ALAN BAKER

MATHEMATICS, INDISPENSABILITY AND SCIENTIFIC PROGRESS

1. INDISPENSABILITY AND JUSTIFICATION

1.1. *Introduction*

Are there good reasons for including mathematical objects such as numbers, sets, and functions as part of our ultimate catalogue of the furniture of the universe? Recent debates within the philosophy of mathematics over this sort of general ontological question have centered on the pros and cons of the so-called *Indispensability Argument.* The basic idea behind this argument is quite straightforward. When faced with a general existence question such as 'Do mathematical objects exist?', we should look to our best available theories of the world for guidance. Our current best theories of the world – by general consensus – are the theories of empirical science. And current science (especially physics) quantifies over mathematical objects. Hence we have good reason to believe in the existence of mathematical objects, unless and until we can do science without postulating them. In short, mathematics is *indispensable* for science. One way of formulating the Indispensability Argument is as follows;

(1) We have good reason to believe in the literal truth of our best scientific theories.

(2) Mathematics is indispensable for science.

(3) We have good reason to believe in the existence of (abstract) mathematical objects.

The Indispensability Argument has been attractive to platonists as a defensive tool because it is an *external* argument for the existence of mathematical objects. It connects the literal truth of mathematics with the literal truth of science by claiming that belief in the literal truth of our

Erkenntnis **55:** 85–116, 2001.

best scientific theories carries with it belief in the literal truth of mathematical claims embedded in these theories. By hitching his waggon to the scientific realist's train, the platonist hopes to gain external support for his ontological claims about mathematics. In particular, the Indispensability Argument has the potential to block sweeping fictionalist or instrumentalist charges that the only arguments for platonism are blatantly question-begging.

Conversely, the Indispensability Argument has been an attractive target for nominalists. If a nominalist can persuade herself that the Indispensability Argument is the *only* good argument for platonism, then this opens up the possibility that platonism can be directly undermined by means of a technical reconstructive program. Remember that the Indispensability Argument depends for its force on a strong modal claim, that science *cannot* be done without mathematics. If the nominalist can come up with a mathematics-free reformulation of science then this shows that science can be done without mathematics, and hence that mathematics is dispensable. Hartry Field's 1980 monograph, *Science Without Numbers*, is probably the best-known attempt to reconstruct science in a purely nominalistic fashion.[1]

A second line of criticism for the nominalist is to attack the validity of the Indispensability Argument, and in particular the purported link between indispensability and truth. Van Fraassen has championed this latter view for the case of concrete unobservables such as electrons, arguing that inference to the best explanation – on which indispensability-style arguments are implicitly based – is not in general valid.[2] More recently Maddy, Azzouni and others have questioned this link in the specific case of abstract mathematical object.[3] Although I think that these latter criticisms raise serious and interesting issues, I will not be addressing them in this paper. My focus instead will be on criticisms of the first sort, which are directed specifically toward the claim that mathematics is indispensable for science.

I shall refer to this latter claim as the *Indispensability Thesis.* It is a thesis that the platonist must defend if she is to secure her position against the Fieldian nominalist attack. It is also a thesis whose truth depends – at least in part – on contingent facts about current and future science, and in this sense it is empirical in nature. The possibility of a knockdown argument either for or against the Indispensability Thesis, proceeding from general philosophical considerations, therefore appears unlikely. However this does not mean that arguments cannot be given, based on philosophical analysis of scientific practice, which may alter our assessment of its plausibility. I am sympathetic to the platonist side of the argument here, and to the reasonableness of the Indispensability Thesis. My aim in this paper is to show that the Indispensability Thesis is less tendentious and more plausible than it is normally taken to be, and I shall thus be defending the Thesis as one that it is reasonable to believe given our current evidence.

Platonists have typically defended the Indispensability Thesis by trying to expose technical flaws in the various nominalist reconstructive programs that have been proposed. The defense I wish to mount is more radical, and if successful then it makes squabbles over technical details largely irrelevant to evaluating the Thesis.[4] I shall argue that recent nominalist programs have missed the true force of the Indispensability Thesis, and that even if such programs could be carried out they would not necessarily undermine it. In the debates that have been sparked off by Field and his sympathizers, both defenders and critics of the Thesis have failed to appreciate the range of ways in which mathematics functions in science. They have concentrated on mathematics as a tool for proving scientific results, and on the question of whether this proof-theoretic power can be adequately reproduced using mathematics-free theories. My claim is that reproducing this function alone is not enough to establish that mathematics is dispensable, for mathematics plays other roles in science that are distinct from its sheer deductive power. In particular mathematics may function as a tool for discovering new results, and as a heuristic aid for the development of new scientific theories. These dynamic features of mathematics are crucial to scientific practice and to scientific progress. Unless and until they too can be adequately reproduced using mathematical-object-free theories, the hypothesis that mathematics is indispensable for science is not undermined.

There is one theme that will surface repeatedly in the subsequent discussion and that I want to stress at the outset. It derives from the insight that – given the naturalistic basis of the Indispensability Argument, which rejects the idea of philosophy as a higher court of appeal for scientific judgments, – the only sensible way of judging alternatives to current science is on scientific grounds. If such alternatives are to be adequate, they must preserve those features of our current scientific theories that are of value to scientists. Many of these features may also be deemed valuable from some broader philosophical perspective. But if there is conflict between the verdicts of the scientist and the philosopher then it is those of the former that must take precedence.

1.2. *The Nominalist View of Theories*

The most definitive way to refute the Indispensability Thesis is to construct a theory (or a concatenation of theories) that is nominalistic and that is at least as good as the totality of our current scientific theories. This is indeed the strategy that is favored by most contemporary nominalist.[5] I shall argue, however, that the standard nominalist conception of the nature of scientific theories has led the bulk of these nominalist projects to be misdirected.

For most nominalists, a scientific theory is essentially a tool for making inferences about the concrete, physical world. When the nominalist looks for alternatives to current science she is concerned above all that this proof-theoretic power be preserved. If every physical fact which is accounted for in the original theory is accounted for in the alternative nominalistic theory, then the alternative is 'minimally adequate' from the nominalist's point of view. In other words, the nominalist is looking for a theory which has the same *physical consequences* as current science, but which does not quantify over mathematical objects.[6]

It is generally conceded – on both sides of the debate –, however, that more than mere duplication of physical consequences is required of a nominalistic theory for it to undermine the Indispensability Thesis. This is because a nominalistic theory which fulfills this condition can trivially be generated using a couple of logical tricks. Let S be current science, viewed in the Quinean manner as a single, monolithic theory. Let T be the *nominalistic restriction* of S, in other words the set of nominalistically statable consequences of T. Then T has (by definition) the same physical consequences as S, and T does not quantify over mathematical objects. If required, T can also be axiomatized using Craig's Theorem, providing that there is a systematic way of distinguishing the nominalistically acceptable vocabulary of T. The existence of this sort of 'alternative' has struck nominalists and platonists alike as philosophically irrelevant to the status of the Indispensability Thesis, principally because the collection of axioms of T will lack coherence and unity. To circumvent this sort of logical trickery, it is normally stipulated that the nominalistic alternatives in question be "reasonably attractive".[7] I have argued, however, (at the end of Chapter I) that the inclusion of this extra condition – even if it can be made precise – is not enough to establish that the nominalistic alternative theory is relevant to the assessment of the Indispensability Thesis. The only way in which this Thesis can be undermined is by showing that there are nominalistic alternatives to current scientific theories that are *at least as good* – judged on scientific grounds – as these theories. My claim is that the features of theories that are scientifically valuable go well beyond their use as tools for proving physical facts.

This view of scientific theories as tools for proving physical facts is an example of what Lakatos called the 'deductivist caricature' of science.[8]

with most caricatures, there may be an element of truth to it. But by focusing solely on this aspect of scientific theories, the nominalist ignores the many other roles which theories play in the context of scientific practice. These functions include the discovery of new results and the development of new theories, and there is no reason to think that such functions will automatically be preserved by the nominalist's alternative theories. By taking a narrow view of scientific theories, the nominalist ends up with a correspondingly narrow view of indispensability according to which the indispensability of mathematics is exhausted by its indispensability as a tool for deriving physical results. Looking more closely at actual scientific practice reveals several other ways in which theories are used in science, and this in turn makes *indispensability* for science a correspondingly richer notion.

Even the arch-nominalist Hartry Field has had occasion to point to this multifaceted aspect of indispensability. Field summarizes the thesis that mathematics is indispensable for science as the thesis that "we need to postulate [mathematical] entities in order to carry out inferences about the physical world and in order to do science".[9] In this passage Field implicitly draws a distinction between 'carrying out inferences about the physical world' and 'doing science', but without indicating that he takes there to be any important difference between these two activities. I hope to show that the practice of science has other important aspects which Field ignores, aspects which go beyond the mere derivation of physical results If this is the case – and if indispensability implies not just indispensability for *proving* physical claims but also indispensability for *doing* science – then the Indispensability Thesis is correspondingly more robust.

The philosophically narrow view of scientific theorizing which underpins contemporary debates over indispensability is especially ironic given the historical pedigree of the Indispensability Argument. For – as we have seen in Chapter 1 – the Indispensability Argument has its roots in Quinean naturalism, a philosophical stance that prides itself on deferring to actual scientific practice and refraining from external critiques of science from the point of view of 'first philosophy'. For Quine, the ontological disagreement between platonists and nominalists is at root a scientific disagreement; the issue is whether our scientifically best theories quantify over abstract objects. What started out with Quine as a debate ostensibly over the *scientific* merits of nominalistic alternatives has gradually evolved into a debate over their *philosophical* merits. Not only this, but the range of features of scientific practice that are deemed to be relevant to this philosophical debate is remarkably narrow.

This shifting of the indispensability debate away from its naturalistic roots has provoked the complaint from certain quarters that the reconstructive programs of contemporary nominalism are simply *irrelevant* to the ontological debate. The nominalist strives to produce alternative mathematics-free theories which are at least as good as our best current theories. But from what perspective are these alternatives to be judged? If the claim is that they are at least as good from the perspective of the working scientist, then we can cut through the debate simply by presenting the nominalist alternatives to scientists and waiting to see if they take them up. Nominalists themselves concede that this would be highly unlikely. If the claim is that the alternatives are at least as good from some other perspective – for example that of the parsimoniously (or nominalistically) inclined philosopher – then the onus is on the nominalist to explain why *this* claim should undermine our confidence in the indispensability of mathematics for science.[10]

I find the basic thrust of this argument very compelling, indeed one aim of this chapter is to analyze in more detail some of the scientifically important features of theories that nominalistic reconstructions have failed to reproduce. To make a forceful case for nominalism – against a broadly naturalistic background – the alternative theories on offer must be judged from the perspective of science, not 'first philosophy'. Whether the actual judgments of working scientists should be taken as incorrigible is a separate issue. It may well be that extraneous social factors have systematic effects on scientists' assessments of alternative theories. Factors such as institutional inertia, epistemological conservativeness, and the costs of 'retooling' will tend to skew judgments in favor of established theories. Also, the apparent simplicity and elegance of an alternative theory may be diminished when considered from the viewpoint of the established theoretical paradigms. These issues may cast doubt on the reliability of simply 'reading off' claims about theory assessment from the behavior of scientists, but the essential point remains that it is the *scientific* merits of alternative theories that make them relevant to the ontological debate.

2. INDISPENSABILITY AND DISCOVERY

2.1. *Mathematics as a Tool for Discovery*

Philosophers who discuss provability tend to think of the concept in purely logical terms, as a deductive relationship between a theory and a sentence. In the context of the indispensability debate, however, this point of view obscures a pragmatic distinction between two separate ways in which proof can function. One function of proof is to verify or justify results which are already known; call this *proof-verification.* An example of proof-verification is the demonstration that Kepler's Laws for the motion of planets around the sun follow from Newton's Laws of Motion. Proofs of this sort help to organize and unify existing bodies of results, even where no new results are derived.[11] A second function of proof is to discover new (i.e. previously unknown) results; call this *proof-discovery.* An example of proof-discovery is Chandresekhar's proof, from the General Theory of Relativity, that any star greater than 1.4 times the mass of the Sun will eventually collapse under the force of its own gravity to form a neutron star. This involved the derivation of a previously unknown result from a pre-existing body of theory.[12]

In general the resources required for proof-discovery exceed the resources required for proof-verification. This point is familiar to anyone who has ever taken a mathematics or physics test. Questions of the form 'show that x' or 'verify that y' are typically easier than open-ended questions where the answer is not given in advance. If you know what you are aiming for then constructing the intermediate chain of reasoning is much more straightforward.

The science of cryptology provides a vivid illustration of how this discrepancy in resources can be exploited. The basic aim in designing a coding system is to make encoding a message as easy as possible, and to make decoding it as difficult as possible, for someone who does not know the secret key.[13] Here encoding is analogous to verification, and decoding is analogous to discovery. One popular current technique – known as the RSA system – is based on the multiplication and factorization of pairs of large primes. The RSA system is a 'public key' system, so-called because the method of encoding messages is made public so that anyone can in principle send a message. The public key consists of some large number, k, which is the product of two large primes. Encoding a message involves multiplying the digitized message by k, which is a comparatively trivial computational task, and then performing some further operations on the result. Decoding a message, however, involves factoring k into its two prime factors. This is a much more complex computational task (since the prime factors of k have not been made public); for instance the factorization of a 100-digit product would take decades at current computing speed.[14]

How does this asymmetry connect up with the issue of indispensability? The point is that if the resources needed for proof-discovery exceed those required for proof-verification then mathematics might be dispensable for the latter task without being dispensable for the former one. It might turn out that our mathematics-free theories enable us to prove all the nominalistic results which we could prove before, but that they do not enable us to prove as many (or any) new results. By concentrating on the context of verification and ignoring the context of discovery, the nominalist ends up operating with an unnaturally narrow conception of indispensability. Even if mathematics can be shown to be dispensable for proof-verification, this does not imply that it is dispensable in the broader sense which includes the role which mathematics plays in the discovery of new scientific results.[15]

2.2. *Mathematics as a Tool for the Development of Science*

It is clear from even the most cursory examination of science over the last three centuries that mathematics has been inextricably bound up not only with the discovery of new results but also with the development of new scientific theories. Moreover there has been a constant interplay between developments in physical theory and developments in mathematical formalism. Sometimes the physics suggests mathematical innovations, as in the case of Fourier analysis.[16] But more often it is the mathematical formalism which suggests the development of new physical theories. One well-known example is the application of group theory to particle physics which allowed the prediction of the existence of whole families of hitherto unobserved subatomic particles.[17,18]

This contribution of mathematics to the development of science over time is ignored in the indispensability debate since attention is focused on a temporal cross-section of science. This time-slice of current science is examined, and nominalistic alternatives to it are suggested. But though these alternatives may mimic the ‘static’ features of current theories, they will not necessarily preserve those features which are crucial to the ongoing development of science. John Burgess puts the point nicely when he complains that if scientists were to put on “nominalistic blinders” then many of the potential avenues for future development may be closed off.

> [T]he physicist who puts on nominalistic blinders may be unable to see certain potentially important paths for the development of science. ... [T]he danger I have in mind is that if science goes nominalistic today, that future theory *may simply never be discovered*.[19]

The argument is a straightforward one. It is all very well for the nominalist to piggy-back her reformulations on each new platonistic theory that is developed, but this does not suffice to show that the corresponding nominalistic theory could have been developed independently of its platonist ‘inspiration’. And if not, it is unclear to what extent the platonistic theories have been shown to be dispensable.

This dynamic aspect of theory development has not been completely ignored in the philosophical literature. Certain passages in Quine’s writings, for instance, indicate that he is sensitive to the way in which mathematical theories can function as heuristic aids for the development of science. The term he uses here is 'fecundity', and he lists it as one of five key virtues which a theory may possess, along with simplicity, strength, unity, and familiarity. Quine defines the fecundity of a theory to be the extent to which "successful further extensions of the theory are expedited", and this seems to point to a concern for how the current resources of a theory may impact its future developmenk.[20] Gödel is similarly sensitive to these issues, and to the role of what he terms 'fruitfulness'. In the 'proto-indispensability argument' articulated in his well-known paper, 'What Is Cantor's Continuum Problem?', Gödel writes,

> [B]esides mathematical intuition, there exists another (though only probable) criterion of the truth of mathematical axioms, namely their fruitfulness in mathematics and, one may add, possibly also in physics.[21]

Fruitfulness – or fecundity – is an important positive feature of scientific theories, and mathematical apparatus can often contribute to the fruitfulness of a physical theory. What happens to a mathematical theory if it is not fruitful for the development of physics? A mathematical theory that is persistently unfruitful may eventually be discarded by working scientists and the physical theory reformulated using different mathematical apparatus.[22] Even when this happens – and it is by no means inevitable – the mathematical theory may well continue to be used and studied by mathematicians for its intrinsic mathematical interest, and its potential to cast light on other areas of mathematics. In some cases the reformulation of a mathematical theory for the purposes of physics is little more than the streamlining of the mathematical notation.[23] In other cases it involves a wholesale replacement of one mathematical theory by another. A particularly clear example of this latter sort of case concerns the 19th-century theory of quaternions and its eventual replacement by modern vector analysis.

2.3. *Case Study: Quaternions*

Quaternions are a noncommutative algebraic number-system developed by William Rowan Hamilton in 1843 while he was searching for 3-dimensional extensions of complex numbers. It had already been established, early in the 19th Century, that complex numbers could be represented graphically using Cartesian coordinates by taking the x-axis as the real component and the y-axis as the imaginary component, in which case multiplication by i corresponds to a rotation of $\pi/2$ about the origin. Hamilton was eager to find an extension of complex numbers, of the form $x + yi + zj$, which could be graphically represented as points in 3-dimensional space and which would preserve most of the important algebraic properties of complex numbers. Since 3 is the number of dimensions of actual physical space, Hamilton hoped that such a system would serve as a useful framework for the formulation of physical theories. Hamilton was unsuccessful in his quest (indeed it was later proved that no such 3-dimensional extension is possible) but he did discover a 4-dimensional extension of complex numbers, of the form $w + xi + yj + zk$; he christened these numbers 'quaternions'. Each quaternion has one scalar and three vector components, and there are simple rules for the addition and multiplication of quaternions. The key property of quaternion multiplication is that it is non-commutative; for example, $ij = k$, but $ji = -k$.

In tracing the subsequent fortunes of quaternions I shall distinguish between the role of quaternions in mathematics and their role in science (more specifically, in physics). From a mathematical point of view, quaternions provided impetus to the early development of algebra as an autonomous subdiscipline within mathematics. Hamilton's discovery of quaternions was of considerable mathematical importance because he was one of the first to identify commutativity as a distinct property, and to realize that there could be mathematically legitimate systems which give it up.[24] As the study of algebraic systems developed through the latter part of the 19th Century, it became apparent that quaternions are just one example of a more general family of algebraic structures known as division algebra.[25] A division algebra is an algebra with two operations, addition and multiplication, in which there are no non-zero divisors of zero. It turns out that there are only four division algebras[26] – real numbers, complex numbers, quaternions, and octonions (which are 8-dimensional extensions of quaternions).

	No. of components	Associative?	Commutative?
Reals	1	yes	yes
Complexes	2	yes	yes
Quaternions	4	yes	no
Octonions	8	no	no

A second way in which quaternions contributed to developments within mathematics was as a stepping-stone on the path to the modern system of vectors and scalars. Quaternions have both a vector and a scalar component; however, in the case of 'pure' quaternions where the scalar part is zero, the structural similarity with modern vectors is much closer. The main difference is that quaternions have a single multiplication rule. The result of multiplying together two pure quaternions is as follows:

$$\begin{aligned}(\mathrm{x}i + \mathrm{y}j + \mathrm{z}k)(\mathrm{x}'i + \mathrm{y}'j + \mathrm{z}'k) &= -\mathrm{xx}' + \mathrm{xy}'k + \mathrm{xz}'j - \mathrm{yx}'k \\ &\quad -\mathrm{yy}' + \mathrm{yz}'i - \mathrm{zx}'j - \mathrm{zy}'i - \mathrm{zz}' \\ &= -(\mathrm{xx}' + \mathrm{yy}' + \mathrm{zz}') + [(\mathrm{yz}' - \mathrm{zy}')i \\ &\quad + (\mathrm{xz}' - \mathrm{zx}')j + (\mathrm{xy}' - \mathrm{yx}')k].\end{aligned}$$

In modern terminology, the result is equal to the sum of the vector product and the negation of the scalar product of the two quaternions. Vector algebra differs from quaternion algebra in having separate operators for the scalar product and the vector product. Yet from a broader perspective – and in comparison to Cartesian geometry – both quaternions and vectors are examples of 'vectorial' systems.[27]

To summarize, the significance of quaternions within mathematics has principally been as a stepping-stone – to algebraic systems on the one hand and to vectors on the other. What about the role of quaternions in science? Hamilton hoped from the beginning that quaternions might form a suitable mathematical model for the analysis of force and motion in 3 dimensions, and that this would in turn lead to fruitful applications in various areas of physics. Indeed it seems that even as he was experimenting with different rules for quaternion multiplication, Hamilton always had a geometrical interpretation in the back of his mind. Hamilton made it a requirement, for example, that any adequate definition of quaternion multiplication satisfy the following 'law of moduli';

(1) $\mathrm{N}(\mathbf{p} \otimes \mathbf{q}) = \mathrm{N}(\mathbf{p})\mathrm{N}(\mathbf{q})$,

where $\mathrm{N}(\mathbf{q})$ is the norm of $\mathbf{q}$ (i.e., the product of q and its conjugate). This law guarantees that every multiplication operation has an inverse. Since Hamilton was hoping that unit quaternion multiplication would correspond to three-dimensional rotation, and since it is always geometrically possible to undo a rotation, the law of moduli follows naturally from Hamilton's quest for a geometrical interpretation for quaternions. It is the fact that this law of moduli holds that makes quaternions a division algebra, since it implies that there are no non-zero divisors of zero.[28] Here we see an interesting example of the potential physical applications of a mathematical theory influencing the logical structure of the theory. It is no accident that quaternions are a division algebra, since Hamilton in effect made this a requirement as he set out to construct them.[29]

Despite their potential for physical applications, quaternions never really established themselves as an indispensable tool for physics, and by the end of the 19th Century interest in quaternions had more-or-less died out among the mainstream of the scientific community. The reasons for their demise were probably as much sociological as mathematical, yet there are several mathematically significant factors that can be pointed to. Part of the problem was that Hamilton – and a number of other early supporters of quaternions – concentrated on developing the mathematical properties of quaternions at the expense of investigating their physical applications. Moreover, Hamilton's choice of geometrical interpretation for quaternions seems to have actively hindered their application to physics. Hamilton identified the imaginary part of a quaternion with a vector. The problem was that he also identified the rotation of a vector with its multiplication by a pure quaternion (i.e. a quaternion whose scalar part is zero). Hence the result of multiplying a vector by a pure quaternion must result in another pure quaternion (since rotating a vector produces another vector). This is fine for cases where a vector is rotated about an axis perpendicular to its direction. However, in the case of conical rotations, the result will not in general be another pure quaternion.[30]

Quaternions were initially presented to physicists as an alternative to Cartesian coordinates. Supporters of quaternions, such as the mathematician and physicist William Tait, argued that they revealed the "physical meaning" of equations more "transparently", and that they avoided "the artificiality of Cartesian co-ordinates". Directly comparing the two ways of formulating physical theories, it is hard to avoid the conclusion that quaternions are syntactically simpler and more elegant. Such advantages were more than outweighed, however, by the unfamiliarity of most physicists with the new (and often arcane) notation of quaternions, and their consequent reluctance to rewrite existing physical theories in quaternionic language. Perhaps given time these pragmatic factors would have receded in importance. Before this could happen, however, quaternions were superseded by a third alternative, namely vectors.

I have already discussed how quaternions laid the mathematical groundwork for the development of the modern system of vectors. As vectorial systems rose to prominence at the end of the 19th Century it became clear that they were in many ways better suited than quaternions to the formulation and development of physical theories. Vectors have several advantages over quaternions. First, the fact that vector multiplication is separated into the scalar (or 'dot') product and the vector (or 'cross') product avoids the unwieldiness of quaternion multiplication. Second, the separate treatment of vectors and scalars often helps in the formulation of specific physical theories. One example is Maxwell's theory of electromagnetism, for which the separation of vector and scalar parts facilitates the independent representation of electric and magnetic effects.[31] Using Cartesian components, Maxwell's theory can be formulated using eight equations; using vectors, this number can be reduced to four.[32] Third, the geometrical interpretation of vectors fitted more closely with physical intuition. One problem with quaternions was that the physical interpretation of the (fourth) scalar component of a quaternion was obscure. It was also unclear why the square of a quaternion with no scalar component should be negative, as the formalism required. The net result of these various factors, combined with the respective states of mathematics and physics at the end of the 19th Century, was that vectors quickly superseded quaternions as the main alternative to Cartesian coordinates. Over time the new vectorial methods gradually gained converts among the scientific community, and by the early decades of this century it is fair to say that vector algebra had established itself as the preferred language of physics.

What does the status of quaternions look like from the perspective of the Indispensability Argument? The first thing to say is that the whole story of the introduction of quaternions, their decline, and their eventual replacement by vectors is a phenomenon which cannot be analyzed in terms of deductive indispensability. This is because both quaternions and vectors are *dispensable* from a deductive point of view, for they are no stronger deductively than the Cartesian geometrical methods in use before the 19th Century. This is a point that is stressed by Michael Crowe in his (excellent) book, *A History of Vector Analysis.* Crowe writes:

> [M]athematically anything that could be done by the application of quaternions in geometry and physics could also be done with the Cartesian methods, though usually by longer processes.[33]

This point applies equally well to vectors. Both quaternions and vectors can be identified with sets of Cartesian coordinates, and suitably modified analogs of operations such as multiplication and differentiation can be defined for them. The resulting calculations may be cumbersome and inelegant, but it can be shown that the surrogate system is proof-theoretically equivalent to the original. Hence anything provable in a quaternion or vector system is also provable in Cartesian geometry.

As with the case of infinitesimals, what is needed for a satisfactory analysis of quaternions is a concept of indispensability that goes beyond sheer deductive power to take account of the way in which mathematical theories can contribute to the discovery and development of new scientific theories. In the latter part of the 19th Century both quaternions and vectors were used to rewrite various existing scientific theories. In this respect vectors turned out to have decisive advantages over quaternions. Though neither theory was *deductively* indispensable, it seems clear that by the early decades of this century vectors were indispensable (in the broader sense discussed above) for physics and thus for science as a whole.

(A) *Twentieth-Century Physics*

Having been eclipsed by vectors as the preferred tool for physics, and with their intrinsic mathematical interest being marginal at best the prospects for quaternions at the turn of the century looked distinctly gloomy. However, the radical and unpredicted changes that shook physics in the early decades of the 20th Century have led – somewhat ironically – to a recent resurgence of interest in quaternions as a tool for physics. Quaternions have turned out to be peculiarly appropriate for two of the major theories that arose out of Einstein's groundbreaking work, namely special relativity and quantum mechanics. In the context of these theories, the mathematical features that distinguish quaternions from vectors - their four-dimensionality and their non-commutativity – turn out to be virtues rather than liabilities. An examination of this potential new role for quaternions serves as an interesting postscript to the debate over their possible indispensability for physics.

(B) *Special Relativity*

The first distinctive feature of quaternions I want to examine is their four-dimensionality. As early as 1844, Hamilton had himself wondered whether the vector part of a quaternion could represent the three spatial dimensions and the scalar part the time dimension. This idea resurfaced with the development of special relativity, which was itself based on the four-dimensional structure of Minkowski space-time.[34] Unfortunately it is not possible simply to represent space-time points as quaternions, because this does not give the correct metric for space-time. In special relativity the separation between two space-time points is equal to their spatial separation minus their temporal separation. The norm of a (real-valued) quaternion, however, is equal to the magnitude of the vector part plus the magnitude of the scalar part [if $q = w + xi + yj + zk$, then $\mathbf{N}(q) = (qq^*)^{1/2} = w^2 + x^2 + y^2 + z^2$]. The easiest way around this difficulty is to represent space-time points using quaternions with complex – rather than real – coefficients. The use of complexified quaternions (also called 'biquaternions') has a couple of drawbacks. The formulation is slightly less compact, since each space-time point is represented by eight numbers, rather than four; also biquaternions do not form a division algebra, since the complex coefficients permit the presence of non-zero divisors of zero.[35]

Quaternions provide a tool for the elegant reformulation of special relativity, and their use in this context has been developed in various different directions. As with earlier physical theories, however, it does not appear that quaternionic reformulations of special relativity are deductively any stronger – they do not permit the derivation of any substantive new results. The difference between special relativity and previous theories, however, is that the [3 + 1] dimensionality of space-time is peculiarly appropriate for modeling using the 3 vector and 1 scalar components of quaternions. As one contemporary physicist remarked, quaternions provide a valuable tool for those "prepared to exploit the accident of having been born in space-time".[36]

(C) *Quantum Mechanics*

Recent applications of quaternions in quantum mechanics have exploited a second distinctive property of quaternions, namely non-commutativity. These quaternionic approaches are of particular interest because they have led to formulations of quantum mechanics which – unlike the case of special relativity – have structural and physical implications that go beyond those of the standard theories. In other words, the role of quaternions in quantum mechanics seems to go beyond that of mere reformulation.

Birkhoff and von Neumann – in a 1936 paper – were the first to point out the possibility of using quaternions as a basis for quantum mechanics. One of the basic tenets of quantum mechanics is the superposition principle for probability amplitudes; this implies that probabilities obey the 'law of moduli', and hence that they form a division algebra. This implies that quantum mechanics can in principle be represented as a vector space over any one of the four division algebras (reals, complexes, quaternions, or octonions).[37,38] However, if – as is generally assumed – probability amplitudes are associative, then this rules out the possibility of using octonions.

The Birkhoff–von Neumann result establishes only that it is *possible* to base quantum mechanics on quaternions and not that it is *advantageous* to do so. In their 1962 survey paper, Finkelstein et al. write;

> We can thus formulate the following precise problem: Which of the three possibilities for the representation of general quantum mechanics is the one most suitable for the description of the actual physical world?[39]

In fact it turns out – for reasons too technical to elaborate here – that real numbers are also flawed as a potential basis for quantum mechanics because the formalism requires the existence of a distinct conjugate pair for each state of the system.[40] We are left, then, with just two possible candidates: complex numbers and quaternions. Standard formulations of quantum mechanics are invariably based on complex numbers. There is a general feeling among physicists that complex numbers can do everything that might be required, and this – combined with their familiarity – makes physicists reluctant to move away from formalizations based on them. In recent years, however, some physicists have begun to question whether complex numbers provide the *optimal* formulation of quantum mechanics in all cases, and this has led them to explore quaternions as an alternative.

It turns out that the property of non-commutativity gives quaternionic formulations of quantum mechanics some very interesting and distinctive properties. Perhaps the most important of these pertains to tensor products. A tensor product is a multiplication of wave functions from different systems, and is used – for example – to calculate interactions between particles of different kinds. Calculating a tensor product requires specifying a common coordinate basis for the two systems, hence tensor products are a so-called *coordinate-dependent* method. Such methods are acceptable only if it can be shown that the result is not dependent on the particular choice of coordinates. If quaternions are used however, then their non-commutativity blocks any such independence result. Hence tensor products are not acceptable in quaternionic quantum mechanics.

Another way of expressing this result is in terms of *complementarity*. Two observable properties are said to be complementary if it is impossible for both properties to be simultaneously determined. [One example is position and momentum, whose complementarity is expressed in the Heisenberg Uncertainty Principle.] In quaternionic quantum mechanics, given any two systems there is a complementarity between at least some of the properties of the systems. In a sense, then, the properties of quaternions rule out the possibility of any two systems being truly independent.[41] This is a striking example of the way in which significant consequences about the nature of the physical world may flow from the choice of the underlying mathematical formalism. In this case it is their distinctive feature of non-commutativity which allows quaternions to play this substantive role.

Results of the above sort have led some physicists to suggest that quaternions might provide a promising framework for the formulation of Grand Unification Theories (or GUTs).[42] In his 1996 paper on this topic, De Leo conjectures that "a successful unification of the fundamental forces will require a generalization beyond the complex".[43] Investigations into using quaternions for GUTs have shown other ways in which the distinctive properties of quaternions have concrete physical implications. De Leo discusses the example of determining which group best represents the quark colors. In the standard model there are three quark colors: red, green, and blue. One possibility that is consistent with the experimental evidence is that the quark color group is a quaternionic group. Another possibility is that the quark color group is a complex group. The interesting point about the quaternionic group is that its adoption implies the existence of a *fourth* quark color, call it white, which is not implied by the choice of a complex group. De Leo writes, "the existence of white quarks is probably esoteric, but not *a priori wrong*". The choice of a quaternionic group as the mathematical basis for quark color makes a prediction which can (at least in principle) be experimentally tested. For the supporter of the Indispensability Argument, the discovery of white quarks would provide good grounds for believing in the existence of quaternions.

I conclude that quantum mechanics provides for the first time the real possibility that quaternions might be deductively indispensable for our best physical theories. If it turns out that quaternionic quantum mechanics becomes widely accepted by physicists, and maybe even confirmed by future experimental evidence, it is likely that the distinctive mathematical properties of quaternions will make them deductively indispensable. Their role in such a theory will not merely be one of streamlining and reformulation, but also to allow the deduction of physical consequences that could not otherwise be derived.

It might be objected that, even in the context of quantum mechanics, quaternions are deductively dispensable because they need not be taken as primitive. For example, quaternions can be defined as suitable sets of quadruples from $\mathfrak{R}$.[44] This situation is different from that of quaternions in 19th-century physics, however, because in that case the alternative – and heuristically preferable – vector-based theory contained no objects which duplicate the mathematical properties of quaternions. In the quantum mechanics case, by contrast, eliminating quaternions in favor of quadruples of real numbers produces a theory which (by design) has the same mathematical structure. In claiming that quaternions may turn out to be deductively indispensable for quantum mechanics, what I mean is that quantum mechanics may turn out to require for its optimal formulation mathematical objects with the structural properties of quaternions.

For a mathematical theory actually to be successfully dispensed with is the best evidence that it is dispens*able*. The strength of the Indispensability Argument derives in large part from the unlikeliness that mathematics as a whole will ever actually be dispensed with by scientists. The case of quaternions is encouraging to the nominalist because it provides a rare example of a piece of mathematics which was dispensed with by working scientists and which also stopped being studied by mathematicians. Of course this case provides no direct support for the nominalist position because quaternions were dispensed with in favor of another *mathematical* theory – namely vector analysis. In addition, recent developments in quantum mechanics suggest that quaternions may be indispensable for science after all.

3. THEORY CHOICE AND ONTOLOGY

3.1. *What Features of Theories are Relevant to Ontology?*

There are a number of ways in which the nominalist might try to counter the arguments I have given. He might – for example – concede that he is using the term 'indispensable' in a narrow sense which ignores many interesting ways in which mathematics functions within science as a tool for discovery and development, but claim that it is only this narrow sense of indispensability which is *relevant* to the issue of ontological commitment

The first thing to say about this response is that it is dubious if offered from the perspective of the scientific naturalist framework within which we are investigating the Indispensability Argument. The nominalist concedes that there are aspects of the use of mathematics in science which are not captured by his narrow concept of indispensability, but dismisses these aspects as irrelevant to his philosophical enterprise. But when the philosopher decides to pronounce on which are and which are not the salient aspects of scientific practice, he is setting off down the start of a very slippery slope. Naturalism involves deferring to scientific practice *in toto*, rather than deferring merely to those aspects of scientific practice which have been deemed philosophically relevant. It may be possible to develop a view based on this sort of 'partial naturalism'. But any such view is going to be vulnerable to charges of arbitrariness and circularity concerning its choice of which features of science to focus on.

If the uses of mathematics in science which I have highlighted are genuine uses, then I think that the onus is on the nominalist either to take account of them or to explain his justification for ignoring them I am doubtful whether any non-question-begging justification can be found. But I want to examine one line of argument that aims to provide grounds for ignoring the discovery-based role of mathematics.

3.2. *Pragmatic versus Theoretical Features of Theories*

In deciding what to believe we are concerned with what is entailed by our best available theories, and this presupposes that we have some sort of systematic way of comparing rival theories. One argument for taking a narrow view of indispensability is that when comparing theories we should distinguish between *theoretical* virtues and *merely pragmatic virtues.* The theoretical virtues of a theory are the only ones which are important when it comes to determining ontological commitment. These theoretical virtues include unity, simplicity, explanatory power, deductive strength, and economy of postulates and primitive terms. The theory which achieves the best balance of theoretical virtues is the one we should use in assessing our ontological commitments. If our 'theoretically best' theory entails the existence of mathematical objects, then we have good grounds for being mathematical realists. Otherwise, not. Theories may also have other sorts of virtues – so-called 'pragmatic' virtues – which are relevant to the practical business of choosing theories. These include features such as manipulative elegance, fruitfulness, familiarity, and perspicuity. These pragmatic features are important when it comes to using theories on a day to day basis, but they should not be taken into account when it comes to deciding which theory to *believe.* Theoretical virtues make a theory a better candidate for truth; pragmatic virtues merely make a theory more useful.[45]

What can be said in support of this claim that the pragmatic aspects of theory choice are irrelevant to ontological decision-making? This view is an initially attractive one – I think – because it is tempting to think of pragmatic features of theories as somehow 'subjective' and dependent on us and our cognitive and computational powers. Theoretical features, on the other hand, seem more 'objective' and therefore more real.[46] By focusing on the theoretical virtues of theories we are transcending our own cognitive limitations. The objectively 'best' theory is the best guide we have to the real furniture of the universe.

I think that something like this chain of reasoning lies behind the logical positivist's dismissal of pragmatic features as philosophically irrelevant. In his influential book, *Experience and Prediction*, Hans Reichenbach distinguishes between "context of discovery" and "context of justification", and claims that "epistemology is only occupied in constructing the context of justification".[47] He argues that the investigation of the process of discovery is properly the task of psychology, not epistemology, since it involves the detailed examination of actual mental processes. What is or is not required for discovery is dependent on the contingent details of our psychological and sensory make-up. This downplaying of the pragmatic aspects of theory development has been part of the enduring legacy of logical positivism, and it is this attitude which has persisted in contemporary discussions of indispensability.

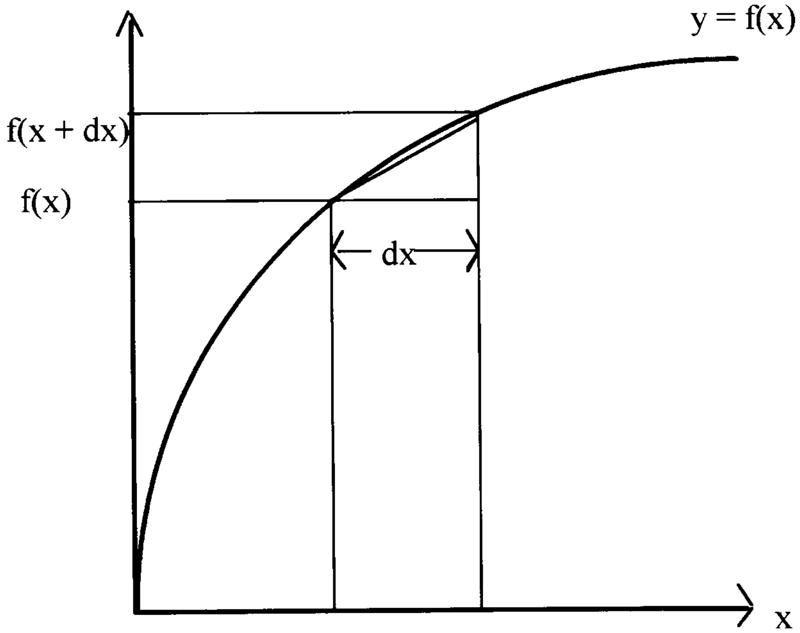
I have said that this view is initially attractive, but I do not think that it can be correct. For it depends on there being some sort of principled division between the theoretical and the pragmatic features of theories, and it seems unlikely that any such division can be convincingly made. Take the feature of simplicity, for example. We tend to prefer simpler theories over more complex ones. But is this merely a quirk of human psychology or does it reflect the deeper fact that simpler theories are more likely to be true? Another example is explanatory power. Van Fraassen argues – in support of the view he calls 'constructive empiricism' – that explanatory power is a pragmatic rather than a theoretical virtue of theories; the fact that people tend to prefer theories with explanatory depth, and make inferences to the 'best explanation', provides no grounds for thinking that there is an objective link between explanation and truth.[48] In general, for any feature which is claimed to be theoretical, 'objective', and a reliable guide to truth, it is possible to make the skeptical counter-claim that it is merely a reflection of human psychological preferences.

Quine's solution to this problem is – in characteristic style – to reject the distinction between pragmatic and theoretical features as confused. For Quine, *all* of our criteria of theory choice are to some degree pragmatic. Science proceeds using pragmatic criteria, and the theories of science are the best we have. Hence it is wrong to dismiss pragmatic criteria as ontologically irrelevant on philosophical grounds, since to do so is to allow 'first philosophy' to trump the internal standards of science. Quine's stress on the legitimacy of pragmatic considerations has gradually disappeared from current debates over indispensability, and this is one reason why contemporary nominalists have been able to shift the focus to purely deductive aspects of the role of abstract objects in science. Once again, I think, the burden of proof lies with those who wish to draw a philosophically significant distinction between the pragmatic and the theoretical to show how and why such a distinction ought to be drawn.

4. CASE STUDY – A BRIEF HISTORY OF INFINITESIMALS

The above points will become clearer if we examine how the twin roles of mathematics for justification and for discovery play out in an actual historical example. I shall focus on the theory of infinitesimals. This is an example which Quine discusses in some detail in the final chapter of *Word and Object*, and his discussion provides a paradigm case of what I referred to as the "narrow" view of the role of mathematics.[49] I shall use Quine's analysis as a starting point, and then go on to argue that it misrepresents the historical realities of the dispute over infinitesimal methods.

Infinitesimals rose to central prominence in mathematics and physics with the invention of the calculus in the late 17th Century. For the first 150 years of its existence the calculus was based around the postulation and manipulation of infinitesimals – magnitudes smaller than any given finite magnitude yet larger than zero – one legacy of which is the '*dx*' notation still in use today. To calculate the slope of a graph at a point, x, the 18th-century mathematician would take the point (x + *dx*) and consider the slope of the graph between these two points.



For example, iff(x) = x^2 then

$$\text{(1)} \qquad S = \frac{(x + dx)^2 - x^2}{dx}$$

$$\text{(2)} \qquad = \frac{x^2 + 2xdx + dx^2 - x^2}{dx}$$

$$\text{(3)} \qquad = 2x + dx$$

$$\text{(4)} \qquad = 2x.$$

The main conceptual problem with this method is that the infinitesimal is treated as being both zero and non-zero at different points in the calculation. At the point where the gradient of the tangent is calculated [steps (1) and (2)], *dx* appears as a divisor and must therefore be non-zero. Subsequently [step (3) to step (4)] a *dx* term appears and is neglected by being put equal to zero.

Eighteenth-century mathematicians who used the calculus knew that there was something arbitrary and unrigorous about their treatment of infinitesimals.[50] But the method worked, and they had no alternative method available. This (unstable) situation persisted until the early 19th Century, when Cauchy and then Weierstrass developed a version of the calculus based on the concept of limit. They showed how infinitesimals could be eliminated in favor of definitions involving the limits of ordinary finite magnitude.[51] Quine argues that this latter theory is superior because it is simpler from a *global* perspective;

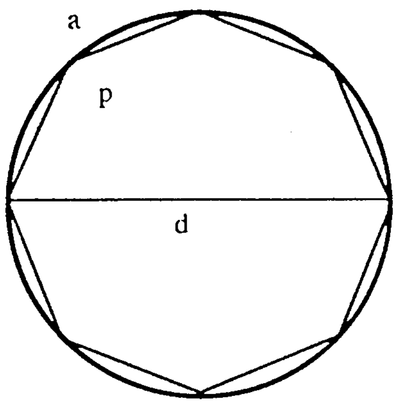
> When we paraphrase ... in the Weierstrassian spirit ..., we are merely switching from a theory that is conveniently simple in a short view and complex in a long view to a theory of the opposite character. Since the latter, if either, is the one to count as true, the former gets the inferior rating of convenient myth ....[52]

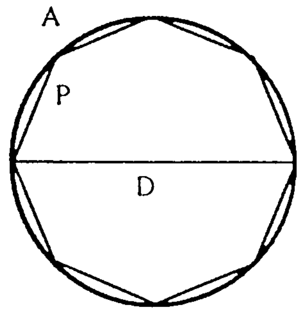
The delta-epsilon notation of Weierstrass's theory is cumbersome and considerably less intuitive than the infinitesimal approach. Its advantage is that it provides a rigorous and logically perspicuous foundation for the calculus, and this allows the theory to be embedded in other more comprehensive theories.

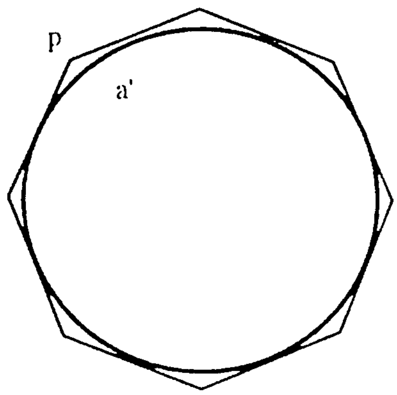
Quine draws from this historical summary two conclusions concerning the epistemological status of infinitesimals. Firstly, before Cauchy and Weierstrass, infinitesimals were (deductively) indispensable for science since the best theories of mechanics and physics required the postulation of infinitesimals. Hence it was rational – according to Quine – for 18th-century mathematicians to believe in the existence of infinitesimals. Secondly, the development of a theory of limits makes infinitesimals (deductively) dispensable, since everything that was provable using infinitesimals can now be proved using limits. Hence there is no reason for us today to believe in the existence of infinitesimals.

I shall argue that both of Quine's central claims are problematic. The root of the problem is his narrow focus on mathematics as a deductive tool, and his equating indispensability for science with *deductive* indispensability. These problems are compounded by the somewhat selective and partial nature of Quine's historical narrative.

Quine's first problematic claim is that infinitesimals were deductively indispensable prior to the development of the Cauchy-Weierstrass theory of limits. In claiming this, Quine seems unaware that most (if not all) of the results obtainable using infinitesimals could also be obtained by an alternative method, the so-called *Method of Exhaustion*. This is a method of proving results about the areas of curved geometrical figures by enclosing the figure in an infinite succession of regular polygons.[53] The area to be calculated is 'trapped' between an inscribed and a circumscribed polygon, each of whose areas can be directly calculated. As the number of sides of the enclosing polygons increases the area is more and more closely approximated. In this way the difference between the curved area and the area of the polygon is 'exhausted'. The Method of Exhaustion was known to ancient Greek mathematicians; Euclid's *Elements* contain several proofs by exhaustion of propositions concerning the areas of circles, and Archimedes made extensive use of the method in his treatises on conics and parabolas. The diagrams below illustrate part of a proof by exhaustion of the hypothesis that the ratio of the area of two circles is equal to the ratio of the squares of their diameter.[54]







Let the areas of the circles be a and A, and let their diameters be d and D. If it is not the case that $a : A = d^2 : D^2$, then there is some other circle of area a' such that $a' : A = d^2 : D^2$. If a' is smaller than a, then it is possible to find a polygon of area p such that $a' < p < a$. We can then inscribe a similar polygon of area P inside the circle of area A. We know that $p : P = d^2 : D^2 = a' : A$. But since $p > a'$, then $P > A$, which contradicts the assumption that P is inscribed in the circle. If a' is larger than a, then an analogous *reductio* can be demonstrated using a circumscribed polygon. Hence $a : A = d^2 : D^2$, by double *reductio ad absurdam.*

As already mentioned, the roots of the Method of Exhaustion lie in ancient Greek geometry, and it was a method that was familiar to 17th- and 18th-century mathematicians. Moreover, given Quine's emphasis on deductive strength and logical rigor, the Method of Exhaustion was *superior* to infinitesimal methods. As we shall see, the Method of Exhaustion was considered to be at least as strong deductively as infinitesimal methods and considerably more rigorous. If this is correct, then the presence of the alternative Method of Exhaustion shows that infinitesimals were not deductively indispensable prior to Cauchy-Weierstrass, contrary to Quine's initial assertion.

Consider first the issue of deductive strength. Judging by contemporary accounts, 17th- and 18th-century mathematicians believed that the Method of Exhaustion was deductively just as strong as infinitesimal methods. Results that had been discovered using infinitesimals were often recast in the double-*reductio* form of a proof by exhaustion, and most mathematicians of the time believed that such a reformulation could always in principle be carried out. Indeed it was Leibniz's belief in the possibility of reformulating infinitesimal proofs as proofs by exhaustion that underpinned his treatment of infinitesimal magnitudes as useful fictions. Leibniz writes:

> There is no need to take the infinite in a rigorous way, but only the way in which one says in optics that the rays of the sun come from an infinitely distant point and are therefore taken to be parallel. ... For, in place of the infinite or infinitely small, one can take quantities as great or as small as one needs so that the error be less than the given error. So that one does not differ from Archimedes' style but for the expressions which in our method are more direct and more in accordance with the art of discovery.[55]

Leibniz never backed up this claim with any formal demonstration that such a prooftransformation is always possible.[56] However, this aspect of Leibniz's position (unlike his fictionalist attitude to infinitesimals) did appear to be widely shared by his contemporaries.[57] The most extended early attempt to derive all the results of the calculus by the Method of Exhaustion is to be found in McLaurin's 1742 *Treatise of Fluxions.* The thesis that any infinitesimal result can be established using an indirect proof by exhaustion implies that infinitesimals are dispensable from a deductive point of view.

Not only was the Method of Exhaustion considered to be deductively adequate, but also – unlike 17th-century infinitesimals – it fully satisfied contemporary standards of rigor. Indeed, Euclid – the very paradigm of rigor – had seen fit to include several proofs by exhaustion in his Elements.

This issue of rigor was the main reason why infinitesimal proofs were so often recast in the form of proofs by exhaustion.[58,59]

From a global point of view, then, the Method of Exhaustion was regarded as superior to the infinitesimal approach. The Method was deductively at least as strong as infinitesimal methods, and possessed the sort of rigor which infinitesimals conspicuously lacked. Use of the Method of Exhaustion allowed a watertight deductive proof to be given while avoiding the contradictions seemingly inherent in the 18th-century concept of infinitesimal. In Quinean terms, then, infinitesimals were already deductively dispensable before the innovations of Cauchy and Weierstrass, for there was a globally better theory available – the Method of Exhaustion – which did not postulate infinitesimal.[60]

This leads me on to my second main point, which concerns the difference between indispensability and *deductive* indispensability. Contemporary debates over indispensability proceed as if there is little or no difference between these two concepts. If this presumption is right, however, and if infinitesimals were already considered deductively dispensable prior to Cauchy-Weierstrass, then why were infinitesimals not dispensed with by 18th-century mathematicians? Wouldn't this have solved the 'crisis' in the foundations of the calculus at one fell swoop?

The answer is no, and the reason why infinitesimals could not be discarded was because – although they were not deductively indispensable – they were indispensable *for the discovery of new results.* The great problem with the Method of Exhaustion is that it is impractical to apply unless the result to be proved is known in advance. This is because a proof by exhaustion proceeds by means of a (two-part) *reductio*, thereby making it an indirect method of proof The logical structure of a proof by exhaustion involves a version of *tertium non datur*; it is shown that if the given area is greater than A then this leads to contradiction, and if it is less than A it leads to contradiction, hence the area must be equal to A. Unless the value of A is known (or somehow guessable) in advance, the Method of Exhaustion cannot be applied.[61] This is in contrast to infinitesimal proofs, which proceed in direct algebraic fashion from premise to conclusion. This difference is what Leibniz is pointing to when he writes, in the passage quoted previously, that infinitesimal methods are "more direct and more in accordance with the art of discovery" than the Method of Exhaustion. Indeed a large part of the motivation for the development of the infinitesimal calculus stemmed from mathematicians' dissatisfaction with the fact that the Method of Exhaustion failed to reflect the way in which results were actually discovered.[62]

Quine is therefore right that infinitesimals were indispensable to 18th-century mathematics, but wrong that they were *deductively* indispensable. The reason that infinitesimals were needed was for the purpose of discovery, not justification.[63] Quine's narrow view of indispensability cannot adequately capture this extra dimension, and this leads him to misrepresent the actual historical situation. The case of infinitesimals is instructive because it provides a particularly clear-cut illustration of the distinction between context of discovery and context of justification. Between the late 17th and early 19th Centuries, mathematicians had a fruitful but non-rigorous theory (infinitesimals), and a rigorous but unfruitful theory (the Method of Exhaustion), but they had no theory which was both rigorous *and* fruitful. The importance of the Cauchy-Weierstrass theory was that it led to methods that possessed both of these virtues. Any analysis of indispensability which does not recognize the interaction between these non-deductive aspects of mathematical theories – as Quine's analysis does not – is bound to end up with a distorted picture of the interactions between mathematics and science.

I mentioned earlier that I had objections also to Quine's second basic claim, that infinitesimals are no longer indispensable for science. The story of infinitesimals has been given a final twist with Abraham Robinson's development of nonstandard analysis in the 1960's, which showed how the infinitesimal calculus can be placed on a rigorous logical foundation using techniques from model theory.[64] Nonstandard analysis combines the rigor of the Method of Exhaustion with the perspicuity of infinitesimal methods to produce a mathematical system of considerable power and flexibility.[65] The pedagogical advantages of nonstandard analysis have been borne out by studies which show that students who are taught using nonstandard analysis learn calculus significantly more quickly than those taught using traditional delta-epsilon methods.[66] The main barrier to more widespread use of nonstandard analysis in mathematics teaching and research seems to be institutional inertia combined with a general lack of familiarity with the new techniques. Maybe, then, the best theory of the calculus does involve the postulation of infinitesimals, in which case infinitesimals would turn out to be indispensable for science after all. Here I am once again speaking of indispensability in a broader sense than just deductive indispensability. Even post-Robinson, there is no question that infinitesimals are deductively dispensable, indeed one of the key results of nonstandard analysis is a proof that nonstandard methods [i.e., methods involving the population of infinitesimals] constitute a conservative extension of classical mathematics.[67]

I want to conclude my discussion of infinitesimals by drawing attention to an interesting difference between their recent rehabilitation and the resurgence of interest in quaternions that was discussed earlier. Let us focus in each case on the stimulus for the reintroduction into science of the discarded theory. In the case of infinitesimals the stimulus came from the mathematical end; it was only once *mathematics* had evolved to the point where infinitesimals could be put on a rigorous logical footing that interest in their use was rekindled. In the case of quaternions the stimulus came from the scientific end; it was only with development of new *scientific* theories – in particular special relativity and quantum mechanics – that fresh attempts were made to use quaternions in physics.

5. CONCLUSIONS

Debates between nominalists and realists over the effectiveness of the Indispensability Argument have focused mainly on the Indispensability Thesis – the claim that mathematics is indispensable for science. This is not the only vulnerable point in the Argument, but it is an obvious point of attack for nominalists who endorse scientific naturalism (of at least the strength encapsulated in the first premise of the Indispensability Argument). My aim in this chapter has been to show that the availability of a reasonably attractive nominalistic theory which captures the physical consequences of our current scientific theories is not by itself sufficient to undermine the Indispensability Thesis. The reason why not is that there is more to the role of mathematics in science than raw deductive power. We have good grounds for believing in mathematical objects if mathematics is indispensable for *doing* science. This concept of indispensability is richer than mere indispensability for proving results about the physical world, because doing science involves the use of theories for tasks other than simple proof. I have concentrated on two other such tasks – discovery of new results, and development of new theories – and argued that they often require more mathematical resources than proving known results. These tasks depend on features of mathematics that are often ignored in debates over indispensability, features that tend to be pragmatic rather than theoretical, local rather than global, and dynamic rather than static. Insofar as these features are crucial to the practice and development of science, any adequate mathematics-free alternative formulation of science must preserve them. The multi-faceted role that mathematics plays in doing science makes the claim that mathematics is *indispensable* more plausible, and this in turn makes the Indispensability Argument more difficult for the nominalist to undermine.

NOTES

[1] Burgess and Rosen (1997) provide a useful overview of the post-Fieldian indispensability debate.

[2] See van Fraassen (1980).

[3] See, e.g., Maddy (1992).

[4] This is not to say that such technical programs are pointless. Establishing that mathematics is dispensable from science, even if only in a narrow, deductive sense, would be important and interesting in its own right.

[5] Or at least those contemporary nominalists who take the Indispensability Argument seriously.

[6] This formulation implicitly equates the concrete part of reality with that part which can be described in purely nominalistic terms. In the context of the debate over the existence of abstract mathematical objects this loose way of talking is unproblematic, but eventually more would need to be said about borderline cases such as properties and relations of concrete objects.

[7] See Field (1980, p. 8).

[8] See Lakatos (1976, pp. 142–144) for an interesting discussion of the 'deductivist style' in mathematics and in science.

[9] Field (1980, p. 5).

[10] See Burgess (1990), Rosen (1992), and Burgess and Rosen (1997) for elaborations of this line of attack on contemporary nominalist strategies.

[11] Another example is the derivation of the Special Theory of Relativity from the General Theory (in the special case where the inertial frame is not accelerating).

[12] The boundary between these two notions is somewhat vague. For example, a case where the mathematics is used to pick out the right answer from a limited range of known alternatives (maybe even just two alternatives) contains elements of both proof-discovery and proof-justification.

[13] For a more sophisticated account of the mathematical aspects of cryptoanalysis see Becker and Piper (1982).

[14] Compare, for example, the relative difficulty of the following two questions:

(i) What are the prime factors of 66887?
(ii) What is the product of 211 and 317?

[15] In rare cases the tables may be turned and empirical resources may be indispensable for *mathematical* discovery. Some theorems about minimum surface areas for irregular shapes were discovered in the 19th Century using wire models and soap bubbles. Another (less clear-cut) case is the use of computers to carry out long and combinatorially complex proofs.

[16] The development of Fourier analysis in the 19th Century was inspired to a significant extent by models of heat transfer along a cylindrical iron bar. Fourier modeled heat flow by calculating the effect of slicing a segment *dm*, off the end of the bar and adding it to the heat sink. By repeatedly carrying out this procedure he came up with an equation for heat flow expressed as an infinite series. For further details see Grattan-Guinness (1972) and (1990).

[17] For more on the application of group theory to physics, see Cornwell (1984), and Banik (1983). A useful discussion of some philosophical aspects of the application of group theory to physics can be found in French (1999).

[18] A more contentious (and less widely known) example is Dirac's 1931 prediction of the existence of magnetic monopoles based on symmetry considerations in the mathematical formalism of quantum mechanics. I will return to examine this example in more detail when discussing the scientific status of Occam's Razor in Chapter 4.
[19] Burgess (1982, p. 99).
[20] Quine (1966, p. 234).
[21] Gödel (1983, p. 485).
[22] The generality of a theory will also be a factor, since a very general mathematical theory has a good chance of finding future application in science even if it is not currently being applied. The prime candidates for discarding are mathematical theories which are both specific and unapplied since such theories will tend to be regarded as both mathematically and scientifically trivial (the theory of chess is one such example).
[23] For example, notational debates in the early development of the calculus over the relative merits of 'dx' notation versus 'x dot' notation.
[24] British algebraists in the early 19th Century subscribed to the so-called 'Principle of the Permanence of Forms' which restricted algebraic systems to those which preserved the basic properties of **N**, **Q**, etc. See Crowe (1985, pp. 15–16) for information on early developments concerning commutativity and other properties of algebraic systems.
[25] For further details concerning division algebras, see Dixon (1994).
[26] I am restricting attention here to *real-normed* division algebras.
[27] Vectors and quaternions are both examples of a broader class of mathematical structures known as 'Clifford algebras'.
[28] The day after his discovery of quaternions, Hamilton wrote that without this property of moduli he would have "considered the whole speculation a failure".
[29] This point is made in an interesting article by O'Neill (1986).
[30] In algebraic terms what Hamilton did was mistakenly identity quaternions with the rotation group in which a rotation of a vector V is expressed as qV, rather than with the spin group, in which a rotation is expressed as qVq^{-1}.
[31] Cf. Anderson and Joshi (1993, p. 312).
[32] The number of equations can be further reduced, to two using tensors or forms, and to a single equation using Clifford algebras.
[33] Crowe (1985, pp. 219–220).
[34] The first quaternionic formulations of special relativity were by Conway in 1911, and by Silberstein in 1912.
[35] Consider, for example, $(1 + \mathrm{i}I)(1 - \mathrm{i}I) = 1 + I^2 = 1 - 1 = 0$ where i is the complex coefficient, and I is the first quaternion vector component.
[36] Rastall (1964), quoted in Anderson and Joshi (1993, p. 316).
[37] See Anderson and Joshi (1993, p. 314), and Finkelstein et al. (1962, pp. 307–8).
[38] There is also a close link between division algebras and Lie groups, which play a central role in modern physics. The four categories of semi-simple Lie groups – orthogonal, unitary, symplectic, and exceptional – are associated with the reals, complexes, quaternions. and octonions respectively. See Anderson and Joshi (1993, p. 313) and Dc Leo (1996b. p. 1827).
[39] Finkelstein (1962, p. 208).
[40] For quantum mechanics to be a special case of classical dynamics, it must be 'symplectic'. One important consequence of this is that degrees of freedom pair up (in technical

terms, a symplectic manifold must have an even number of dimensions). Using the real numbers as a basis does not give this property, hence such a basis is inadequate.

[41] This account of tensor products and complementarity relies heavily on the discussion given in Finkelstein (1962, p. 211).

[42] See Adler (1995), and De Leo (1996b).

[43] De Leo (1996b, p. 1821).

[44] See, e.g., Bold and Wayne (1972), section 6.13 for details.

[45] Michael Crowe points to a tension between pragmatic and theoretical considerations in the debate over quaternions versus vectors:

Many 19th-century physicists I took what may be described as a pragmatic approach to the question of which system was to be preferred. Many of their arguments were on grounds of expressiveness congruity with physical relationships, and ease of understanding. The quaternionists, on the other hand. put somewhat greater stress on mathematical elegance and algebraic simplicity. (Crowe (1985, p. 217))

[46] They also tend to be more precise, and therefore easier to specific exactly.

[47] Reichenbach (1938, p. 7).

[48] See van Fraassen (1980).

[49] See Quine (1960, pp. 248–250).

[50] This is not to say that the 18th-century infinitesimal calculus was demonstrably *in*consistent. There are two basic senses in which a theory can be unrigorous. On the one hand it may be formally inconsistent, as for example was the case with Frege's naive set theory. On the other hand it may be an informal theory for which no consistent formalization has been found. This latter situation better describes that of the early calculus.

[51] Thus, for example, a function f(x) is continuous for a value x = t iff given any $d > 0$ there is an $e > 0$ such that $[f(x +/- e) - f(x)] < d$.

[52] Quine (1960, p. 250).

[53] See, e.g., Eves (1955, pp. 316–7) for an example and a statement of the basic Archimedean principle on which the method is based.

[54] This example is taken from Euclid XII, 2. My presentation of the proof follows the one given in Boyer (1949, p. 341).

[55] Leibniz (1701, pp. 270–1). By "Archimedes' style" Leibniz means indirect proofs by exhaustion.

[56] Indeed it is difficult to see how any formal demonstration could have been given at that time, since it was precisely the lack of a formal, rigorous basis for the new 'science of infinitesimals' which was the source of much of the controversy.

[57] Mancosu (1996, p. 171) writes: "The claim of being able to recast any proof involving infinitesimals into a proof in the style of Archimedes – a proof using the method of exhaustion – was extremely suggestive, but it was never developed in print in a completely convincing way".

[58] Newton's *Principia* is often cited as an example of this phenomenon, although there has been some controversy recently over whether Newton originally discovered his results using infinitesimals (what he called 'fluxions') or not.

[59] The debate over the rigor of infinitesimals finds interesting parallels in the development of methods of proof involving so-called "indivisibles" earlier in the 17th century. One example was the Italian mathematician Evangelista Torricelli who, though himself

a supporter of indivisible methods, recast many of his theorems in order to satisty his (presumably more skeptical) readers. Alter proving one such theorem, Torricelli writes

I consider the previous theorem sufficiently clear in itself.... However, in order to satisfy the reader who is scarcely a friend of indivisibles, I shall repeat its demonstration at the end of the work with the usual demonstrative method of ancient Geometers which, although longer, in my opinion is not for that more certain. (Torricelli (1644, vol. I, p.194)).

[60] A further point against the Quinean reconstruction of the Cauchy-Weiertrass episode is that Cauchy's new methods are themselves not entirely free of infinitisimals.
[61] See Mancosu (1996, p. 36).
[62] See Mancosu (1996, p. 37ff.) for more on earlier 16th- and 17th-century attempts to avoid the Method of Exhaustion by using indivisibles or infinitesimals.
[63] Boyer (1949) writes of the Method of Exhaustion that it "was not a tool well adapted to the discovery of new results" (p. 48) and that it "directed attention toward the synthetic form of exposition rather than toward an analytic instrument of discovery" (p. 36).
[64] See Robinson (1974).
[65] A valuable discussion of the use of nonstandard analysis in physics can be found in Salauskis & Sinaceur (1992, section 4).
[66] See Dauben (1988, pp. 190–3).
[67] Not all developments of nonstandard analysis have this property. Bell (1998) discusses certain new systems of infinitesimals that have arisen out of work in synthetic differential geometry. These systems contain first-order differentials, dx, such that dx = 0 but (dx.dx) = 0, and thus are not conservative over classical analysis.

REFERENCES

Adler, S.: 1995, *Quaternionic Quantum Mechanics and Quantum Fields*, Oxford University Press, Oxford.

Anderson, R. and G. Joshi: 1993, 'Quaternions and the Heuristic Role of Mathematical Structures in Physics', *Physics Essays* **6**(2), 308–319.

Baruk, A. (ed.): 1983, *Quantum Theory, Groups, Fields, and Particles*, Kluwer, Boston.

Becker, H. and F. Piper: 1982, *Cipher Systems*, Northwood.

Bell, J.: 1998, *A Primer of Infinitesimal Analysis*, Cambridge University Press, Cambridge.

Bold and Wayne: 1972, *Number Systems*, Litton.

Boyer, C.: 1949, *The History of the Calculus and Its Conceptual Development*, Dover, New York.

Burgess, J.: 1983, 'Why I Am Not a Nominalist', *Notre Dame Journal of Formal Logic* **24**, 93–105.

Burgess, J.: 1990, 'Epistemology and Nominalism', in Irvine (1990), pp. 1–15.

Burgess, J. and G. Rosen: 1997, *A Subject With No Object: Strategies for Nominalistic Interpretation of Mathematics*, Clarendon Press, Oxford.

Cornwell, J.: 1984, *Group Theory in Physics*, London Academic Press, London.

Crowe, M.: 1985, *A History of Vector Analysis*, Dover, New York.

Dauben, J.: 1988, 'Abraham Robinson and Nonstandard Analysis: History, Philosophy, and Foundations of Mathematics', in Asprey and Kitcher (1980), pp. 177–200.

De Leo, S.: 1996a, 'Quaternions and Special Relativity', *Journal of Mathematical Physics* **37**, 2955–2968.

De Leo, S.: 1996b, 'Quaternions for GUTs', *International Journal of Theoretical Physics* **35**(9), 1821–1837.

Dixon, G.: 1994, *Division Algebras: Octonions, Quaternions, Complex Numbers and the Algebraic Design of Physics*, Kluwer, Dordrecht.

Eves, H.: 1955, *An Introduction to the History of Mathematics*, Rinehart, New York.

Feferman, S.: 1978, *The Logic of Mathematical Discovery vs. the Logical Structure of Mathematics*, in Asquith and Hacking (1978), pp. 309–327.

Field, H.: 1980, *Science without Numbers: A Defense of Nominalism*, Blackwell, Oxford and Princeton University Press, Princeton.

Finkelstein, D., J. Jauch, S. Schiminovich, and D. Speiser: 1962, 'Foundations of Quaternion Quantum Mechanics', *Journal of Mathematical Physics* **3**(2), 207–220.

French, S.: 1999, 'Models and Mathematics in Physics: the Role of Group Theory', in J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, Cambridge University Press, Cambridge.

Gödel, K.: 1983, 'What is Cantor's Continuum Problem?', in Benacerraf and Putnam, 2nd edition (1983), pp. 470–485.

Grattan-Guinness, I.: 1972, *Joseph Fourier*: 1768–1830, MIT Press, Cambridge, MA.

Grattan-Guinness, I.: 1990, *Convolution in French Mathematics, 1800–1840*, Birkhäuser Verlag, Basel and Boston.

Lakatos, I.: 1976, *Proofs and Refutations: The Logic of Mathematical Discovery*, Cambridge University Press, Cambridge.

Leibniz, G.: 1702, 'Mémoire de Mr. Leibniz Touchant son Sentiment sur le Calcul Differential', *Journal de Trévoux* 270–272.

Maddy, P.: 1992, 'Indispensability and Practice', *Journal of Philosophy* **89**, 275–289.

Mancosu, P.: 1996, *Philosophy of Mathematics and Mathematical Practice in the Seventeenth Century*, Oxford University Press, Oxford.

Millard, A.: 1997, *Non-Commutative Methods in Quantum Mechanics*, Ph.D. Dissertation: Princeton University.

O'Neill, J.: 1986, 'Formalism, Hamilton and Complex Numbers', *Studies in the History and Philosophy of Science* **17**, 351–372.

Quine, W.: 1960, *Word and Object*, MIT Press, Cambridge and New York.

Quine, W.: 1966, *The Ways of Paradox and Other Essays*, Random House, New York.

Reichenbach, H.: 1938, *Experience and Prediction*, University of Chicago Press, Chicago.

Robinson, A.: 1974, *Non-standard Analysis*, North-Holland, Amsterdam.

Rosen, G.: 1992, *Remarks on Modern Nominalism*, Ph.D. Dissertation: Princeton University.

Salanski, J.-M. and H. Sinaceur (eds.): *Le Labyrinthe du Continu*, Springer-Verlag.

Torricelli, E.: 1644, 'De Solido Hyperbolico Acuto', in *Opera Geometrica*, Florentiae.

van Fraassen, B.: 1980, *The Scientific Image*, Clarendon Press, Oxford.

Wolfson College
University of Cambridge
Cambridge CB3 9BB
U.K.
E-mail: arb40@cam.ac.uk