a categorical equivalence. The question of generalisations of this viewpoint to higher 'dimensions' is still a very active area of research. He also developed a categorical notion of space[11] that has provided one of the ingredients in the recent development of non-commutative geometry.

In all this, categorical notions, developed to describe the way mathematical structures behave and interact, have led to new mathematical structures of interest in their own right. A final but striking example of this is the idea of a monoidal categories. This again is part of the *internalisation*, and *categorification* tendency we have mentioned several times. They are like monoids, but are defined on a category with the multiplication being given by a functor not a function, the equations such as that for associativity, which in a monoid involve equality are here replaced by expressions involving an equivalence, and so on. The amazing thing is that this wonderfully abstract idea gives amazing analogies between

[11] A beautiful description of Grothendieck's contribution to categorical ideas of space can be found in Cartier's article, [4].

the behaviour of rings and of knots, and can also be analysed algebraically in interesting finite or countable cases. It has revolutionised parts of physics and geometry, as well as revealing new structures and analogues in the theory of group representations.

It has earlier been said that monoids are categories with one object, and conversely categories are monoids with many objects. Similarly, groupoids are 'groups with many objects'. This extra 'spatial component' given by the objects has been found to have profound implications, and to give valuable tools for modelling and expressing intuitions, and even for more effective calculation. We can impose numerous different structures on 'space', and likewise with this next level of categories. This again leads to structured categories and groupoids, à la Ehresmann. If these objects have to be structured as a category or groupoid, then we are led to higher categories and also to Grothendieck's higher dimensional analogues of Galois theory. There is currently a burgeoning interest in this field, which, it is suggested by some, may be a major theme of mathematics in the 21st century.

Conclusion

We do not claim to have answered the question, but via our themes of comparison and analogy, and their relation to abstraction and unification, perhaps we have shed some light on 'What is Category Theory?' We have not really touched on the aspects of Category Theory as a language, or on the very strong links between category theory and certain parts of logic, but have wanted, to some small extent, to point out the links between the algebraic and combinatorial or geometric aspects of the subject. Perhaps the beauty of Category Theory is the way in which it has enriched its constructions and theories with intuitions from a very wide family of subject areas and, by comparison and analogy, has abstracted some of the essence from each. This aids the unification of large parts of mathematics and related subjects[12].

References

[1] J. Baez, Categorification, available on the ArXiv: math.QA/9802029.

[2] R. Brown, Topology : A Geometric Account of General Topology, Homotopy Types and the Fundamental Groupoid, Ellis Horwood, Chichester, 1988.

[3] R. Brown, N. D. Gilbert and T. Porter, *Mathematics and Knots* exhibition, Bangor, 1989, enlarged web-version (http://www.popmath.org.uk), also available on *RPA Maths* CD-Rom (version 1.1, by R. Brown, T. Porter and M. Yates).

[12]If a reader doubts the reality of this, come to a category theory conference. The wide range of backgrounds of people contributing to the discussions and the interactions this gives is always very stimulating.

[4] P. Cartier, A mad day's work: from Grothendieck to Connes and Kontsevich, The Evolution of Concepts of Space and Symmetry, Bull. American Mathematical Society, 38, (2001), p. 389-408.

[5] R. Corazzon, Ontology: A resource guide for philosophers, at URL: http://www.formalontology.it

[6] C. Ehresmann, Oeuvres complètes et commentées, available from http://perso.wanadoo.fr/vbm-ehr/CT/CT21.htm

[7] N. D. Gilbert and T. Porter, Knots and Surfaces, Oxford University Press, 1994.

[8] T. R. Gruber, A translation approach to portable ontologies, Knowledge Acquisition, 5 (1993), p. 199-220.

[9] T. R. Gruber, What is an Ontology? at URL http://www-ksl.stanford.edu/kst/what-is an-ontology.html