CHAPTER 9

The importance of mathematical conceptualisation

> A mathematician, like a painter or a poet, is a maker of patterns. If his patterns are more permanent than theirs, it is because they are made with *ideas*. (Hardy 1940: 24)

> All these difficulties are but consequences of our refusal to see that mathematics cannot be defined without acknowledging its most obvious feature: namely, that it is interesting. (Polanyi 1958: 188)

9.1 VALUES IN MATHEMATICS

As with any academic community, mathematicians must devote a significant part of their time to promoting their research activities. This occurs both *externally*, with a view to improving the standing of mathematics relative to other disciplines, and *internally*, with a view to establishing the importance of specific research programmes. What is very noticeable when one encounters such promotion exercises is the enormous variety of qualities alluded to and the differences in emphasis placed on these qualities. Even if, as seems to be the case, a considerable consensus has persistently taken certain moments in the history of mathematics to be pivotal, importance does appear to be a time-dependent notion. Something vitally important for one generation may not seem quite so crucial for the next. But, alongside these rather predictable variations between generations of mathematicians, one also finds considerable dissimilarities between mathematicians of any given era. We would expect, then, that an exploration of contrasting opinions about what constitutes an important advance would reveal much about the evolution of competing images of mathematics and the tensions existing between them.

A first step is to glean what we can from mathematicians who meet with the question of importance in their roles as researchers, teachers, referees, textbook writers, grant body panellists and doctoral supervisors. As I mentioned above, what we quickly discover is that there are many criteria for judging the importance of a development and that there are differences among mathematicians as to the value they place on success according to these various criteria. For the purposes of this chapter, I propose to arrange these criteria, several of which we matched to Lakatos's notions of progressiveness in chapter 8, into five broad categories:

(1) When a development allows new calculations to be performed in an existing problem domain, possibly leading to the solution of old conjectures.
(2) When a development forges a connection between already existing domains, allowing the transfer of results and techniques between them.
(3) When a development provides a new way of organising results within existing domains, leading perhaps to a clarification or even a redrafting of domain boundaries.
(4) When a development opens up the prospect of new conceptually motivated domains.
(5) When a development reasonably directly leads to successful applications outside of mathematics.

Naturally, some developments may be rated highly according to several or perhaps all of the categories. In particular, it may happen that a reformulation of a body of existing theory or a unification of existing theories leads to new results in an already existing domain, but also points the way forward to a new area. Of course, there is a fine line between clarifying the boundaries of an old domain and extending beyond them into a new domain, but there are cases which are clearly on one side or the other.

What needs to be brought under close scrutiny is the tacit weighting given by the mathematical community to these different criteria. My perception is that, very reasonably, if a development is seen either to be doing well or to have the potential to do well according to the majority of the criteria, then interest is guaranteed. Take, for example, the Atiyah–Singer index theorem, a formula demonstrated in the early 1960s which links analytic information concerning an elliptic differential operator with topological information on an associated vector bundle, thereby relating important constructions in algebraic topology to the domain of partial differential equations. This immediately scored extremely highly on (1) and (2), highly on (3), and had the potential to score well on (5). Sure enough, the theorem later found its uses in quantum field theory.

A problem arises, however, when a development appears to do well on one front, but poorly on the others. An example here is the computer assisted proof of the four-colour theorem. Few now deny that the theorem is true, or that the various computer proofs warrant our belief in it, and yet a widespread feeling persists that unless there is some more conceptual success, for example, by linking the theorem to other branches in illuminating ways,[1] then little has been achieved. What interests me more, however, are situations where the conceptualists are in the minority position. We can express their concern as follows: despite the greater difficulty of scoring according to categories (3) and (4), success here is not in general given sufficient weighting. Their worry is, in other words, that the conceptual aspect of mathematical activity is on occasions undervalued in that the acquisition of results is favoured over the reorganisation and elaboration of concepts.[2]

To be what I have called a conceptualist is not indicative of any specific view as to the ultimate goal of mathematics. As a conceptualist, you may well believe that the proper organisation of mathematical ideas is an end in itself, but equally it could be that you see it as the most appropriate way of providing tools to model the natural world, which you view as the essential purpose of mathematics.

Complaints of a lack of conceptual appreciation are not hard to find. For instance, although his work is generally considered to be extremely important, Mikhael Gromov considers that a book of his containing fundamental insights on partial differential equations

> is practically ignored because it is too conceptual. (Berger 2000: 187)

Now, it is no easy business defining what one means by the term *conceptual*. One radical position, represented by a book such as *Conceptual Mathematics* (Lawvere and Schanuel 1997), sees category theory as providing much of the answer. Where set theory picks up on a few of our everyday structural concepts (collection, membership, union, etc.), category theory does so more extensively in such a way that its concepts can be found in a multiplicity of contexts. Without wishing to take sides here, I think we can say that the conceptual is usually expressible in terms of broad principles. A nice example of this comes in the form of harmonic analysis, which is based on the idea, whose scope has been shown by George Mackey (1992) to be immense, that many kinds of entity become easier to handle by decomposing them into components belonging to spaces invariant under specified symmetries. In the case of Gromov's book, on the other hand, the conceptual core is expressed in terms of the *h-principle*, which holds, roughly speaking, that, in many geometric situations, obstructions to the construction of solutions to partial differential equations arise only from topology.

[1] This is being done. See Thomas (1998).

[2] This imbalance is also noted by Laugwitz (1999: 22).

In a fascinating paper, which provides an excellent counterpoint to this one, the mathematician Timothy Gowers (2000a) stands up for the kind of mathematics which earned him a Fields's Medal. To some the field of what he refers to as 'combinatorics' appears as a collection of wholly unrelated problems, each requiring some clever trick to solve it, while fields such as algebraic number theory contain many general unified results. But by contrasting the problem solving to the theory building components of mathematical activity and identifying the field of combinatorics as one where the former prevails, Gowers does not mean to suggest that there is no common ground between the ways of arriving at results in combinatorics. Instead, he notes that the solution of combinatorial problems often leads to the production of 'somewhat vague general statements' (Gowers 2000a: 72) which then open up other problems for solution.[3] It would be interesting to observe the extent to which these implicit general principles can evolve to become explicit unifying theories.

To my mind the most straightforward access we can gain to these issues is via a case study analysis. What we require then is an example of a development whose fate hangs or has hung in the balance. In this respect an account of quantum groups, fascinating though this would be for our understanding of mathematical physics in the late twentieth century, will not fit the bill. Given the centrality of Lie groups and Lie algebras to mathematical physics, the pleasantly surprising discovery that 'quantum' deformations of examples of the latter exist was never going to be seen otherwise than as an important breakthrough. What we need to observe is an idea which, it is claimed by some, has suffered neglect because of a lack of immediate success in the more 'practical' categories, (1) and (5) of those outlined above.

The case I have chosen to treat in this chapter concerns the question as to whether the group concept should be extended to, or even subsumed under, the groupoid concept. Over a period stretching from at least as long ago as the early nineteenth century the group concept has emerged as the standard way to measure the degree of invariance of an object under some collection of transformations.[4] The informal ideas codified by the group axioms, an axiomatisation which even Lakatos (1978b: 36) thought unlikely to be challenged, relate to the composition of reversible processes revealing the symmetry of a mathematical entity. Two early manifestations of groups were

[3] For example: 'if one is trying to maximize the size of some structure under certain constraints, and if the constraints seem to force the extremal examples to be spread about in a uniform sort of way, then choosing an example randomly is likely to give a good answer' (Gowers 2000a: 69).

[4] See Wussing (1984).

as the permutations of the roots of a polynomial, later reinterpreted as the automorphisms of the algebraic number field containing its roots, in Galois theory, and as the structure-preserving automorphisms of a geometric space in the Erlanger Programme. Intriguingly, it now appears that there is a challenger on the scene. In some situations, it is argued, groupoids are better suited to extracting the vital symmetries.[5] And yet there has been a perception among their supporters, who include some very illustrious names, of an unwarranted resistance in some quarters to their use, which is only now beginning to decline.

My claim is that, although groupoids did well at reformulating old domains and pointing to new areas for exploration, they suffered from leading to too little in the way of new techniques for solving old, circumscribed problems. Thus, their early adoption required an inclination towards being conceptually adventurous. However, now that programmes using groupoids have becoming established, researchers can use them to work with more of an air of what we might call 'normal mathematics' within these programmes.

9.2 WHAT IS A GROUPOID?

When promoting a mathematical concept, it is never a bad idea to think up an illustration from everyday life. Ronald Brown (1999: 4), a leading researcher in groupoid theory, has provided us with a good example by considering possible car journeys between cities of the United Kingdom. Now, one approach to capturing the topology of the British road system is to list the journeys one can make beginning and ending in Bangor, the Welsh town where Brown's university is located. This possesses the advantage that the members of the list form a group under the obvious composition of trips, where the act of remaining in Bangor constitutes the group's identity element.[6] However, for a country so dominated by its capital city, it might appear a little strange to privilege Bangor and the act of staying put there. Each city might be thought to deserve equal treatment.

[5] If they succeed, then my account of the deficiencies in Lakatos's philosophy of mathematics discussed in chapter 7 will be supported. Lakatos's belief that 'elementary group theory is scarcely in any danger [of heuristic refutation]' arises from his idea that 'the original informal theories have been so radically replaced by the axiomatic theory' (Lakatos 1978b: 36). But extended informal notions of symmetry which arise from working with axiomatised theories elsewhere in mathematics may provide such a refutation.

[6] Note that trips are being considered here only 'up to homotopy'. In particular, taking a trip and then retracing one's steps is to be equated with staying at home.

Pleasant as it is to remain in Bangor, staying put in London should surely be seen as another identity element. Moreover, if you want to know about trips from London to Birmingham, it would seem perverse to have to sift through the set of round trips from Bangor which pass through London and then Birmingham, even if all you need to know is contained therein. And if ferry journeys are excluded, this method is perfectly hopeless for finding out about trips out of Belfast. More reasonable then to list all trips between any pair of cities, where ordered pairs of trips can be composed if the destination of the first trip matches the starting point of the second. Something group-like remains but with only a *partial* composition. On this basis Brown can claim that:

> [t]his naïve viewpoint gives rise to the heretical suggestion that the natural concept is that of groupoid rather than group. (Brown 1999: 4)

As the mention of heresy indicates, the suggestion is far from universally accepted within the mathematical community. We read that Alain Connes, the French Fields's medallist, considers that 'it is fashionable among mathematicians to *despise* groupoids and to consider that only groups have authentic mathematical status, probably because of the pejorative suffix oid' (1994: 6–7, my emphasis). This explanation of the origins of such a strong sentiment may seem implausible, but there can be little doubt that the climate towards groupoids has not been exactly favourable.

Brown reproduces a passage from a letter sent to him by Grothendieck in 1985:

> The idea of making systematic use of groupoids . . . , however evident as it may look today, is to be seen as a significant conceptual advance, which has spread into the most manifold areas of mathematics . . . In my own work in algebraic geometry, I have made extensive use of groupoids. (Quoted in Brown 1999: 7)

One might have expected that eleven years later the matter would have been settled, the 'evident' idea would have spread, but according to Alan Weinstein, a noted geometer, by 1996 the message had still not got through:

> Mathematicians tend to think of the notion of symmetry as being virtually synonymous with the theory of *groups* . . . In fact, though groups are indeed sufficient to characterize homogeneous structures, there are plenty of objects which exhibit what we clearly recognize as symmetry, but which admit few or no nontrivial automorphisms. It turns out that the symmetry, and hence much of the structure, of such objects can be characterized algebraically if we use *groupoids* and not just groups. (Weinstein 1996: 744)

To counteract resistance to their use we find that three articles have been written and two Internet websites constructed with a view to their promotion. Such explicit promotion is quite unusual, although mathematicians are aware of the need to market their wares. In a humorous subsection of his book, entitled 'Commercial break', the algebraic geometer Miles Reid tells us that:

> Complex curves (= compact Riemann surfaces) appear across a whole spectrum of maths problems, from Diophantine arithmetic through complex function theory and low dimensional topology to differential equations of math physics. So go out and buy a complex curve today. (Reid 1988: 45)

Now, this book is aimed at undergraduates – complex curves have not stood in need of any PR campaign for the purposes of recommending them to professionals for many decades. In fact, it would be hard to count yourself a professional mathematician without agreeing that complex curves are a good thing, even if your research never brings you particularly close to them. Groupoids, on the other hand, despite generating sufficient interest for an annual 'Groupoid Fest' to be held in their honour, still require some salesmanship.

Two of the promotional articles are due to Brown (1987, 1999), the first appearing in the long-established *Bulletin of the London Mathematical Society*, which publishes research and expository articles, while the second forms the opening article of the first issue of a new journal *Homology, Homotopy and Applications*. The other article, Weinstein (1996), appears in the *Notices of the American Mathematical Society*, a more informal journal, which includes, besides less technical exposition, articles on the teaching of mathematics and administrative issues. This informality is reflected by the choice of cover picture for the edition containing Weinstein's article. Next to the title of this article one sees a photograph of a herd of zebra. No explicit explanation is offered for its presence, nor is one needed. The received account as to why zebras sport stripes is that when they stand in a herd, a charging lioness is presented with a strongly patterned visual array, making it very difficult for her to detect the outline of a single member of the herd. The rationale for the choice of this picture, in which one imagines Weinstein played a part, rests in his idea that groupoids are better than groups at detecting the inner symmetry of patterns of this kind. This idea Weinstein explicitly illustrates in the article itself with a discussion of the symmetries of a set of bathroom tiles. In contrast to this rather mundane concern of the mathematician contemplating the pattern of the grouting while enjoying a soak, the cover picture makes clear that such inner symmetry is a matter of life and death. As any zebra will tell you, 'symmetry capturable by groupoids but not by groups saves lives'.

Let us now consider the definition of a groupoid and its motivation.[7] A groupoid is composed of two sets, A and B, two functions, a and b, from B to A, and an associative partial composition, $s \cdot t$, of pairs of elements of B with $a(s) = b(t)$, such that $a(s \cdot t) = a(t)$ and $b(s \cdot t) = b(s)$. Furthermore, there is a function, c, from A to B such that $a(c(x)) = x = b(c(x))$ and such that $c(x) \cdot s = s$ for all s with $b(s) = x$ and $t \cdot c(x) = t$ for all t with $a(t) = x$. Finally, there is a function, i, from B to B such that, for all s, $i(s) \cdot s = c(a(s))$ and $s \cdot i(s) = c(b(s))$.

This may seem like a highly convoluted definition, but it can be illustrated simply in Brown's picture. We simply take A to be the set of cities, while B is the set of trips. The start and finish of a trip are given by applying a and b, respectively. Applying c to a city results in the staying-put trip. Finally, i sends a trip to the same trip in reverse.[8] This illustration should prompt anyone acquainted with category theory to realise that a groupoid may be defined concisely in its terms. Indeed, a groupoid is just a small category in which every arrow is invertible. This much curter definition points to an important association of groupoid theory with category theory, as we shall see later. From this perspective, groups can be seen to be special cases of groupoids, that is, they are groupoids with only one object. Alternatively, in terms of the definition above, a group may be represented as a groupoid in which the set A is a singleton, and where B corresponds to the set of group elements seen as permutation maps on the group.

9.3 HOW GROUPOIDS COMPARE WITH GROUPS

The fact that groups are just a type of groupoid raises the possibility that groupoids comprise a more conceptually basic variety of object. The first mathematician into whose consciousness groupoids explicitly appeared seems to have been H. Brandt. In his research on quaternary quadratic forms he found that he could define a composition on classes of forms, but unlike in the binary case where a group is involved, this composition was only

[7] Note that the term 'groupoid' is also used to denote a set with a binary operation on it satisfying no further conditions. This minimalistic structure is used by Saunders Mac Lane as an example of what he calls a 'mathematical dead end' (Mac Lane 1992: 10).

[8] Again these trips are being considered here only 'up to homotopy'. Higher-dimensional groupoids can used to maintain the distinction between homotopic paths. Note also that one-way streets are being overlooked.

partial. He named the corresponding structure a *gruppoid* (Brandt 1926). As this was a continuation of a programme begun by Gauss, quite possibly groupoids might have been defined earlier. One can speculate that, had the course of history run differently, we would find what we now call groups being designated by some epithet as a type of what we now call groupoid, rather than having, as is now the case, groupoids seen as 'not-quite-groups'. On the other hand, it seems very likely that category theory would have had to have been invented first.

Historical counterfactuals do not take us far. What we need now is a comparison of the characters of the group and groupoid concepts. Given that the group structure *had* already been isolated, our question is how much was there to gain by generalising to groupoids. Groupoids will have to confront the charge that anything they achieve was already inherent in the idea of groups. Before they are allocated some of the goodwill earned by their relatives, they will need to prove sufficiently different to enable their users to do new things and to do old things more straightforwardly. To the extent that the monoid concept is a generalisation of the group concept, in that the requirement that each element has an inverse is dropped, one might imagine that it might be in a similar situation. However, it is easy to argue that the character of monoids is very different and that monoids will have to make their own way in the world. In that the idea of an inverse is central to any concept of symmetry, groupoids, as their name was designed to suggest, would appear to be lesser distortions of groups than are monoids.

In favour of the idea that groupoids are similar in spirit to groups, we find that only a small modification is required:

> Thus the groupoid I, which at first sight seems unworthy of notice, plays a key role in the theory of groupoids, and in applications. A failure to extend group theory so as to include the use of I, on the grounds that I is a trivial object of only formal interest, is analogous to failing to use the number 0 in arithmetic, a failure which in fact held back mathematics for centuries. Of course, if you allow I, then in effect you allow all groupoids since any groupoid is a colimit of a diagram of copies of I, in the same way as any group is a colimit of a diagram of copies of $\mathbf{Z}$. (Brown 1987: 121)

The groupoid I is composed of two objects, identity arrows and an arrow passing in each direction between the objects. Think of two cities with a single road between them. Thus, overcoming our resistance to groupoids is likened to that monumental moment when zero was recognised as a number – a small change with large ramifications.

The opposition[9] can pick up on the size of this change, however. One of the few explicit criticisms of groupoids portrays them as only a minor variant of groups, the real essence of the notion of symmetry. After all, it is the case that a (transitive) groupoid is isomorphic to the product of the vertex group at any object and the coarse groupoid on the elements of A, that is, the groupoid with a single arrow between each pair of elements of A. To carry out this reduction in terms of Brown's illustration, for each city select a path leading to it from Bangor. Then any trip from, say, London to Birmingham can be recreated from the designated paths from Bangor to each of these two cities and the group of round trips from Bangor. The complexity of a groupoid appears to be already contained within any of its isomorphic vertex groups, the arrows looping around at a given point. Compare this to the intricate structure theorems of groups themselves or of von Neumann algebras.

A related point is that the naming of examples of a certain class of entity acts to give the definition of that class a greater sense of importance. For instance, the largest of the sporadic simple finite groups is known as the *monster*. In addition to its vast size it has recently received additional fame through the connections established between its representations and the j-function as we saw in chapter 4. Elsewhere, we can find noted von Neumann algebras such as the hyperfinite type II_1 factor and among Lie algebras E_8 attracts much interest. With groupoids, on the other hand, no individual stands out that is not a group. One might point to the simplest groupoid which is not a group, the one we denoted I above, but it does have a very simple structure.

Against the 'trivial classification' criticism, Brown and Weinstein produce the same two counter-arguments, the first of which runs to the effect that if this criterion is to be applied rigorously then important entities such as finite vector spaces become vacuous as they are categorised simply by a natural number. What is vital for vector spaces is the linear maps between them. So it is with groupoids. Notice also that no finite vector space stands out.

The second line of defence argues that especially interesting things happen when you add extra geometric structure to groupoids. Groupoids come in several varieties: topological, measurable, differentiable, Lie, Poisson, symplectic, quantum, algebraic, etc. The geometric structure often interacts with the groupoid structure in a more complicated way than in the corresponding situation with only a group structure.

[9] Most of the opposition takes the form of a reluctance to use groupoids. Some comes in the form of anonymous referees' reports on grant proposals. We shall see some of the small amount of explicit opposition below.

I have yet to read or hear of any riposte to these lines of defence. This of course does not mean that such a thing is impossible, but it does indicate that mathematicians, unlike philosophers, have no particular inclination to engage in sustained argumentative activity. I am inclined to believe that this is due to a deficiency in mathematical training, rather than because it is unnecessary. Lakatos drew a similar conclusion when through the voice of the student Gamma he wonders:

> Why not have mathematical critics just as you have literary critics, to develop mathematical taste by public criticism? (Lakatos 1976: 98)[10]

However, while it is difficult to discover arguments passing through several turns of criticism and defence, it is quite straightforward to find a considerable range of lines of argument put forward to support a construction.

Let us now turn to consider some of the reported advantages of groupoids. There is only space in this chapter to touch on a few of these advantages, which may be classed as follows:

(1) As generalisations of groups, they fully capture the one-dimensional aspects of a situation
(2) As generalisations of equivalence relations, they cope well with the symmetries of 'bad' inhomogeneous spaces
(3) Applications in physics for groupoids have been found which go beyond the use of group theory
(4) Higher-dimensional groupoid theory is richer than higher-dimensional group theory and allows new geometric features to be measured.

9.4 THE FULL EXPLOITATION OF ONE-DIMENSIONALITY

Groups do not fully exploit all the path-like behaviour that is present in a situation, because they do not capture the intermediate stages of reversible processes. We can see this in the following example. Recall from algebraic topology that the fundamental group of a space at a base point is the set of classes of closed paths in the space beginning (and ending) at that point, where two paths belong to the same class if one can be continuously deformed to the other within the space.[11] Van Kampen's theorem tells you that

[10] See also Brown (1994: 50): 'Does our education of mathematicians train them in the development of faculties of value, judgement, and scholarship? I believe we need more in this respect, so as to give people a sound base and mode of criticism for discussion and debate on the development of ideas.'

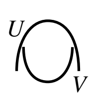
[11] It is not essential to gain a thorough understanding of the mathematics which follows. For those who wish to see a more leisurely presentation of this material I can recommend Gilbert and Porter (1994).

if you know the fundamental groups of two spaces, U and V, and of their intersection, $U \cap V$, at x, a base point in the intersection, and if all three spaces are path connected, then you can calculate the fundamental group of $X = U \cup V$ at x. This theorem is used to pass from the fundamental group of simple spaces such as the disc and circle to more complicated spaces. The fundamental group of the disc for any base point is trivial, since any loop can be deformed to the constant loop. In the case of the unit circle, loops which pass a given number of times around the origin are equivalent, so may be classified by an integer. Composition of paths then corresponds to integer addition.

Van Kampen's theorem tells you that the fundamental group of the union of U and V has as generators those of U and V, but that in addition to the relations already in place new ones may be imposed. These arise from equating the two representations of a loop situated in the intersection, according to whether it is viewed as belonging to U or to V. For example, a torus may be taken as the union of two spaces: an open rectangle, U, and a union of two annuli, V. The fundamental group of U is trivial and that of V is the free group on two generators, since it is retractable to the join of two circles. The loop c around the intersection of U and V is collapsible in U, but is homotopic to the path $a^{-1}b^{-1}ab$ in V. Therefore, in the fundamental group of the union the latter path must be put equal to the identity, or in other words, the relation $ab = ba$ is imposed and we can conclude that the fundamental group of the torus is the Abelian group on two generators:

$$\pi(U, x) = \langle | \rangle, \pi(V, x) = \langle a, b | \rangle \text{ and } \pi(U \cap V, x) = \langle c | \rangle.$$
$$\text{In } U, c \sim \text{constant path at } x. \text{ While in } V, c \sim a^{-1}b^{-1}ab.$$
$$\text{Therefore, } \pi(\text{Torus}, x) = \langle a, b \mid a^{-1}b^{-1}ab \rangle = Z \oplus Z.$$

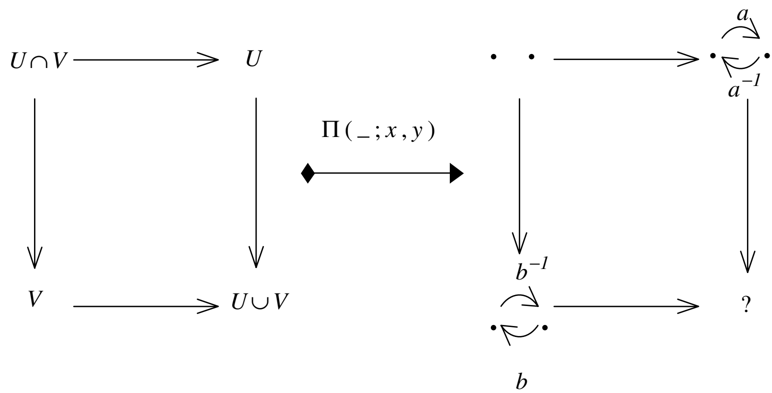
But how can the result that the fundamental group of the circle is isomorphic to the integers under addition be derived? There are several ways of doing this, none of which is as straightforward as might be expected for such a basic shape. It might have been hoped that it could be found by applying van Kampen's theorem to the circle seen as the union of two overlapping open intervals:



That this will not be possible is clear from the observation that, in the words of Rogers and Hammerstein, 'nothing comes from nothing' – the fundamental groups of U and V being trivial and so providing no generators. The problem arises from the fact that the intersection of U and V is not connected, but is composed of two disjoint intervals. An arbitrary choice must be made as to which component will contain the base point. As we saw earlier with the disconnected United Kingdom, fundamental groups do not cope well with such spaces as they can measure only one connected component. This presents something of an anomaly since open regions of n-dimensional space are the basic building blocks of manifolds in modern topology and geometry. There is, of course, nothing wrong with the various ways of establishing the fundamental group of the circle. No Lakatosian proof analysis conducted on any of them will discover a counter-example. One also cannot deny the interesting connections with many other branches of mathematics, e.g., the winding number of a function about a point in complex analysis. The point is, however, that the van Kampen philosophy just ought to work there. To call this case a 'heuristic counter-example' would be to stretch the meaning of the term beyond that given it by Lakatos (1976: 83), but I think he would have approved.

In fact it can be made to work but only if one extends the fundamental group idea to allow loops at several base points and paths between them. One for each component would be enough. But after this extension we shall no longer be dealing with a group since composition will not be possible for each pair of paths. The fundamental groupoid of a space X with respect to a given set of base-points has this set as A and equivalence classes of paths between two such points as the elements of B. Groupoid status is assured owing to the fact that each path may be run backwards.

One can now prove a van Kampen theorem for fundamental groupoids. I shall not enter into details, but note that its phrasing in category theoretic terms is very simple. The category of topological spaces has *pushouts*. A pushout may be thought of as a kind of sum of two objects which identifies or keeps separate precisely what ought to be identified or kept separate. In particular, the pushout of the injections of the intersection, $U \cap V$, into U and into V is their union, $U \cup V = X$. The fundamental groupoid construction provides a *functor* from the category of topological spaces to the category of groupoids which preserves pushouts. With U and V overlapping intervals forming a circle, and x and y points in the components of the intersection, we have:



This tells us that insofar as we are interested in one-dimensional data, much about the compositional structure of topological spaces has been captured algebraically. Indeed, the fundamental groupoid of the circle with a pair of base-points is the pushout of the diagram of groupoids (denoted by the question mark) and one can prove that the vertex group at each object in this groupoid is isomorphic to the infinite cyclic group.[12]

Although groups are more familiar to mathematicians, the restriction to one base point may also lead to an unwieldy presentation in terms of generators and relations, rather as it would be inconvenient to view all British road trips in relation to Bangor. As Grothendieck remarks on the benefits of groupoid presentations:

> people are accustomed to work with fundamental groups and generators and relations for these and stick to it, even in contexts when this is wholly inadequate, namely when you get a clear description by generators and relations only when working simultaneously with a bunch of base-points chosen with care – or equivalently working in the algebraic context of *groupoids*, rather than groups. Choosing paths for connecting the basepoints natural to the situation to one among them, and reducing the groupoid to a single group, will then hopelessly destroy the structure and inner symmetries of the situation, and result in a mess of generators and relations no one dares to write down, because everyone feels they won't be of any use whatever, and just confuse the picture rather than clarifying it. (Quoted in Brown 1987: 118)

Groupoids even provide new information about groups themselves, because they possess some important properties not shared by groups. For example,

[12] The identity arrows of the groupoids have not been shown. The composition of arrows a and b is an arrow from x to itself. As nothing tells us to equate either it or any iterate of it to the identity arrow at x, we do not. Indeed, it forms a generator for the vertex group at x. Notice how the simple groupoid I, which marks the gap between groups and groupoids, crops up here.

the category of groups cannot support several useful constructions:

> One of my hopes in preparing the text was to convince students of group theory that it is often profitable to cross the boundary between groups and groupoids. The main advantage of the transition is that the category of groupoids provides a good model for certain aspects of homotopy theory. In it there are algebraic analogues of such notions as path, homotopy, deformation, covering and fibration. Most of these become vacuous when restricted to the category of groups, although they are clearly relevant to group-theoretical problems. (Higgins 1971: vii)

A further significant flaw with groups is that:

> One of the irritations of group theory is that the set Hom(H, K) of homomorphisms between groups H, K does not have a natural group structure. However, homotopies between homomorphisms of groupoids H, K may be composed to give a groupoid HOM(H, K) with object set Hom(H, K). (Brown 1987: 122)

This construction leads to a groupoid isomorphism HOM($G \times H$, K) HOM(G, HOM(H, K)), an example of a very widespread structural law which is found even in the simplest theories, such as arithmetic, $a^{(b \times c)} = (a^c)^b$, and propositional logic, A & $B \vdash C$ if and only if $A \vdash B \rightarrow C$. The desirability of this property is also the reason some topologists give for working with compactly generated topological spaces.[13] Use of this construction again provides information about groups.

9.5 GROUPOID ALGEBRAS USED TO COMPENSATE FOR BAD SPACES

First let us consider a simple way in which the algebra of complex $n \times n$ matrices may be reinterpreted from a groupoid perspective. Take the pair (or coarse) groupoid $A \times A$, where $A = \{1, 2, 3, \ldots, n\}$. This is the groupoid where for each pair of members of A there is a single arrow passing from the first to the second. Next, take the algebra of complex valued functions on the arrows of this groupoid. A type of multiplication known as convolution may now be defined on these functions, generalising a similar construction used for groups. The value of the convolution of two such functions, f and g, denoted f^*g, on a pair (i, k) is the sum over j of products of the form $f(i, j) \cdot g(j, k)$. This is completely equivalent to matrix multiplication, where if the ijth entry of a matrix, M, is $f(i, j)$ and similarly for a matrix N and the function g, then the matrix corresponding to f^*g is simply $M \cdot N$.

[13] Cf. Mac Lane (1971: 184).

Convolution algebras can be defined similarly for all groupoids. In view of the importance of group convolution algebras, which over the integers underlie the harmonic analysis of Fourier series, we can see the potential for this generalisation. Indeed, groupoid convolution algebras play a major part in the field of non-commutative geometry. This field is based on the observation that a commutative algebra may be construed as the collection of functions on a space. By analogy the non-commutative algebras are seen as arising from functions on a *non-commutative* space. Such a space is often characterisable as the orbit space of a groupoid. To recapture a space of functions on what might be termed a 'bad' space with an inadequate collection of ordinary set-based functions, the focus is shifted from the space to the groupoid representing it, and from there to a suitable convolution algebra.

To take a simple example, consider the topological space formed from two intervals, $\{0\} \times [0, 1]$ and $\{1\} \times [0, 1]$, by identifying pairs of points $\{(0, a), (1, a)\}$ for all $a \neq 1/2$. This looks very much like just one interval, except for a small split half-way along it. It is a perfectly legitimate topological space, although not Hausdorff. But then if we try to characterise the space by the continuous complex functions it can support, we find that the points $(0, 1/2)$ and $(1, 1/2)$ cannot be distinguished. From the point of view of continuous functions the space is indistinguishable from a simple interval. However, this space can be reformulated as the orbit space of a topological groupoid where the objects are the points along the two intervals, and besides the identity arrows, there are pairs of inverse arrows between the points to be identified. This groupoid is not equivalent (homotopic) to the groupoid of identity arrows on an interval, as becomes apparent by the difference between their convolution algebras of complex continuous functions. This may be represented in the case of the 'bad' non-Hausdorff space as the algebra of 2×2 matrices of continuous complex functions on $[0, 1]$ which are diagonal at $1/2$.

This is a simple example of an important use for groupoids, which Weinstein claims:

> leads us to the following guiding *principle* of Grothendieck, Mackey, Connes, Deligne, . . .
>
> Almost every interesting equivalence relation on a space B arises in a natural way as the orbit equivalence relation of some groupoid G over B. Instead of dealing directly with the orbit space B/G as an object in the category S_{map} of sets and mappings, one should consider instead the groupoid G itself as an object in the category G_{htp} of groupoids and homotopy classes of morphisms. (Weinstein, 1996: 748, my emphasis)

Here we see once more the expression of a broad principle, which is what I took earlier to be the mark of the conceptual. Notice the vagueness of the wording. There is plenty of scope here for argument about what counts as 'almost every' and as 'interesting' and as to whether it really is so very 'natural'. As we saw in chapter 8, this vagueness is reminiscent of the kind of language with which the aims and means found at the heart of a research programme are articulated.

Groupoids as generalisations of sets, equivalence relations, groups and group actions permit a unified reformulation of these concepts. This fact by itself is not sufficient reason to adopt them; concepts must work harder to pay their way. Whether a reformulation constitutes a clarification and whether a unification may be considered important are two deep questions. For Weinstein the approach to orbit spaces outlined in this section *has* led to an important unification and he claims that Alain Connes's book on non-commutative geometry, which utilises this construction:

> shows the extent to which groupoids provide a framework for a unified study of operator algebras, foliations, and index theory. (Weinstein 1996: 745)

9.6 APPLICATIONS AND OLD CONJECTURES

Having applications in the sciences, computing or engineering is incontrovertibly a good thing for a piece of mathematical theory. However, questions remain concerning the importance of the application and whether a particular theory is indispensable in a given application. In the case of groupoids, we hear in an announcement for a 1998 conference, 'Groupoids in Physics, Analysis and Geometry', that:

> The uses of groupoids in physics come from two main sources. The first is Alain Connes' theory of noncommutative geometry, in which groupoids are a main source of examples of noncommutative spaces. This theory is being studied very actively by physicists, and by mathematicians. Bellisard's work studying the quantum Hall effect via noncommutative geometry has led to the study of connections between solid state physics and noncommutative geometry models associated with tilings.
>
> The second major source of the use of groupoids in physics is the general theory of quantization in mathematical physics. A theory of quantization has been introduced by V. Maslov and A. Karasev, and a version due to Alan Weinstein has been actively developed by him and his collaborators. One step in this program is to associate a symplectic groupoid to a given Poisson manifold. (Kaminker 1998)

As for the first of these uses, Bellisard, a solid state physicist, uses non-commutative geometry to explore the non-commutative Brillouin zone of an aperiodic medium. While discussing the way mathematicians succeeded in capturing some forgotten intuitions of Heisenberg in the context of the C^*-algebra approach to quantum mechanics, he claims that:

> The breakthrough went with the notion of a groupoid . . . which is nothing but the abstract generalization of the notion of transition between stationary states as defined by Bohr and Heisenberg. (Bellissard 1992: 551)

This relates to the reformulation we saw above of matrix algebras as the convolution algebras of groupoids. It is Connes's position that for Heisenberg the groupoid idea came first, albeit implicitly, and hence his requirement of matrices. As he explains in a section of his book (1994: 33–9) written to 'remove this prejudice [towards groupoids]' (1994: 7), the groupoid idea is present in the case of an electron's transitions between energy levels in the atom – the transitions from level i to level j and from level k to level l may be composed iff $j = k$.

With the physical world providing only a very indirect constraint on mathematical theorising, mathematicians have worried that pieces of research, although perfectly correct, may be of little or no value. They have, therefore, sought ways internal to mathematics of adjudicating whether a theory is on course. One way of doing this is to set up, as Hilbert famously once did, a series of problems to be solved. Then a theory's success in solving any of these problems can be taken as a token of its worth. Thus:

> Often a test for the value of a new theory is whether it can solve old problems. *De facto*, this limits the freedom of a mathematician, in a way which is comparable to the constraints imposed on a physicist, who after all doesn't choose at random the phenomena for which he wants to construct a theory or devise experiments. (Borel 1983: 14)

One senses here the concern that without such constraints mathematicians may find themselves wandering aimlessly through a world of mathematical possibility.

Now, on this score there appear to be no clear successes for groupoids,[14] but remember that groupoids can be used to discover new properties about familiar things, namely, groups. Just because these properties were unforeseen, and so no conjectures made about them, seems to be no reason to mark groupoids down.

[14] It may be argued, however, as Brown has, that Grothendieck's reliance on groupoids means that they are due some credit for Wiles's proof of Fermat's Last Theorem.

As a second response, we might say that while it is reassuring that a theory solves famous old problems, mathematics must also be about opening up new areas by the elaboration of mathematical ideas. Frequently, for those venturing into unknown territory there is no shortage of constraints resembling the 'old problems' one. For example, you may find that a definition that seems to go in the right direction unexpectedly makes contact with older work, or that the method you are using to overcome obstacles which are preventing you from performing a construction analogous to an earlier one gives you a much clearer picture of the whole domain. We now turn our attention to some new areas.

9.7 NEW PROSPECTS: HIGHER-DIMENSIONAL ALGEBRA

The fundamental group of a space need not be Abelian. Imagine yourself based at the crossover point of a figure of eight. The path which takes you clockwise round the upper loop, then around the lower one is not equivalent to the path taking the loops in the opposite order. However, higher homotopy groups *are* always Abelian. Here, rather than throwing loops into our space to see what we can catch, we are throwing spheres (two-dimensional and higher). Just as we can see a loop belonging to the fundamental group as a line where the endpoints are identified, we can see an element of the second homotopy group as a map of a square into the space where the perimeter gets mapped to the base point. Think of a net having being cast by a fisherman, who now holds its opening. Then we can set up a composition in two directions, corresponding to the two dimensions of a square.

Let us give an idea of what happens when we compose in one direction. Imagine two square nets joined along one edge and pinned to the table along their perimeters. Push all the raised part of the left net into its upper half so that the rest lies flat, and all the raised part of the right net into its lower half. Then make the raised parts swap sides and permits them to reform their original shapes. You may be able to see that the two multiplications coincide in a single commutative operation.

The question then arises as to whether this commutativity is due to the higher dimensional homotopic nature of spaces or whether it is a failure on the part of groups to capture this nature. On the face of it there is no reason to expect homotopy to become simpler in higher dimensions, suggesting that the fault lies with the algebra, which must be refined to detect deeper features of geometric reality. As part of the process of capturing these deeper features, in the 1940s the topologist J. H. C. Whitehead devised what are known as crossed modules. Brown has succeeded in using them in this way and he notes that:

> information about even such an apparently simple computation as a second absolute homotopy group of this mapping cone is tightly bound to information on crossed modules. There is at present no alternative description [to crossed modules] of this second homotopy group in algebraic terms. This highlights some basic difficulties of homotopy theory, and also suggests that homotopy theory is an essentially non abelian subject. The abelian homotopy groups, even as modules over the fundamental group, give only a pale shadow of the homotopical structures. (Brown 1999: 32)

Now, crossed modules turn out to be equivalent to groupoid objects within the category of groups, i.e., groups on which there is a compatible groupoid structure. Unlike in the purely group theoretic case, the two structures interact non-commutatively.

The next step is to look for double groupoids, groupoid objects within the category of groupoids – or, if you prefer, two interacting groupoid structures. The simplest way to catch a glimpse of what is happening here is to think of mapping a square into a topological space as we did above, but this time with no restrictions on where its perimeter lands. We still have multiplications running in two directions but, in the spirit of groupoids, only if the paths corresponding to the adjoining sides of two square are equal. Brown managed, after years of effort, to achieve a van Kampen-style theorem in two dimensions. These mark some early steps of an enormous programme we shall discuss in chapter 10.

We should note that multiple and higher-dimensional groupoid theory has not penetrated into the non-commutative geometry mentioned in sections 9.5 and 9.6, although it is starting to be used in differential geometry (see, e.g., Mackenzie 1992). In view of the fact that Charles Ehresmann was exploring such ideas in the late 1950s, we may wonder why the development has been so slow. Has there been an undervaluing of the conceptual?

9.8 THE CONCEPTUAL AND THE NATURAL

Recall my quoting earlier Brown making 'the heretical suggestion that the *natural concept* is that of groupoid rather than group' (Brown 1999: 4, my emphasis). The philosophical treatment of the notion of a mathematical concept is still to be done, but it is interesting to note that the category theorist William Lawvere, co-author of *Conceptual Mathematics* (1997), has expressed the view (Brown 1987: 129) that the term *group* should be taken to refer to what is now covered by *groupoid*. Most mathematicians will find this hard to accept, having been taught to accept the group concept as the natural one. Let us approach these matters by discussing the idea of *naturalness*.

The epithet 'natural' is never far from mathematicians' lips when they describe their favourite constructions. It even appears in mathematical terms such as natural number, natural transformation and natural deduction. Only for the first of these can some connection with the physical world be claimed, as the natural numbers constitute possible responses to questions of the kind 'How many elephants are there in this National Park?'. Mathematicians sometimes play humorously on this idea. In what must be one of the wittiest mathematics textbooks ever written, Frank Adams discusses the situation where two spaces have the same homology groups, yet their fundamental groups are 'wildly' different. He tells us that one of the spaces can even be taken to have trivial higher homotopy groups:

> By now we have theorems saying that this situation is common; Kan and Thurston show that given almost any space Y, you can approximate it homologically by an Eilenberg-Mac Lane space EM(π, 1) for some weird and artificial group π. However, we should perhaps be more concerned with cases where this situation arises in *nature*. (Adams 1978: 84, my emphasis)

Without wishing to labour the point, you are not going to meet with this situation while on safari in a National Park. You won't even meet with it while doing theoretical physics. But you may encounter it, without artificially engineering it, while working in reasonably well frequented regions of mathematics. Adams's sentiment is that for a type of construction to be worth defining or for a type of situation to be worth describing, there ought to be examples readily available.[15]

For Adams, if an instance of the situation he describes occurs and he decides not to count it as arising in nature, he is not thereby banishing that instance from its fellows. This may be contrasted with what Lakatos (1976: 23) designated as 'monster-barring', when a proposed counter-example which may be thought to have refuted a claim about a class of entities is declared not to belong to that class. Here, the 'monster' is unnatural – it does not have in its nature what it takes to be a member. Adams's reaction is more

[15] A similar idea is expressed by Robert Solomon when he points out that, despite the fact that the majority of finite groups are nilpotent of nilpotence class 2, and so far from being simple, 'experience shows that most of the finite groups which occur "in nature" – in the broad sense not simply of chemistry and physics, but of number theory, topology, combinatorics, etc. – are "close" either to simple groups or to groups such as dihedral groups, Heisenberg groups, etc. which arise naturally in the study of simple groups' (Solomon 2001: 347).

typical of the contemporary mathematician, who knows enough about the conceptual twists and turns that have occurred in the discipline since the mid-nineteenth century not to take talk of unnatural monsters too seriously and illustrates an important point about the changing conceptions of mathematicians towards the role of definitions. It would be a valuable exercise to make comparisons with claims of naturalness from earlier times.

We can see the modern attitude illustrated in the following example from non-commutative geometry. While discussing the set, X, of Penrose tilings,[16] Alain Connes notes that, although it is clearly a spatial entity, when it is treated with the classical tools of point-set topology it cannot be distinguished from a point. This is because among the peculiar properties of X we find that, given any two distinct tilings, a finite portion of one of them of whatever size will be found occurring infinitely often within the other. Hence:

> [t]he natural first reaction to such a space X is to dismiss it as pathological. (1994: 6)[17]

This may sound rather like monster-barring, but what Connes means here is that:

> To a conservative mathematician this example might appear as rather special, and one could be tempted to stay away from such spaces by dealing exclusively with more central parts of mathematics. (1994: 94)

So, it is not a question of excluding X by modifying the definition of a topological space. Either one accepts it as an odd sort of space and then ignores it or, like Connes, one brings new tools to bear upon it, in this case a convolution algebra on the associated groupoid. The situation may be summarised well by describing X as a *heuristic counter-example* to the notion that classical topology is adequate to deal with all topological spaces.

Another way of arguing for the naturalness of a concept is in terms of the inevitability of its discovery. There seems to be a widespread feeling that

[16] These are the quasi-periodic tilings of the plane with local 5-fold symmetry whose patterns have been found to occur in the natural world in what are termed *quasi-crystals*. X may be interpreted as the orbit space of a groupoid.

[17] Notice here how 'natural' is being used about the mathematician rather than about the mathematical entity. Perhaps *natural deduction* involves both. It turns the reasoning processes of mathematicians into an entity which may be investigated mathematically.

however mathematics was to reach anywhere near the level of sophistication we see today, a basic concept such as that of a group was bound to be formulated, while there is disagreement over whether the same could be said for groupoids. Powerful evidence that a concept was inevitably going to be forged is to show that it was required independently by researchers working in different fields. A convincing case can be made that this was so for groups. As for groupoids, after their introduction by Brandt in 1926, researchers have for their own reasons deemed it worthwhile to introduce them into the theory of field extensions, non-commutative ring theory, algebraic logic,[18] partial differential equations, category theory, differential geometry, differential topology, foliations, non-Abelian cohomology and ergodic theory.

Independence of use is most marked when the researcher coins a new name for the concept as when George Mackey working in ergodic theory used the term 'virtual group' to refer to what amounts to a groupoid. A more vivid illustration of this phenomenon is reported in a survey article on Lie algebroids and Lie pseudoalgebras[19] by Kirill Mackenzie. He remarks (1995: 100) that the notion of a Lie pseudoalgebra had been devised independently by well over a dozen researchers in almost identical fashion, each with a different name.

The mathematicians' notion that some ideas are fundamental and will inevitably emerge reveals a degree of faith resembling that motivating the scientists' discursive line, treated by Gilbert and Mulkay (1984), that 'Truth will out'. These ideas are deemed to possess such intrinsic value that they can overcome the vagaries of the human research effort. The strongest form of this sentiment would maintain that some concepts will necessarily appear and that, by virtue of their nature, they rather than the user will determine their use. Opposed to this faith that methods of research will not stand in the way of important ideas and their proper deployment is the notion that even good ideas that have at some time surfaced into the awareness of a mathematician may be lost to future generations. For Gian-Carlo Rota this is no rare event:

> On leafing through the collected papers of great mathematicians, one notices how few of their ideas have received adequate attention. It is like entering a hothouse and being struck by a species of flowers whose existence we did not even suspect. (Kac *et al*. 1986: 1)

[18] In the early 1950s, Jónsson and Tarski required *generalised Brandt groupoids* to capture the calculus of binary relations. These were not required to satisfy transitivity.

[19] The former are related to Lie groupoids as Lie algebras are to Lie groups and are special cases of the latter.

Presumably, then, for Rota some of these powerful ideas may be lost for a long time, and possibly forever.

This picture offers the mathematician the opportunity of presenting their work as allowing the recovery of some of this lost treasure. Some mathematicians are appealed to more than others in this respect. Someone like Sophus Lie working in a branch of geometry at a time when standards of rigour had not become well established, but when mathematicians were closer to 'nature', makes for an excellent target. Hence:

> [t]he concept of groupoid is one of the means by which the twentieth century reclaims the original domain of application of the group concept. The modern, rigorous concept of group is far too restrictive for the range of geometrical application envisaged in the work of Lie. (Mackenzie 1987: vii)[20]

There are various ways of responding to claims of naturalness. I may suggest a new concept to be the natural development of an earlier one, or the natural idea on which to base an attack on an important problem, or the natural way to illuminate some phenomenon in the physical sciences. You reply by claiming it to be an unnecessary modification of a perfectly serviceable idea, with little to be gained from its acceptance other than as a boost to my publication record. A colleague then chips in with her view that while she does not believe it to have achieved what I claim, it may prove useful as a temporary measure, and may lead to a better reconceptualisation of the field. This threefold distinction – fundamental, convenient and pointless – is quite common. In the following quotation we see groupoids consigned to the middle category:

> [definitions] like that of a group, or a topological space, have a fundamental importance for the whole of mathematics that can hardly be exaggerated. Others are more in the nature of convenient, and often highly specialised, labels which serve principally to pigeonhole ideas. As far as this book is concerned, the notions of category and groupoid belong in this latter class. It is an interesting curiosity that they provide a convenient systematisation of the ideas involved in developing the fundamental group. (Crowell and Fox 1977: 153)

Naturally, Brown sees this as much more that an 'interesting curiosity'. Rather, the elegance of the systematisation was read by him as a clue that groupoids could play a very large role in this area.

The *convenient* class is very broad. Crowell and Fox judge groupoids to be at the lower end, bordering on the pointless. Meanwhile, in a discussion of

20 Lie is also selected by two exponents of synthetic differential geometry who claim to be able to allow his intuitive reasoning to be fully captured in a rigorous framework (Moerdijk and Reyes 1991).

the role of groupoids in differential geometry, Kumpera has them straddling the boundary between the fundamental and the convenient:

Bundles are of course extremely useful objects but, as Ehresmann would probably say, groupoids are somehow closer to the truth. As for connections, they are an extremely useful algorithm whereas groupoids (and algebroids) are an extremely useful concept. (Kumpera 1988: 359)

Rota goes as far as to say of the notion of groupoids that it 'is one of the key ideas of contemporary mathematics' (Bergeron *et al*. 1997: vii).

We see here the idea of a split between something intrinsically worth studying and something valuable as a tool to be used in the study of something else, even if this means that the tool needs to be studied to know if it is up to the job.[21] But this distinction is not permanent. In the course of time the status of mathematical entities may change from being viewed as useful tools to becoming fully fledged objects. In the search for a solution to a problem, means are introduced which can then become interesting in their own right, and further means will then be necessary to study them in turn:

Once we have a genuine need for some mathematical idea as a matter of language, that idea has arrived; it is hardly necessary to discuss its status as a useful technical tool. (That sentence is not intended to exclude the possibility that some authors may try to introduce language we can do without.) (Adams 1978: 79)

The criticism which may meet this kind of promotion is that means have been unjustifiably raised to the level of ends.

Category theory is often picked out for this treatment as when Miles Reid remarks:

The study of category theory for its own sake (surely one of the most sterile of all intellectual pursuits) also dates from this time; Grothendieck himself can't necessarily be blamed for this, since his own use of categories was very successful in solving problems. (Reid 1988: 116)[22]

As a contemporary algebraic geometer, Reid could not possibly deny category theory its 'useful' status. Even in a textbook aimed at undergraduates (Reid 1988) he allows it to make an occasional appearance. What he appears

[21] One might be tempted to equate the class of that which is worthy of study for its own sake with the *natural*, but mathematicians might easily talk of a piece of what they term *machinery* as natural. Topologists actually refer to the apparatus for converting between spectra and spaces as 'machinery'. See Adams (1978, ch. 2). I have sometimes been asked why the methodologist need bother studying (difficult) contemporary theory development. The answer is clear: if she wants to find out how today's mathematical 'technology' works, she has no choice.

[22] For robust rebuttals of Reid's remarks see Brown (1994: 49).

to be criticising here is the view that categories are like meta-Abelian groups, braids, projective varieties or sets, entities to which a mathematician today may devote her whole career without being required to demonstrate their applicational virtues. His is probably a majority position at the present time,[23] but again within this position finer distinctions may be made ranging from those who think category theory works well as a convenient means of representing a body of theory to those who think its principles can at times provide very strong guidance for future research. Moreover, promotion for categories may be imminent. With the race to develop a theory of weak n-categories currently proceeding at a frenetic pace, one may argue that important entities of intrinsic conceptual interest are being carved out.

This mention of category theory leads us to a narrower, more technical, sense of the term 'natural'. As we saw above, the classification of a transitive groupoid requires the choice of an object, and the choice of an arrow from that object to each other object. These choices cannot, however, be said to be natural, insofar as there was no good reason to favour them over any others. In other words, the reduction of a groupoid to a group suffers from the need to make an arbitrary, or unnatural, choice. This might remind the reader of a similar unforced choice in the theory of vector spaces. To establish an isomorphism between a finite vector space and its dual, the space of linear maps to the ground field, one must make an arbitrary choice of a basis. By contrast, when it comes to establishing an isomorphism between a vector space and its double dual, there is a natural map, namely, the map which sends a vector to the map equivalent to evaluation at that vector. This last example is part of the category theoretic folklore. It provided Eilenberg and Mac Lane with the notion of a natural transformation between functors, in this case between the identity and double dual functors on the category of vector spaces.

Here we see again the idea that it is preferable not to make arbitrary choices. It accords with the notion that one should not privilege a member of a collection without good reason. Recall that we constructed the fundamental groupoid of a space precisely to avoid creating a privileged base-point. This kind of privileging occurs in the notion of a principal fibre bundle from differential geometry where one fibre is singled out. A more even-handed or 'democratic' treatment of this very important geometric idea is to work with Lie groupoids, those for which the sets we have

[23] Of course, the positing of a piece of research as 'for its own sake' is open to challenge.

denoted *A* and *B* are smooth manifolds satisfying some further conditions. This is not just done for convenience – it makes a concrete difference:

> The need for privileged fibres has an important consequence when one needs to consider group actions. A single automorphism of a principal bundle can be transported to its Lie groupoid and vice versa, using a chosen reference point, but for general (nontrivial) groups of automorphisms it is impossible to choose reference points consistently. The automorphism groups of a principal bundle and its Lie groupoid therefore do not correspond and there is a notion of group action for Lie groupoids which makes no sense for principal bundles. (Mackenzie, personal communication)[24]

In sum, a full analysis of the use of the term 'natural' by mathematicians through the ages would require a book-length treatment. As used today it possesses several shades of meaning, which blend into each other to some extent, relying as they do on a sense of freedom from arbitrariness and artificiality. Promoters of groupoids see them as a natural concept since:

(1) One comes across them in the course of carrying out research in many areas of mathematics, without resorting to artifice
(2) They embody a simple, non-artificial idea, which permits them to measure the symmetries of families of objects
(3) They permit one to model situations without requiring that arbitrary choices be made.

9.9 CONCLUSION

This chapter should be seen as an early foray into an extremely complex subject. We have discovered that mathematicians will on occasion argue for the acceptance and further study of a piece of theory by indicating a panoply of good qualities. Philosophers should note that in the case treated here arguments for the 'existence' of groupoids did not figure in the array surveyed. This is through no oversight on my part – mathematicians make no use of the idea in their advocacy of the groupoid concept. On the other hand, turning to the arguments they do use, it is reasonable to wonder why so many different types are employed. I would explain this by pointing out that individual mathematicians weight the candidate criteria for progress idiosyncratically. One sets greatest store by the unificatory power of a concept, another by the potential for applications, a third by its ability to help resolve outstanding problems. Brown, Weinstein and other promoters of groupoids may have their own preferences as to the

[24] See Hitchin (2001) for the 'democratic' advantages of groupoids.

reasons why groupoids should be accepted, which reflect their sense of what mathematicians should be aiming to achieve, but they wish to cater for as many tastes as possible. Something akin to what Laudan terms the *dominance* of a theory over another is at stake. For Laudan (1984), even if the aims of scientists are varied, a theoretical development will still be held unanimously to constitute progress when it satisfies each scientist's criteria.

I believe there is much we can learn from parallel work in science studies. For example, Jardine's *The Scenes of Inquiry* (Jardine 2000) presents a highly sophisticated pragmatist philosophy, which stresses the importance of what he terms 'calibration', the act of making a theory measure up to earlier ones. The considerations of importance treated in this chapter would seem to be mathematicians' forms of calibration, where solving old problems is just one form. A particular concern of this chapter has been a disagreement about the weighting of these forms of calibration. We have encountered the thought that, whether seen as a goal in itself or as a means to a further end, a bias against conceptual reformulation and development have acted to delay the acceptance of groupoids. From this perspective, it took the efforts of conceptually daring mathematicians, such as Grothendieck and Connes, to set up programmes in which the use of groupoids became a matter of course.[25] A still more daring act of faith has been needed to pursue higher-dimensional groupoids. This research has provided key insights into how to develop higher-dimensional algebra, whose revolutionary credentials we shall be considering in chapter 10.

It is worth noting that for debates about the value of groupoids even to begin, there must be some shared ground, one or more absolute presuppositions, held by the participants. We can detect one fairly clearly – the presupposition that there is a distinction between concepts of great mathematical importance and concepts which are mathematically pointless. People disagree about how this distinction fits with the space of mathematics, but they do not question that it exists. Now, we could plausibly claim that the idea of such a distinction has been operating at least as far back as the Greeks. It is curious, then, that many contemporary varieties of philosophy of mathematics disregard it.

One vision I share with Lakatos about how the future of mathematics might be shaped involves encouraging both a heightening of the level of historical awareness among the mathematical community and a facilitation of the expression of critical attitudes. This is not to say that the present

[25] See Cartier (2001) for some insight into why groupoids link the geometric visions of these two giants of mathematics.

situation is hopeless, but it can only be good for the health of mathematics to make improvements. One might also hope for useful communication with those philosophers who are well-versed in areas of modern mathematics. Unfortunately, the latter are not too numerous at present, but successful exchanges should encourage an expansion. Eventually one might hope to construct arenas where mathematicians and those working in the history, philosophy and sociology of mathematics can come together to permit sustained discussion of common concerns to take place.