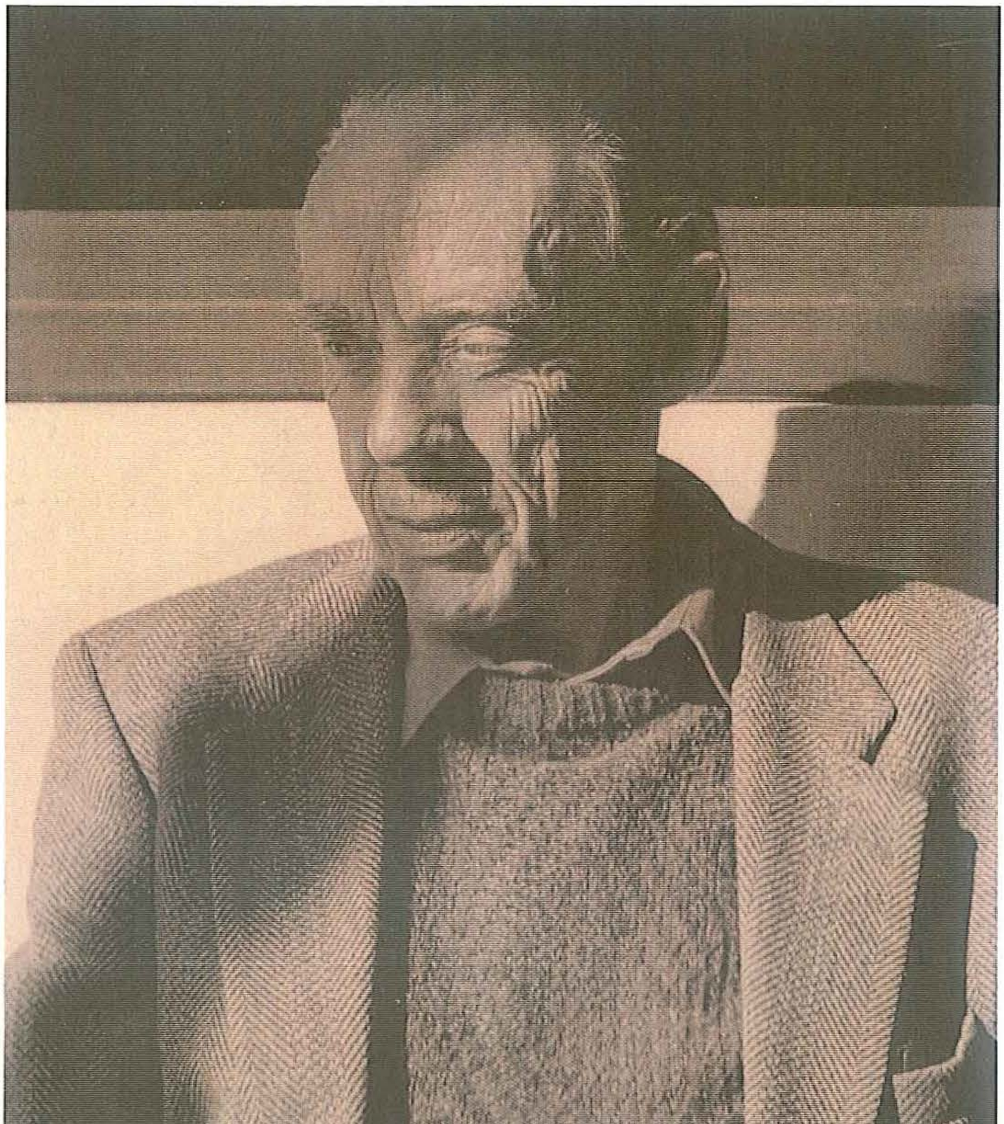
Contexts

Proceedings of ANPA 31

Arleta D Ford, *Editor*



Published by ANPA

June 2011

Contexts

Proceedings of ANPA 31

Arleta D. Ford, ***Editor***

Proceedings of the 31st Annual International Meeting of the
Alternative Natural Philosophy Association

Wesley House, Jesus Lane, Cambridge
August 2010

Contexts: Proceedings of ANPA 31
Arleta D. Ford, *Editor*

ISBN 978-0-9562148-1-2

Printed in Great Britain by the MPG Books Group,
Bodmin and King's Lynn

CONTENTS

EDITORIAL

The title of this proceedings, “Contexts” is taken from Clive Kilmister’s paper “Process, the Combinatorial Hierarchy and Similarity” included here. He wrote the paper in November 2009, and the paper is remarkable in many ways. He wished it to be published in the proceedings, knowing that he would not be able to attend the meeting. However, for a number of reasons – some of them justified, I haven’t included it in the last year proceedings. Referring to the paper he writes :

> Red Tiles Cottage
> High Street
> Barcombe
> Lewes
> BN8 5DH
> 11 – 11 – 09
>
> Dear Arleta,
>
> Here it is at last! I'm not as pleased with it as I hoped but it is the best I can do for the moment. You had an earlier version of the first half some months ago. It would be best if you could destroy it, to prevent confusion. I haven't made many changes to that first eleven or so pages but what I have made are quite important.

Indeed, what he has written in (the second part of) the paper is very important. First of all, he introduces several new key concepts which were never before used in the work on the Combinatorial Hierarchy; one of them is what he calls “context(s)”. As one would guess knowing Clive, the meaning of the concept is significantly different from what the word itself would suggests. With the help of the new concepts Clive is able to raise a number of important questions about the construction of the CH; the kind of questions which for years have been swept under a carpet as it were. To some of them he proposes answers which greatly clarify the whole construction. He also comments on his and Ted Bastin's recent book "The Origin of Discrete Particles", and finishes the paper pointing to the "way ahead". The first part of the paper goes into the traditional subject of the origin and the history of different constructions of the CH which could be of benefit to people who are new to the subject.

The photograph of Clive Kilmister which is on the cover of the proceedings was made by Keith Bowden few years ago.

This reminds me to repair my unforgivable lapse with regard to the previous year proceedings (no 30) of not mentioning that the picture on the cover was (obviously!) created by Prof Louis H. Kauffman. As we see, in Lou's hands the Wheeler Eye underwent a significant twist, transmuting into a self-pondering Universal Knot. To find the details of and reasons for such a transmutation of Wheeler's picture of the Universe you would need to study Lou's paper "Reflexivity" in the previous year proceedings.

Wishing everybody great meeting!

Arleta D Ford

PROCESS, THE COMBINATORIAL HIERARCHY AND SIMILARITY

Clive Kilmister

10 November 2009

1. Introduction

I am going to return to a number of questions of interpretation of the Combinatorial Hierarchy (CH). Some are new, others have been troubling me at the back of my mind for many years. My reason for doing this is twofold, in our two books (Combinatorial Physics (CP) (1995) and The origin of discrete particles (ODP) (2009)) Ted Bastin and I made progress in explicating how Frederick Parker-Rhodes' CH, and its corrected non-associative version, might have physical significance by using the idea of process. The process is that by which knowledge of the physical world is obtained. We assumed that this arises discretely and speak of items of knowledge, or simply *items.* In ODP we came to the conclusion that the original CH construction with its changes of level was not the only one. We introduced the "transverse construction" as well. More needs saying about the details of this since it is unclear in ODP. Many of the mathematical details in the following are not new, so I have relegated them to the notes. My second reason is that corrected algebra opens the way to a calculus of similarities that needs explanation.

In Frederick's original construction of the CH the basic step is to determine whether two items are the same or different. How is this done in

CH? Consider an analogy from high-energy physics. Go back a little way historically to when cloud chamber photographs were examined. The question "Is this a μ - meson or not?" becomes "Has it the right quantum numbers?". Now all this is done by computer but this does not affect the analogy. The machine is conducting the same questions for you. The CH description simply abstracts from all the physical details and, in the simplest case, the way it does it is given by the mathematical properties of the theory. I say "in the simplest case" because level change produces complications but equally there the mathematical structure is the means of making the distinction. The advent of the transverse construction has made me look more closely at these structures both in the original CH and in the corrected version for which the CH is the "skeleton". I begin with this skeleton.

2. Mathematical structures

Items are represented by elements in the structure; as elements arise they are labeled. When two elements are different, this determination is a new item, so is represented by a new element, w say, which is then adjoined to the others. The mathematical structure describing the determination of whether two elements u, v are the same or different is therefore a binary operation, $u, v \rightarrow w = u.v$. To carry out the determination the value of $u.v$ when u, v are the same must always be different from when u, v are different. Notice that it is not being said that the elements are a set closed under a binary operation. There are two reasons for this, which I shall explain shortly. The first is that new labels can always be introduced as needed and are, so that to say that the structure is closed adds nothing.

The second reason is more important. It concerns what happens when u, v are the same and therefore the values of $u.v$ when $u = v$. In this case there is nothing to be adjoined. An obvious way of signifying that $u = v$ would be to take $u.v = u$, but this will not do from a process point of view because the next step would then have to determine that the output u was in fact the same as the input u. That is, the same problem again, which would lead to an infinite regress.

Moreover, that u, v are equal is not the same item as u itself. For these reasons $u.u$ must be a non-element, which we call a *signal*. The process can then recognize this without further determinations. There is no loss of generality in assuming there is just one signal, z say, so named because for Frederick it was the zero string. That signals are not elements is the second reason for the system not being closed under the binary operation.

If now $u.v = z$, then there is nothing to be adjoined, that is, v is rejected. Here an important difference between the skeleton and the full hierarchy arises. In the skeleton no distinction is drawn between $u.v$ and $v.u$, so if v is to be rejected, so must u. The advent of another u removes the original one, so that elements vanish as well as arising. In the full hierarchy there is a distinction which is captured by $u.v \neq v.u$, so that only the second element is rejected. This does not change the overall structure but it does mean that the skeleton tends to need more steps before complex structures arise and even then they may vanish again.

Since the binary operation is to characterize inequality, if $v \neq w$ then $u.v$

must be unequal to $u.w$. These considerations can be summed up in this:

Definition

1. If $u = v$, then $u.v = z$;
2. If $u \neq v$, then $u.v = w$ where w is different from u, v, $u'.v$, $u.v'$; u', v' being any labels that have already been given by the definition.

It is useful to have a particular ordered set of labels to capture more precisely the idea of labels that have already been given. A convenient set is that of the ordinals 1, 2, 3, ... with the usual ordering. Then the definition can be transformed into

Conway's Rule

If $u \neq v$, the label of $u.v$ is the least one different from u, v, $u'.v$, $u.v'$, where u', v' are any labels less than u, v.

The least label is chosen as a matter of convenience. With that proviso the binary operation is called *discrimination.* It is easily proved to be commutative and associative. I should make it clear that this is a "sanitized" version of the process[1]. It is of what will finally come about according to a "generalized ergodic principle" that if something is possible it will eventually happen. My vision is of a kind of mist of discriminations and through the mist structures with some degree of permanence.

The first such permanent structure is that generated by two elements. The structure has three elements in all and it has the property that it is closed under the discrimination of two different elements. A set with this property is called a discriminately closed subset (dcs). Following the rule the Cayley table has the form[2] :

	1	2	3
1	z	3	2
2	3	z	1
3	2	1	z

This is the table that results from two elements 1, 2 arising and 3 is then the item 1.2. The process cannot give rise to values of $z.z$, $z.u$, $u.z$ because z is not an element so one is free to assign any values, and in particular z, u, u respectively and the result[3] is to give an extended table which is that of the quadratic group, S = $C_2 \times C_2$, where C_2 is the cyclic group of order 2. This accounts for Frederick's original (and puzzling) expression of the theory in terms of bit-strings. The rule and the notation together result in $u.v$ being labeled with the Nym sum of the labels of u and v.

Mathematicians prefer to deal with groups and this does not give rise to any difficulties so long as we remember that the identity is to be omitted. For reasons that will become clear below, it is convenient to have a name for a dcs with three elements. I shall call it a *context*. It is important to note that the symmetry between 1, 2, 3 in the context (1, 2, 3) , though it can be useful in physics, conceals the fact that there remains a difference, important for the CH, between the two generators 1 and 2 and the derived 3.

3. The next step

The rule allows the mathematical machinery to go on further, but it requires further investigation whether the continuation is physically useful. If one takes three generators 1, 2, 4 then the same argument determines a dcs with 7 elements[4], or equivalently $C_2 \times C_2 \times C_2$. Similarly for r generators the number of elements is $2^r - 1$ which I denote here and later as r^*. The possible numbers of elements are therefore 3, 7, 15, 31, ... corresponding to $(C_2)^r$.

These higher structures cannot arise in this way in the process though they can enter in another way. The reason for this is that if a context, say C = (1, 2, 3) has arisen and another element x, arises and if, for the sake of argument, x is 1 (but of course has not yet been identified) then a member of C has to be selected to discriminate with x. There is a probability 1/3 that 1 is selected and so x is determined, and 2/3 that a new element will be generated by the discrimination. Following this up in the second case, there is now a probability of 1/4 of finding $x = 1$, and so on.

The probability of success after k trials is $2/[(k + 1)(k + 2)]$. Hence the mean value of the number of trials needed is twice the sum of $k/[(k + 1)(k + 2)]$ which is evidently divergent.[5] (If one decides to give up after n trials, fixing n beforehand, then the mean number of such trials is roughly $2[\log(n + 2) - 1.4228)]$). I shall call such higher order dcss *virtual*, since they are not produced directly by the process.

Two methods of proceeding further have been suggested. Whether there are others I do not know. Each requires a definite change in procedure. I deal with each in turn.

4. Level change

This is the method introduced by Frederick. From the process point of view one has to say that the process may consist of a level change at this point or not. Instead of dealing with the original elements 1, 2, 3 level change deals with the dcss generated by 1 and 2 (so here the importance of distinguishing between generators and derived elements arises). The sanitised version of level change is:

Each of the three dcss [1], [2], [1, 2, 3] is labelled by the corresponding map of elements into elements which preserves discrimination. The label of each dcs is then that map which leaves it and only it invariant. In this way the infinite regress lurking in simply labelling the dcss 1, 2, 4 and discriminating is avoided.

Another effect, which is important for us, though it was evidently not in Frederick's mind, is that a dcs is now protected against becoming incomplete by the advent of a new element coming in which is one of its members.

A suitable notation for the maps is (u, v) where u, v are the results of applying the map to 1, 2 respectively. With suitable values for the signal z combined with elements the resultant structure is $C_2 \times C_2 \times C_2$.[6] Thus we have

Theorem 1

The structure generated as the next level by the level change is isomorphic to the virtual structure produced by using Conway's rule on three generators.

The reader may consider this too trivial a result to be designated a theorem. My reason is that the later results for the actual hierarchy are more complex. Any two of the three elements generate a context and there are seven of these in all. With r generators there would be $\frac{1}{3}r * (r-1) *$ contexts.[7] An important corollary is this: At every level the structure is symmetric, meaning that every context is isomorphic to the original (1, 2, 3).

Frederick then proposes to repeat the construction. The three generators define seven dcss, so that the next level would have seven generators. Continuing in this way the number of elements will be $u_n = 3, 7, 127, \ldots$ where $u_{n+1} = u_n*$. There is an important difference from the first level change. The first set of maps was uniquely determined. At the next stage this is not so. For example, each 1-element dcs can be labeled by one of 14 maps and every 3-element dcs by one of three.[8]

It is these possibilities that allow the calculation of the accurate value of the fine-structure constant, so they are not to be avoided in some way, if such could be found. Discrimination between dcss is now defined by discrimination between labels, so that once a dcs has been labeled, it must retain that label, for otherwise discrimination would be affected.[9]

5. The transverse construction

This construction arises when the process fails to take up the option of level change, so to some extent it seems an automatic alternative, about which more need not be said. This was how Ted and I looked at it when we came across the need for it in our discussion of our approximation to continuous space. But we soon saw that there were other applications of it that emphasized its importance. So my purpose in this paper is to find what further mathematical machinery is involved.

The description of the construction in ODF simply says: one dcs (a context) is left and another started. This is incomplete and it is also a sanitized version as in the other constructions. But this time the situation is slightly different, because it is now the random discriminations that generate the structure. This is in the nature of things; it is not because we came upon the need for the construction before the details were clear. Consider to begin with what happens in level change. The essential feature is that a halt is called to the original process of discrimination. A new step is introduced, that of labeling dcss by maps and then discriminating between these maps. In the transverse case it is similarly easiest to split the story into two parts. (In the actual construction these take place at random and so simultaneously.) In the first part elements arise and generate a context. Continuing, new elements arise and generate a second context, and so on. In this way a *spray* of contexts arises.

Any two contexts may be a) the same, b) overlapping (i.e. have one element in common), c) different, which suggests a new kind of discrimination.

But the complete determination of such a discrimination would again run into the infinite regress problem. This points up the essential problem to be tackled for the construction: it is that, although as in the level change, the process is really concerned with dcss, unlike the level change, it is still operating with elements.

The most that is possible, then, is for the elements of various dcss to continue to discriminate and it may then transpire that an element of one dcs, A say, is the same as one of another, B. In order to say in this case that A and B overlap one more step is needed. In level change the switch to maps from dcss protects the integrity of the dcss.

So here, the integrity of the dcss has to be protected in some way. The further step to do this is to require that the process of determining overlap should differ from ordinary discrimination by retaining both of two identical elements in two dcss. Then the contexts continue discriminating in this way, and this is a *web* of contexts. As more discriminations take place the web grows. A spray will be called *fully webbed* when any two contexts in it overlap.

There are two kinds of fully webbed spray. The first kind has an element common to all the contexts. Take this element as 1 so the spray is [(1, a, b)] and the number of contexts in the spray is the number of pairs (a, 1.a) which, for r generating elements is

$$\frac{1}{2}(r^* - 1) = (r - 1)^*$$

Thus for $r = 2$ this is a single context, for $r = 3$ there are 3 and for $r = 4$ there are 7, and so on. For our purposes the second kind is more important, because it gives rise to the basic unit in the transverse construction, which we shall call a *nucleus*. In this kind there is no common element to all the contexts. Take one context as $(p, q, p.q)$. The next can be taken as $(p, r, p.r)$. If there is to be no common element then there must be at least one context omitting p by itself, say: $(q, r, q.r)$ or $(q, p.r, q.p, r)$ or $(p.q, r, p.q.r)$ or $(p.q, p.r, q.r)$ and these four mutually overlap and with the first two. Finally $(p, q.r, p.q.r)$ overlaps all these but no more are possible. This second kind of fully webbed spray has 7 contexts, and this is independent of the number of generators so long as it exceeds 2. Thus

Theorem 2

The nucleus is isomorphic to the virtual structure produced by using Conway's rule on three generators, as in the case of level change.

6. The correct hierarchy

I turn to the full hierarchy, in discussing which the results for the skeleton will be a useful guide. The description of the structure at the beginning of §2 still holds, up to but not including the definition of discrimination. The problem with this definition from a process point of view is that it identifies the result of an element v arising and being discriminated with a u that has already arisen with the result of u arising and being discriminated with an already existent v. It does this because the definition assumes that $u.v = v.u$. Now $u.v$ and $v.u$ label different determinations but they have something in common, viz. that each is between the same two elements. This therefore introduces the germ of an idea of *similarity*. If there exist elements u, v such that $w_1 = u.v$ and $w_2 = v.u$ then we say that w_1, and w_2 are *similar*. A suitable notation for the present is to write $w_1 \cong w_2$. At a later stage this notion will develop into a more recognizable similarity relation.

This notion of similarity is touched on in ODF but only in passing. (It was, of course, absent in CP.) This was an error. It is fundamental to the correct structure, and indeed the investigation of the correct structure for 3 or more generators will turn out to be based on the three principles: finiteness (as usual), similarity and a restricted version of similarity, duality, to be described below.

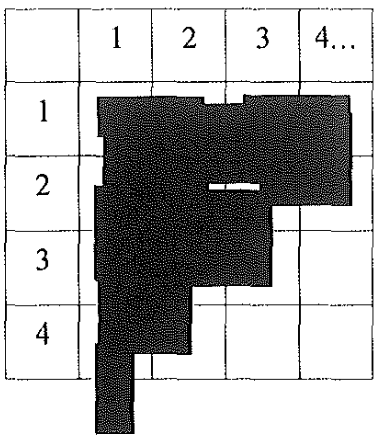
As a result of this definition of similarity there are now three relations possible between two elements: (i) equal, (ii) different but similar, (iii) different and not similar, or one could simply say, dissimilar. I shall use the term similar to include equality so that it can later be postulated to be transitive (as well as its obvious properties of reflexivity and symmetry). I shall also later want to postulate that the similarity relation is consistent with discrimination, i.e., that, if $u \cong v$ and $x \cong y$ then $u.x \cong v.y$. I shall call two contexts similar if every element of one is similar to an element of the other.

What is happening here is this: the discrimination operation turns out to be non-associative and these postulates of transitivity and consistency restrict the non-associativity in a certain way. It is not profitable to express this restriction directly in terms of the binary operation.

It is convenient to have a further term to cover the case when two elements are similar but different. I shall then write $w_2 = w_1{}^*$ (or equally $w_1 = w_2{}^*$) and call w_1 and w_2 *dual*.[10]

We have to decide now how the process point of view can deal with this. Just as before, in case (i) there must be a signal z. For (iii) there must be a new element labeling the discrimination. For (ii) the argument in the skeleton for equality holds equally here, another signal, y say, must be introduced, so that $w.w^* = y$. Since duality is symmetric it follows that $w^*w = y$ and so dual pairs must be excepted from the rule that $u.v \neq v.u$. We now have to investigate how the definition leading to Conway's rule has to be changed. There are several steps.

(i) Because the binary operation is non-commutative, the process has to be carried out in a definite order in constructing the Cayley table. I shall use the one shown in the diagram, which I shall call the *standard path*.[11]



(ii) An obvious choice for the new definition would be If $u \neq v$ then $u.v$ is the least label different from u, v, $u'.v$, $u.v'$, or $v.u$.

This will not do because it fails to take the existence of duals into account. The table for 2 generators is :

	1	2	3	4
1	z	3	4	2
2	4	z	1	3
3	2	4	z	1
4	3	1	2	z

Its defect is that, from the table, $3 \cong 4$, $4 \cong 2$, $2 \cong 3$, $1 \cong 4$, $3 \cong 1$ and $1 \cong 2$, so that all elements are similar.

(iii) An examination of why this failure occurs serves to show what might be a better definition. Going along the standard path the trouble begins at 1.3. Why should this not be 4? Because although 4 differs from 1 and 3, it is similar to 3. (One already knows this since 3 = 1.2 and 4 = 2.1.) This suggests replacing "different from" with "dissimilar" (though of course with the exception of $v.u$, since $u.v$ is by definition similar to $v.u$). Here the Cayley table begins:

	1	2	3	4	5	6	7	8
1	z	3	5	6	10	9	y	11
2	4	z	1	7	y	12	6	10
3	2	7	z	y	8	13	10	
4	8	1	y	z	11	2		
5	9	y	6	12	z			
6	10	11	14	5				
7	y	8	9					
8	12	9						
9	13							

where

$3 \cong 4$

$2 \cong 5$

$1 \cong 7$

$6 \cong 8$

$9 \cong 10$

$11 \cong 12$

$13 \cong 14$

and it is clear that this will be an infinite table.

(iv) The scheme in (iii) was an over-reaction to the defects of (ii) which suggests the preferred definition:

Definition

i. If $u = v$, then $u.v = z$

ii. If $u \cong v$ and $u \neq v$, $u.v = y$

iii. If u, v are dissimilar, then $u.v$ is the least label

(a) dissimilar from u and v, and

(b) different from all $u'.v$, $u.v'$ and $v.u$, where u', v' are any already labeled elements

It is clear that (b) must have "different from" rather than "dissimilar" because an earlier $u.v'$ (say) might come from the dual of v and of course uv^* should not be excluded. The Cayley table is now:

	1	2	3	4	5	6
1	z	3	5	2	4	y
2	4	z	1	6	y	3
3	2	6	z	y	1	5
4	5	1	y	z	6	2
5	3	y	6	1	z	4
6	y	4	2	5	3	z

Here $3 \cong 4, 2 \cong 5, 1 \cong 7$

In terms of the dual notation, so that 4 = 3*, 5 = 2*, 6 = 1*, the Cayley table shows that duality has the properties

$$u^{**} = u, \quad u.v^{*} = (u.v)^{*} = u^{*}.v = v.u$$

Using these properties, the Cayley table can be re-written:

	1	2	3
1	z	3	2*
2	3*	z	1
3	2	1*	z

In this form I call the structure Q* (because of a partial resemblance to quaternions, which I call Q).[12]

The structure Q* (which is obviously non-associative) replaces the notion of context in the skeleton. So a context now has 6 members and has the general form (u, v, $u.v$, u^*, v^*, $u.v^*$). We shall continue to use the same notation for contexts as before, so that (1, 2, 3) stands for (1, 2, 3, 3*, 2*, 1*), and call (1, 2, 3) its *label*.

7. The next step

I now retrace the steps of §3 to show how the notion of similarity determines what virtual structures are possible with 3 generators. Here again the structures are called virtual since the direct use of a process would lead to infinite regress. But first the notion of dual needs a little explanation. The * notation is very convenient but can be misleading. Consider the element 1* in the last section. This really means: express 1 as 2.3; then 1* = 3.2. This arises because 1 is seen as belonging to the context (1, 2, 3). Now 1 also belongs to other contexts, one of which is (1, 4, 5). In this context the meaning of 1* changes to: express 1 as 4.5; then 1* = 5.4. If we continue to use the original concept of dual, this has the effect of identifying 3.2 and 5.4, and these with 7.6.

Similarly with other dual terms. The restriction is possible (and will prove important later on) but seems severe, so I turn to other possibilities. In §4 I noted that the virtual process in the skeleton, or equally level change, produced a *symmetric* structure (every context isomorphic to S). My first simplifying assumption is to limit attention to correct structures that have the corresponding property:

Assumption

The structure is symmetric in the sense that every context (dcs generated by two generators) is isomorphic to Q*.

There is such a structure with a single dual operation.[13] It therefore has 2.7 = 14 elements (i.e. $2r^*$) and two signals. I believe such structures exist for any number of generators, a fact which is useful in extending the theory.

Such an identification is only one possibility; there are others less restrictive. There is no a priori reason why $3.2 = 5.4$, so the two duals need to be distinguished by a new notation. I begin by labeling the 7 contexts listed in note 7 with suitable letters: $a = (1, 2, 3)$, $b = (1, 4, 5)$, $c = (2, 4, 6)$, which I call *basic* contexts, $d = (1, 6, 7)$, $e = (2, 5, 7)$, $f = (3, 4, 7)$, and $g = (3, 5, 6)$, and I call all of the 7 *initial contexts*. I then replace the * notation by writing 1^a for 1* in context a and so on for b and d and similarly for any other contexts.

Naturally the properties $u.v^p = u^p.v = (u.v)^p$ still hold, where p is any one of the context labels and u, v belong to the context p. Since every element is in three initial contexts it has 3 duals. The number of elements is therefore at least $(3 + 1).7 = 28$. This extension in the notion of dual produces a corresponding one in similarity:

Provisional definition

$u \cong v$ if and only if there is a context p in which $u = v^p$

I call this definition provisional because it assumes the one-to-one relation between contexts and duals. I show below that to maintain this relation leads to an infinite structure. Even with restriction to symmetric structures there is more to the story for many more contexts arise. For example, consider the context generated by 1^a and 2^c. Then $1^a.2^c \cong 3$ so that by the provisional *definition* there must be a p for which $1^a.2^c = 3^p$.

The collection of contexts evidently splits into seven similarity classes. Each class has one initial and twelve other members, so there are 7 x 13 = 91 at this stage. For consider the initial context $(1, 2, 3)$. Now 1 could be retained or replaced by 1^a, 1^b, 1^d and 2 by 2^a, 2^c, 2^e. Writing 1^z for 1 4 of the 16 possible pairs give only a repetition of the original: for *z,z, z,a, a,z, a,a*. If each new context introduces a new dual and the new duals then generate new contexts, the number of elements obviously increases without limit.

Some restriction is needed to restrict the structure within finite limits. That is to say, the new elements being generated must be identified with existing ones. New duals inevitably lead to new contexts, so the identification must tackle the other branch of the dilemma; new contexts must be restricted from giving rise to new duals. This needs to be done in some systematic way. I shall now describe one such set of identifications which produces what I call *binary-constrained structures* (bc-structures).

I shall argue that these structures arise in a natural way because the structure of discrimination leads directly to a definition of a binary operation between the labels of initial contexts. This then induces in an obvious way the corresponding operation between similarity classes of contexts.

Any three initial contexts have a single element in common, so one can define: $(u, v, w).(u, x, y) = (u, p, q)$ where p, q are the remaining elements from the whole set [1, 2, 3, 4, 5, 6, 7] other than u, v, w, x, y.

The various properties (such as that two contexts must have a single common element) derive from discrimination, which is why I say that the new operation derives from it. An alternative definition yielding the same result is to define (u, p, q) as the set (u, $v.x$, $v.y$, $w.x$, $w.y$) which will turn out to have repetitions; this is inconvenient to work out and not so intuitive, but it serves to emphasize the derivation from discrimination.

If the definition is completed by defining the product of any context with itself as an identity the new operation is evidently commutative and it is quite easy to prove that it is associative, notwithstanding its derivation from discrimination. It is therefore the operation of a group structure in which every element is of order 2, and so of $C_2 \times C_2 \times C_2$.

Lest it should be thought that this binary operation is simply a trick that would work only for three generators, I should say that it can be extended to any number, but the natural operations for higher r are not between contexts but between what I will call (borrowing a term from the algebraic geometers) *primals*, that is, dcss with one less generator than the number for the whole structure.[14]

I shall now show how the binary operation can be used to keep the structure finite. There are several steps. Evidently $1^a.2^c = 3^{a|c} = 3^p$, where $a|c = p$, some dual. The first step is to define the binary operation $a|c$ as the same operation as $a.c$ between the labels of initial contexts. This is a considerable restriction since there are other ways of defining $a|c$. For example, $1^a.4^c \cong 5$ and so the assumption that $a|c$ is in general a binary operation this must be 5^p. Similarly $1^a.6^c = 7^p$. But to what context would p then refer? There is no context similar to (3, 5, 7). We must therefore allow that the structure will have new elements (in the apparent form of duals) which are not defined by contexts. That is, further elements similar to given ones but not defined by contexts. Although this introduction of new elements may seem discouraging, the binary operation can be used to keep the structure finite.

The provisional definition must now be changed:

Definition

$u \cong v$ if and only if there is a dual operation p for which $u = v^p$.

Then the second step is to define similar contexts to have identical dual operations. Then $2^c.1^a = (3^a)^p = 3^{a \times p}$ (say) where $a \times p$ is some dual. The third step is then to define this new binary operation, $a \times c$ as $a.c$ as well.

I turn to the way in which this natural binary operation keeps the structure finite. Consider again $1^a.2^c$. Since

$$a.c = (1, 2, 3).(2, 4, 6) = (2, 5, 7) = e$$

we have $1^a.2^c = 3^e$. Notice that 3 is not otherwise defined, since 3 does not belong to the context e. Similarly, by the second step,

$$2^c.1^a = (3^a)^e = 3^c$$

because $a.e = c$, with the same remark about 3 and the context c.

Overall, then, bc-structures achieve finiteness by removing the tight connection between contexts and dual operations in two ways: different contexts can have the same duals, and some duals do not arise from contexts at all. None the less, enough is left for the notions of context and dual to still be recognizable.

The easiest way to look at this bc-structure is to begin with the three labels of the basic similarity classes *a*, *b*, *c* as the generators of the group. The full Cayley table for these three similarity classes is easily calculated as:

	a	*b*	*c*	*d*	*e*	*f*	*g*
a	*z*	*d*	*e*	*b*	*c*	*g*	*f*
b	*d*	*z*	*f*	*a*	*g*	*c*	*e*
c	*e*	*f*	*z*	*g*	*a*	*b*	*d*
d	*b*	*a*	*g*	*z*	*f*	*a*	*c*
e	*c*	*g*	*a*	*f*	*z*	*d*	*b*
f	*g*	*c*	*b*	*e*	*d*	*z*	*a*
g	*f*	*e*	*d*	*c*	*b*	*a*	*z*

But the extent to which this defines the full Cayley table for the three generators needs more investigation. If the generators are, as usual, 1, 2, 4 we can define: 3 = 1.2 5 = 1.4 6 = 2.4.

Different definitions, as for example, trying to preserve circular symmetry by defining $5 = 4.1$, would lead to isomorphic tables. But there are three significantly different definitions of 7: $1.6 = 7$, $2.5 = 7$, and $3.4 = 7$ (with duals as further possibilities). If I begin with $1.6 = 7$ the following partial table easily results:

	1	2	3	4	5	6	7
1	z	3	2^a	5	4^b	7	6^d
2	3^a	z	1	6	.	4^c	.
3	2	1^a	.	.	.	.	.
4	5^b	6^c	.	z	1	2	.
5	4	.	.	1^b	z	.	.
6	7^d	4	.	2^c	.	z	1
7	.	.	.	.	.	1^d	z

If I were to continue by setting $2.5 = 7$, this makes the identification $2.5 = 1.6$. Since $2.5 \cong 7$ it is also possible that $2.5 = 7^p$ for some p. Because of the non-associativity, p can be taken fairly arbitrarily. A concrete example of such a structure comes up in the next section.

Each element now has 7 duals so the similarity classes of elements have 8 members and the whole structure has 8 x 7 = 56 elements (as well as 8 signals, one for identity and 7 for the various dualities). Thus even in the case $r = 3$ the similarity relation becomes a little more recognizable.

In the Combinatorial Hierarchy it would be natural to go now to $r = 7$ but in looking at similarities it is more natural to look at all values of r. For $r = 4$ there are evidently 35 initial contexts of which 6 are basic. But here (and so for larger r) it is necessary to begin with the primals (dcss with 7 members) of which there are 15. It is between these that the natural binary operation holds. For general r there are r^* primals and any two of them have $(r - 2)^*$ elements in common. For the present I confine attention to $r = 3$.

8. Level change and the transverse construction.

The last section was concerned with ensuring finiteness in any structure with three generators. Frederick's approach as set out in §4 is one alternative way of doing this in the skeleton. In the actual hierarchy we again follow this; the three dcss [1], [2] and [1, 2, 3] are the basis, though now, of course, these labels are shorthand for the full dcss with duals included as [1, 1*]. Each dcs is then labeled with that discrimination-preserving map which leaves it and only it invariant. In note 6 these were specified for the skeleton as (1, 3), (3, 2), and (1, 2), where (u, v) is the map carrying 1 to u and 2 to v. This notation is inadequate for the actual hierarchy because of the non-commutativity, but the necessary changes are not difficult to make. Begin with the skeleton and change the notation for maps to what I will call the prolongation: this is the set of three elements into which the matrix maps 1, 2, 3 resp. So the three prolongations are: (1, 3, 2), (3, 2, 1), (1, 2, 3).

Now in generating the next level one has to add these automorphisms (for that is what they are). Because discrimination is being taken as commutative it is a simple theorem that the prolongation of a sum is the sum of the prolongations. So really the idea of a prolongation is not needed in the skeleton but becomes important in the actual construction.

There the three maps, being automorphisms, become (1, 3, 2*), (3, 2, 1*) and (1, 2, 3). Call these 1, 2, 4 at the next level and label the 7 contexts that they produce as in §7. Further define 1.2 = 3, 1.4 = 5 and 2.4 = 6 and choose out of the 3 ways in which 7 can be defined 1.6 = 7. Then

$$3 = (2^*, 1^*, 3^*),\ \ 5 = (z, 1^*, 1^*),\ \ 6 = (2,\ z,\ 2)\ \text{and}\ 7 = (3, 3, y).$$

Any dual operation in these 7 contexts is then easily calculated. For example, 1.2 = (2*, 1*, 3*), 2.1 = (2, 1, 3), so that the dual consists of dualising each of the lower level components: $(p,\ q,\ r)^a = (p^*,\ q^*,\ r^*)$. Then[15] the Cayley table comes out to be:

	1	2	3	4	5	6	7
1	z	3	2^a	5	4^b	7	6^d
2	3^a	z	1	6	7^g	4^c	5^b
3	2	1^a	z	7^e	6^d	5^c	4^d
4	5^b	6^c	7^d	z	1	2	3^e
5	4	7^b	6^c	1^b	z	3^d	2^g
6	7^d	4	5^d	2^c	3^c	z	1
7	6	5^g	4^e	3^d	2^b	1^d	z

Surprisingly, then, the analogue of Theorem 1 follows:

Theorem 3

The structure generated at the next level by level change is isomorphic to one of the virtual bc-structures produced by using the extension of Conway's rule for three generators.

Turning to the transverse construction, the version for the skeleton in §5 can be taken over with only the obvious slight changes and so yields:

Theorem 4

The nucleus is isomorphic to one of the virtual bi-structures produced by using the extension of Conway's rule for three generators.

9. A way ahead (tentative!)

The generalisations to higher r are fairly evident but something needs saying about symmetry. Instead of importing the idea afresh each time it seems better to make a slightly stronger definition. Consider a tower of structures. At each level r there is a minimum structure with $2r^*$ elements (& 2 signals) and a single dual. Call this $Q_r{}^*$ (so $Q^* = Q_2{}^*$). Then there is the extension, with r^* dual operations and $(r^* + 1)r^*$ elements (& $r^* + 1$ signals). Call this R_r. Then define this as a symmetric tower if, at every level, every primal of $Q_r{}^*$ and of R_r is isomorphic to $Q_{r-1}{}^*$.

Notes

1. The process really consists of discrimination going on randomly at all stages: sometimes elements vanish again. It begins like this: an element arises, so is called 1. If 1 arises again, next, 1.1 = z so both elements vanish. Eventually 1, 2 arise in succession; then 1.2 = 3 by the rule. If 2, say, then arises again, there are three possibilities for the next step:

1.2 = 3 so that [1,2,3] remains;
2.2 = z so only 1, 3 remain;
3.2 = 1 so that [1,2,3] remains.

As explained in the text, this is different from the actual hierarchy, when a distinction can be made between already established elements and putative new ones. Then 2.2 = z means that the new 2 is dropped but the old one is retained.

2. There are two generators 1, 2. Now 1.2 cannot be 1 or 2 so is 3. The same for 2.1. Similarly 2.3 cannot be 2 or 3 but can be (and so is) 1 and so on.

3. The new Cayley table is

	z	1	2	3
z	z	1	2	3
1	1	z	3	2
2	2	3	z	1
3	3	2	1	z

4. The dcs given by Conway's rule is, easily,

	1	2	3	4	5	6	7
1	z	3	2	5	4	7	6
2	3	z	1	6	7	4	5
3	2	1	z	7	6	5	4
4	5	6	7	z	1	2	3
5	4	7	6	1	z	3	2
6	7	4	5	2	3	z	1
7	6	5	4	3	2	1	z

(The entries in the top left hand corner are as before. Then $1.4 \neq 1, 4, 2, 3$ so $1.4 = 5$. $1.5 \neq 1, 5, 2, 3$ so $1.5 = 4$ and so on.)

5. For

$$2\sum_{k=1}^{\infty}\frac{k}{(k+1)(k+2)} = 2\sum_{k=1}^{\infty}(\frac{1}{k+1} - \frac{1}{(k+1)(k+2)})$$

The first term is the divergent harmonic series, the second is

$$2\sum_{k=1}^{\infty}\frac{1}{(k+1)(k+2)} = 2\sum_{k=1}^{\infty}(\frac{1}{k+1} - \frac{1}{(k+2)}) = 1$$

If n is fixed as a maximum number of trials, the mean value becomes

$$2[H_{n+2} - 3/2] - 1 + 1/(n + 2)$$

where H_{n+2} has the asymptotic value $\gamma + \log(n + 2)$, and γ is Euler's constant. For large n this gives the result in the text.

6. The actual process consists of random discriminations and random recognitions of dcss which can be labeled. Through the mist will gradually appear the elements of the next level.

The three dcss are represented by the maps (1, 3), (3, 2) and (1, 2). The discrimination of maps is induced in the usual way:

$$[(u,v).(w,y)]x = (u,v)x.(w,y)x \qquad \text{for any } x$$

This means that

$$(u,v).(w,y) = (u.w, v.y)$$

Calculating the Cayley table and relabeling the elements that arise as: 1 for (1, 3), 2 for (3, 2), 3 for (2, 1), 4 for (1, 2), 5 for (z, l), 6 for (2, z), and 7 for (3, 3) gives the same table as that of note 4 and this shows that the dcs is $C_2 \times C_2 \times C_2$.

7. For r = 3 the 7 contexts are:

(1, 2, 3), (1, 4, 5), (1, 6, 7), (2, 4, 6), (2, 5, 7), (3, 4, 7), (3, 5, 6).

More generally, there are r^* elements so the first number of a context can be chosen in r^* ways, the next in $r^* - 1$, the third is then determined. But this could have arisen in 3! = 6 ways, so the number is

$$r^*(r^* - 1)/6 = \frac{1}{3} r^*(r - 1)^*$$

since obviously $r^* - 1 = 2(r - 1)^*$.

8. For example, the dcs [1] can be labeled by the map (1, *a*, *b*) where *a*, *b* are restricted in this way: $a \neq 1$ (or the map would be singular) $a \neq 2$ (2 would then be invariant) and so on:

Element	1	2	3	4	5	6	7
becomes	1	*a*	1.*a*	*b*	1.*b*	*a*.*b*	1.*a*.*b*

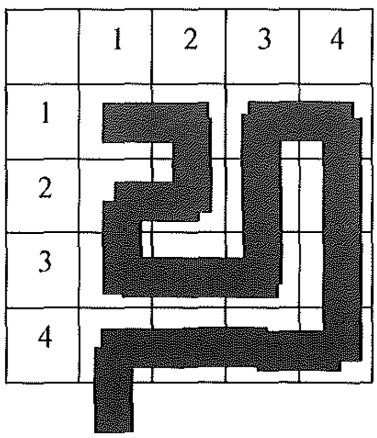
So in all $a \neq 2,\ 1,\ b,\ 6.b,\ 1.b$; $b \neq 4,\ 1,\ a,\ a.6,\ 1.a$. Possible pairs then begin with $a = 3, 4, 5, 6, 7$ and, in turn, if $a = 3$ then $b \neq 3,\ 5,\ 2$ so b is 6 or 7 and continuing in this way gives 14 possible pairs.

9. For example, note 8 shows that (1, 3, 6) and (1, 3, 7) both label the dcs [1], but (1, 3, 6).(1, 3, 7) = (*z*, *z*, 1) instead of (*z*, *z*, *z*). One could say that Frederick's construction in its later stages is about the labels of dcss, not the dcss themselves. There are $3^3.14^3 = 74088$ ways of labeling the 7 dcss but not all give rise to 127 elements at the next level. In fact, only 61160 do so, leaving 12928 cases when 63, 31 or 15 elements result.

As an example consider the labeling of the 4 dcss [1], [1, 4, 5], [1, 2, 3] and [ALL]. One set of labels would be (1, 3, 6), (1, 3, 4), (1, 2, 6) and (1, 2, 4) and these discriminate to (z, z, z). Hence if this set of labels is present, less than the full set of 127 elements will result from discriminate closure. Frederick required the *choice* of one of the 61160 sets but this is impossible from the process point of view.

10. In ODP "essentially different" was used for dissimilar. The whole subsequent discussion there was regrettably cursory compared with what is given here.

11. Label the squares as pairs of coordinates, so the path in the text starts from 1,1 to 1,2 → 2,1 → 3,1 →... . If successive determinations are numbered off 1, 2, 3, ... the function $a,b \rightarrow n$ is obviously recursive. If the path 1,1 → 2,1 → 1,2 → 1,3 →... were chosen instead, the resultant structure with the same process rules, would obviously be isomorphic (an "anti-isomorphism"). It is otherwise with a path "with corners" like:



as is proved in note 12.

12. It is now possible to take up the question of isomorphisms from note 11. The path with corners, shown there, gives a Cayley table

	1	2	3	4	5	6
1	z	3	6	2	4	y
2	4	z	5	1	3	y
3	2	1	z	y	6	5
4	6	5	y	z	2	1
5	y	4	2	6	z	3
6	3	y	1	5	4	z

$3 \cong 4,\ 1 \cong 5,\ 2 \cong 6$

To compare this with Q* it is best to begin by interchanging 5 and 6 (and the corresponding rows and columns). This gives:

	1	2	3	4	5	6
1	z	3	5	2	4	y
2	4	z	6	1	y	3
3	2	1	z	y	6	5
4	5	6	y	z	1	2
5	3	y	1	6	z	4
6	y	4	2	5	3	z

Here of course 5 and 6 do not mean the same as before.

$3 \cong 4,\ 1 \cong 6,\ 2 \cong 5$

This table respects duality just as Q* does, so use * to denote dual here to get an abbreviated form:

	1	2	3
1	z	3	2*
2	3*	z	1*
3	2	1	z

I shall now show that this is not isomorphic to Q*. For the only possible isomorphisms are interchange of rows and columns and the renaming of u as u*. Now consider all possible 3x3 tables with three stars in them. Of course, duality requires 3, or, for r generators,

$$\frac{1}{2} r * (r-1) *$$

A convenient notation for these would be to treat * as 1 in Z_2 and number off the columns, so that Q* would be 241, whereas the new table is 203. Then the eight possible schemes fall into 2 groups: 241 and 412 are each isomorphic to Q* but 013, 051, 203, 640, 602, 450 are all isomorphic to each other but not to Q*.

13. One such (possibly not the only one) is:

	1	2	3	4	5	6	7
1	z	3	2*	5	4*	7*	6
2	3*	z	1	6	7*	4*	5
3	2	1*	z	7*	6*	5	4
4	5*	6*	7	z	1	2	3*
5	4	7	6	1*	z	3*	2*
6	7	4	5*	2*	3	z	1*
7	6*	5*	4*	3	2	1	z

That this is symmetric can be seen by listing seven contexts, one from each similarity class and verifying that each is Q* as in note 12.

14. It is sufficient to exhibit this in the case $r = 4$. Further extension will be obvious. Take the 4 generators as 1, 2, 4, 8. There are the 4 (i.e. r) basic primals, (1 2 4 | 3 5 6 7), (1 2 8 | 3 9 10 11), (1 4 8 | 5 9 12 13), (2 4 8 | 6 10 12 14) and 11 (i.e. $r^* - r$) other initial ones. Any two initial primals have overlap of 3 (i.e. $(r - 2)^*$) elements and so together a set of 2.4 = 8 (i.e. $2[(r - 1)^* - (r - 2)^*] = 2^{r-1}$) other elements. The binary operation is defined in just the same way. It is instructive to start with the four initial primals above and verify that the set of all products gives exactly $C_2 \times C_2 \times C_2 \times C_2$.

15. A convenient notation here is to use 1 to represent dual and 0 to represent leaving the element unchanged.

So one can write, for instance, $(p, q, r)^a = (1\ 1\ 1)(p, q, r)$. Then treat the 1 1 1 as an element of Z_2 and so define the operator [7] = (1 1 1), and $u^a = [7]u$. The duals for each context are then evidently:

a	*b*	*c*	*d*	*e*	*f*	*g*
7	6	5	1	2	3	4

To make this quite clear, $u^c = [5]u$ means that $(p, q, r)^c = (1\ 0\ 1)(p, q, r) = (p^*, q, r^*)$.

The calculation of the Cayley table then begins:

	1	2	3	4	5	6	7
1	z	3	2^a	5	4^b	7	6^d
2	3^a	z	1	6		4^c	
3	2	1^a	z				
4	5^b	6^c		z	1	2	
5	4			1^b	z		
6	7^d	4		2^c		z	1
7	6					1^d	z

Now consider the remaining places defining the other 3 contexts. Take *e* first. Then

$$2.5 = (3, 2, 1^*).(z, 1^*, 1^*) = (3, 3, z) = 7^g$$
$$5.2 = (7^g)^e = 7^b$$

The first of these results is not 7 because we chose to define 7 as 1.6 and not 2.5. The Cayley table for the e-context is:

	2	3	7
2	z	7^g	5^b
3	7^b	z	2^g
7	5^g	2^b	z

Perhaps we should verify that this is indeed Q*. For this purpose, replace 5 by a new element $q = 5^g$. Then the table becomes

	2	3	7
2	z	7	q^b
3	7^e	z	2
7	q	2^e	z

which is evidently Q*. The remaining contexts are dealt with in the same way, so giving the complete Cayley table which the reader is invited to complete.

' P u C H '

A Purely 'Combinatorial' Combinatorial Hierarchy

'Only what we can Count is accountable'

. . . A Report on Work in Progress . . .

John Amson

Anstruther, KY10 3DY, Fife, Scotland : john.amson@uwclub.net

[20110-Jun-04 : 1800 Ver : V12.S1.6]

On the one-thousandth-eight-hundredth-and-thirty-sixth night* the Princess Scheherazade was again with the King :

"Auspicious King, let me ask you another riddle : What is it that is greater than everything, the dead eat it, and if the living eat it, they die?".

"Wait !", cried the King, "Your riddle has no answer. Nothing is greater than everything; and it is blasphemous to say anything else !"

"My King, I do not blaspheme", replied Scheherazade, "and you have just answered my riddle".

(*) And there are approximately 1836 electron masses in a proton's rest mass.

ABSTRACT

A new Combinatorial Hierarchy [**PuCH**] is introduced to replace the 'ANPA classical' one [**CH**]. It has no specific algebraic or discriminatory-process structure, only a combinatorial one. It is constructed by the simplest iteration : each new Level is just the set of all non-empty subsets of the previous Level. The Initial Ur-Level is a set with just 2 elements : one = 'nothing', the other = 'the set containing nothing'. The Levels then have the familiar **CH** sizes : 2, $2^2-1 = 3$, $2^3-1 = 7$, $2^7-1 = 127$, $2^{127}-1 \approx 10^{38}$ (these integers are all Mersenne Prime Numbers). The iteration continues until the next candidate Level has a non-prime number of members (*cf.* Parker-Rhode's 'linear independence' algebraic 'Stopping Rule'). Similarities are noted with the Gödel Constructible Universe (1940), and the von Neumann Universe, or Hierarchy of Sets, sometimes called 'The Cumulative Hierarchy'(c.1955) both of which are far earlier than the Parker-Rhode's Combinatorial Hierarchy (1965). Every physically interesting non-dimensional constant (ratio) can be identified with some pair of subsets (in some Level) the ratio of whose sizes is the physical constant in question to <u>precisely</u> the current known empirical accuracy. The rôle of the extraordinarily large number of 'Prime Subsets' is conjectured to be structurally important (in a physical, biological or information context) but in a direction not yet understood.

---

1 Prologue.

In the beginning and at the ending was Nothing.

'Nothing'. The paradox of Nothing.

As soon as we have Nothing we have Something — Nothing.

The Void. The 'Empty Set'.

The **Ur** set.

Having Nothing we have Something. A new set. A non-empty set.

The set that contains Nothing, the set that contains the Empty Set.

$$\{\ \emptyset\ \}$$

We shall label it $\mathbf{S}_0$.

How many members has the Empty Set : *None*

How many members has the Set that contains Nothing : *One*

We are now in business — we can now start to build our Hierarchy.

Simply by "COUNTING", just "COUNTING".

....

We shall need only to be able to count and to be able to appeal to these four Axioms from naive Set Theory

[1] `Every non-Empty set contains a subset.`

[2] `Every non-Empty set contains at least two subsets :`
`the Empty Set and the set of all its members.`

[3] `There are at least two sets and we can form their union set.`

[4] `To every set is associated a unique cardinal number :`
`the 'count' of how many members it has.`

As usual, we call the set of all subsets of a set S the *Power Set* $\mathcal{P}(\mathrm{S})$ of S.

Unlike as usual we shall call the Power Set which contains only those subsets of a set S which have a <u>countable</u> number of members (*i.e.*those with the Empty Set excluded since counting starts with '1') the *Countable Power Set* $\mathcal{P}^\star(\mathrm{S}) = \mathcal{P}(\mathrm{S}) \setminus \emptyset$.

For 'count sizes' we have $|\mathcal{P}(\mathrm{S})| = 2^{|\mathrm{S}|}$, and $|\mathcal{P}^\star(\mathrm{S})| = 2^{|\mathrm{S}|} - 1$.

2 Counting.

Start counting : [1]

S_0 above is a set with 1 member. It has 2 subsets : the Empty Set $\emptyset$, and the set of all its members, *i.e.* its only member $\emptyset$, *viz.*: $\{\emptyset\}$. So its 'set of subsets' has 2 members; it is a new set, $\left\{ \emptyset, \quad \{\emptyset\} \right\}$.

If we expelled the 'Empty Set' member $\emptyset$ we would be back again with a "set with a single element". This would stifle any new iterative generative 'counting' process. So we retain that 'Empty Set' member here (and only here).

We shall give to this new set of subsets of S_0 the label $\mathbf{S}_1$.

$S_1 = \left\{ \emptyset, \quad \{\emptyset\} \right\} = \mathcal{P}(S_0)$.

Its counting size is $|\mathbf{S}_1| = 2$. This is a Prime Number.

....

Continue counting : [2]

S_1 above is a set with 2 members. It has 4 subsets. (Note: $4 = 2^2$)

(1) the Empty Set $\emptyset$ — but it has no members waiting to be counted,

(2) the subset with the member $\emptyset$, *viz.*$\{\emptyset\}$,

(3) the subset with the member $\{\emptyset\}$, *viz.*$\{ \{\emptyset\} \}$,

(4) the subset consisting of both members, *viz.*$\{ \emptyset, \{\emptyset\} \}$

Having more than 2 subsets we can now expel the 'Empty Set' subset $\emptyset$ and retain the other 'non-Empty' subsets, without regressing to a single member set.

We shall give to this new set of subsets of S_1 (excluding the $\emptyset$) the label S_2. Thus $S_2 = \left\{ \{\emptyset\}, \{ \{\emptyset\} \}, \{ \emptyset, \{\emptyset\} \} \right\} = \mathcal{P}^\star(S_1)$.

Its counting size is $|S_2| = 3 = 2^2 - 1 = 2^{|S_1|} - 1$.

In binary notation it is 11.

This is a Prime Number, the first Mersenne Prime, $\mathbf{M}_2$,

and the first Catalan-Mersenne prime `CM(1)` (see below).

`REMINDER` A 'Mersenne Number' is a number of the form $2^p - 1$; if it is prime then it is called a 'Mersenne Prime' (and denoted by M_p) and then p must also be a prime. But a number $2^{prime} - 1$ need not be prime; *e.g.* $2^{11} - 1 = 2047 = 23 \times 89$. [See Wikipedia for more information.]

....

Continue counting : [3]

S_2 above is a set with 3 members. It has $8 = 2^3$ subsets, including the Empty Set. We will not display them here. Having more than 2 subsets we can again expel the 'Empty Set' subset (with no members) without regressing to a single member set again.

We shall give to this new set of subsets of S_2 (excluding the $\emptyset$)

the label $S_3 = \mathcal{P}^*(S_2)$.

Its counting size is $|S_3| = 7 = 2^3 - 1 = 2^{|S_2|} - 1$.

In binary notation it is 111.

This is a Prime Number, the second Mersenne Prime M_3,

the first Double Mersenne Prime M_{M_2},

and the second Catalan-Mersenne Number $\mathtt{CM}(2)$.

REMINDER A Double Mersenne Number is a number of the form $M_{M_p} = 2^{M_p} - 1$ (a 'Mersenne of a Mersenne Number'); if it is prime then it is called a Double Mersenne Prime. If a Double Mersenne Number is the 'Mersenne of a Double Mersenne Number' then it is called a *Catalan-Mersenne Number*. These are integers defined iteratively by the algorithm :

$$\mathtt{CM}(0) = 2; \quad \text{and for each} \quad n = 0, 1, 2, \ldots, \quad \mathtt{CM}(n+1) = 2^{\mathtt{CM}(n)} - 1. \qquad [\ddagger]$$

(which generates the sequence A007013

2, 3, 7, 127, 170141183460469231731687303715884105727, ...

in 'OEIS', the 'On-Line Encyclopedia of Integer Sequences' on the www). It is said that the French/Belgium mathematician Catalan (1814–1894) came up with this sequence after the discovery in 1876 of the primality of $M_{M_{(7)}} = 2^{(2^7-1)} - 1 = M_{(127)} = 2^{127} - 1$ by the French mathematician Lucas (1842–1891) (noted for his work on the Fibonacci numbers). Although the first five terms in the Catalan-Mersenne iterative sequence (up to $\mathtt{CM}(4) = M_{M_{(7)}}$) are prime, the general consensus appears to be that no known methods can decide if any more of these Catalan-Mersenne numbers are prime (in any reasonable time) simply because the numbers in question are too huge, unless a factor of the next one $\mathtt{CM}(6) = M_{M_{(127)}}$ is discovered.

....

And continue counting : [4]

S_3 above is a set with 7 members. It has $128 = 2^7$ subsets, including the Empty Set. We will not display them here. Having more than 2 subsets we can again expel the 'Empty Set' subset (with no members) without regressing to a single member set again.

We shall give to this new set of subsets of S_3 (excluding the $\emptyset$)

the label $S_4 = \mathcal{P}^\star(S_3)$.

Its counting size is $|S_4| = 127 = 2^7 - 1 = 2^{|S_3|} - 1$.

In binary notation it is `1111111`.

This is a Prime Number, the fourth Mersenne Prime M_7,

the second Double Mersenne Prime M_{M_3},

and the third Catalan-Mersenne Number `CM`(3) .

....

And still continue counting : [5]

S_4 above is a set with 127 members. It has 2^{127} subsets, including the Empty Set. We cannot possibly display them all here. Having more than 2 subsets we can again expel the 'Empty Set' subset (with no members) without regressing to a single member set again.

We shall give to this new set of subsets of S_4 (excluding the $\emptyset$)

the label $S_5 = \mathcal{P}^\star(S_4)$.

Its counting size is $|S_5| =$

$170141183460469231731687303715884105727 = 2^{127} - \; = 2^{|S_7|} - 1.$

In binary notation it is

```
1,1111111,1111111,1111111,1111111,1111111,1111111,
  1111111,1111111,1111111,1111111,1111111,1111111,
  1111111,1111111,1111111,1111111,1111111,1111111.
```

This is a Prime Number, the <u>twelfth</u> Mersenne Prime M_{127},

the <u>fourth</u> Double Mersenne Prime M_{M_7},

and the <u>fourth</u> Catalan-Mersenne Number `CM`(4).

....

Can we still continue counting ? : [6]

S_5 above is a set with 170141183460469231731687303715884105727127 members.

It has $2^{170141183460469231731687303715884105727}$ subsets, including the Empty Set.

'So how many's that, now, Jimmy ?'

'Yon's an awfee muckle heap, y' ken, tha' noo, Janet !'

But, having a *gey wee bittee* more than just 2 subsets we can again expel the 'Empty Set' subset without regressing to a single member set again.

We shall give to this new set of subsets of S_5 (excluding the $\emptyset$)

the label $S_6 \quad = \quad \mathcal{P}^\star(S_5)$.

Its counting size is $|S_6| =$

$2^{170141183460469231731687303715884105727} - 1 = 2^{|S_{127}|} - 1$.

In binary notation, well, it starts with

```
1,1111111,1111111,1111111,1111111,1111111,1111111,
  1111111,1111111,1111111,1111111,1111111,1111111,
  .......,.......,.......,.......,.......,........
```

.... and the printout goes on, and on, and on, for about 10^{33}km . . . . and way past the nearest star, Proxima Centauri, the red dwarf star in the constellation of Centaurus, which is a mere 4.2 light-years (3.97×10^{13} km) distant.

This is a Mersenne Number $M_{170141183460469231731687303715884105727}$.

And a Double Mersenne Number $M_{M_{127}}$,

And a Catalan-Mersenne Number $\mathbb{CM}(5)$.

Ah, bu' wait on Jimmy — is it a Prime Number now ?

Fair's fair, Janet — we canna tell one way or th'other — there's nae ony body oot there can get their bonny heid aroond it, there ah' fa' scunnered, and it looks richt set t' be playin' that way, till we all graw auld . . . — or cauld.

....

3 A new Combinatorial Hierarchy.

What we have just done is to construct a seemingly endless sequence of ever larger sets, by a simple process of 'emergence'.

The process starts with a primitive 'Ur-based' set S_0 consisting of nothing but the set containing the Empty Set : $\{ \emptyset \}$.

Using this process of counting, of enumerating — and by forming subsets and collecting them into new sets — we created a 'Purely Combinatorial Hierarchy" of sets.

No pre-empted notions of algebraic relationships or structures, no philosophical question-begging notions of pair-wise 'discriminations'.

Just the naive pre-conception of a set, and of a set having subsets, and of indentifying the set of all subsets (with or without the subset consisting of just the Empty Set).

In an earlier version[1] I called this Purely Counting (Combinatorial) construction '**PuCH**' — **CH** is a familiar abbreviation for 'Combinatorial Hierarchy'; **Pu** is short for 'Pure'. We shall continue to use that name here.

Here is a summary of the 'Counting Process' in the last section (apart from some minor modifications it also summarizes features in my earlier **PuCH**). As in all things **CH** we call the stages of the construction the 'LEVELs' of the Hierarchy :

LEVEL	many members 'size'		Catalan-Mersenne	Prime number
S_0	2	$= 2$	CM(0)	Yes
S_1	$2^2 - 1$	$= 3$	CM(1)	Yes
S_2	$2^3 - 1$	$= 7$	CM(2)	Yes
S_3	$2^7 - 1$	$= 127$	CM(3)	Yes
S_4	$2^{127} - 1$	$170 \ldots 727 \sim 10^{38}$	CM(4)	Yes
S_5	$2^{\sim 10^{38}} - 1$	?.?.?.?.?.?.?.?.?	CM(5)	?

TABLE 1. **PuCH** LEVELS

A seemingly important feature of the original **CH** construction was the Frederick Parker-Rhodes 'stopping rule' which prevented the construction of a Level S_5 in the **CH**. In my previous 2008 construction of **PuCH** a di ff erent

[1] *'PuCH', a Purely Combinatorial Combinatorial Hierarchy — or — A Midsummer Night's Dream ?*, presented ANPA 2008 in Cambridge.

'stopping rule' was proposed by me, one based on the fact that the there was simply insufficient 'physical time' to 'count' the huge number of members in Level $\mathbf{S}_5$ in **PuCH**. The candidate number was greater than my ULTIMOL, the 'largest possible integer that could be counted in any subjectively reasonable time.

I argued that in the worst case we cannot increment a ¡count¿ in any time shorter than *Max Planck's Second* which is of the order of $5.3912427 \times 10^{-44}$ seconds. That means that the greatest number we could count up to in **1 second** is of the order of 1.202×2^{144}

or about 2^{160} in one day of 86400 seconds

or 2^{168} in one Julian year of 365.25 days

or 2^{175} in one Julian Century of 36525 days

These numbers are vastly smaller than the number $2^{10^{38}}$ needed to count Level $\mathbf{S}_5$, and I concluded that 'no such Level $\mathbf{S}_5$ could therefore exist in any "constructible sense" '.

It has recently been pointed out to me (by Louis Gidney, in a private communication, 2010) that there might also be another less 'subjective' argument for 'stopping' the growth of my **PuCH** system : namely, that the three Double Mersenne Prime Numbers $\mathrm{M}_{\mathrm{M}_{(2)}} = 2^3 - 1$, $\mathrm{M}_{\mathrm{M}_{(3)}} = 2^7 - 1$, $\mathrm{M}_{\mathrm{M}_{(7)}} = 2^{127} - 1$ are three of the only four Double Mersenne Numbers known to be prime. These are the sizes of the Levels $\mathbf{S}_2$, $\mathbf{S}_3$, $\mathbf{S}_4$.

The Double Mersenne Number $\mathrm{M}_{\mathrm{M}_{(5)}} = 2^{(2^5-1)} - 1 = 2^{31} - 1 = 2147483647$ is the other one known to be prime[2] but this does not seem to play have played any part in **PuCH** as yet.

But we have now recognized, above, that the sizes of all the first Levels $\mathbf{S}_0$, $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$, $\mathbf{S}_4$, are the Catalan-Numbers that are all known to be Prime, and that the 'Primeness' of the size of the next putative level $\mathbf{S}_5$, although the next Catalan-Number in the sequence, may not be 'testable' (simply because it is too large for practical test algorithms). If it should ever be tested and found to be Non-Prime, then this would provide a very significant 'finite-ness' feature of **PuCH**.

[2]See *e.g.* Wolfram MathWorld website at
`http://mathworld.wolfram.com/DoubleMersenneNumber.html`

In all immediate circumstances it would seem that the question of the Primeness of the size of Level S_5 must remain open.

Nevertheless, this does not prevent us from proposing :

NEW **PuCH** STOPPING RULE :
continue constructing new Levels from the old until the newest candidate fails to have a prime number of members.

....

It is useful to call a set whose 'size' (count of members) is a Mersenne Prime a *Mersenne Prime Set* [MP-set]. The Levels in **PuCH** are Mersenne Prime Sets. We immediately see that there are other Mersenne Prime Sets distributed amongst ours, as shown in the Table below.

MP-sets	n	Levels	members
(—	1	S_0	$= 2^\dagger$ seed)
$M_{(1)}$	$2^\dagger$	S_1	$= 3^\dagger$
$M_{(2)}$	$3^\dagger$	S_2	$= 7^\dagger$
$M_{(3)}$	5		$= 31^\ddagger$
$M_{(4)}$	$7^\dagger$	S_3	$= 127^\dagger$
$M_{(5)}$	13		$= 8191$
$M_{(6)}$	17		$= 131071$
$M_{(7)}$	19		$= 524287$
$M_{(8)}$	$31^\ddagger$		$= 2147483647$
$M_{(9)}$	61		$= 2305843009213693951$
$M_{(10)}$	89		$= 2^{89} - 1 \sim 10^{28}$
$M_{(11)}$	107		$= 2^{107} - 1 \sim 10^{34}$
$M_{(12)}$	$127^\dagger$	S_4	$= 2^{127} - 1 \sim 10^{38}$

TABLE 2. THE FIRST 12 MERSENNE PRIME SETS

(The 'Double Mersenne Primes' $M_{M_{(n)}}$ are marked with † or $\ddagger$.)

There are many other Prime Sets that are not Mersenne, but the Mersenne ones are those that are already the Set of Subsets (excluding the Empty Set) of some smaller Prime Set. For instance, 8191, the number of members in $M_{(5)}$ is the count, $2^{13} - 1$, of the number of non-empty subsets in the Prime Set with just 13 members.

....

4 Curious numerical ratios

Another especially promotional feature of the original **CH** was that certain numbers intrinsic to the construction (*e.g.* the Cumulative Counts above), or generated by curious algebraical manipulations, were numerically in agreement with certain experimental numbers to do with ratios of physical quantities. For example :

(i) 3 was the number of spatial dimensions in the observable physical world; and its reciprocal is of the order of a strong coupling constant[3];

(ii) 10 is also of the order of strong coupling constant[3], it also turned out to be the number of dimensions thought to be meaningful in the context of String Theory;

(iii) 137 was the integer part of the reciprocal of the Fine Structure Constant $\alpha = 137.035,999,068(96)$ that characterizes the Electromagnetic coupling constant;

(iv) 1.7×10^{38} was 'of the order of magnitude' of the reciprocal of the ratio of the Gravitational to the Electromagnetic force.

What was never made clear was why it was only these four numerical 'coincidences' that emerged from the **CH** construction, and why there was no obvious mechanism or intrinsic structure within the construction that could supply addition numerical coincidences.

If numerical 'coincidences' between on the one hand, (a) the sizes of certain subsets in the construction, and on the other, (b) the empirical numerical values of certain 'physical scale constants' (or their ratios), is judged to be an indication of the metaphysical significance of the **CH** structure, then we have to ask ourselves this question :-

[3]Ted Bastin (in *On the Origin of the Scale Constants of Physics.* Bastin, E.W., Studia Philosophica Gandensia 4, 1966, Gent. | a TEXversion is available on request from John Amson |) says that "Reference to the table at the end of §3 shows that the theory provides for two types of strong interaction, associated with couplings 1/3 and 1/10. There is considerable doubt as to the right way to define strong coupling, and all one can say is that these constants are of the right order of magnitude[....]".

Question : "Can we now productively 'data-mine' my new purely combinatorial **PuCH** for these and other other numbers ? Other ratios ? Using Combinatorial Tools and only Combinatorial Tools" ?

The answer is : "Indeed we can. And there are some immediately obvious tools" :

C1 Each level can have its own Census of how many ways a subset with a given number of members can be selected within the Level : this Census is none other than the familiar listing of the relevant binomial coefficients $\binom{n}{r}$. There are $\lfloor n/2 \rfloor = n/2$ or $(n-1)/2$ distinct such numbers (depending on n even or odd).

C2 Each level can have a $2nd$ catalogue of the ratios of those binomial coefficients :

$$\binom{n}{s} \div \binom{n}{r} = \frac{n!}{(n-s)!\,s!} \times \frac{(n-r)!\,r!}{n!} = \frac{(n-r)!\,r!}{(n-s)!\,s!} \qquad [\star]$$

There are $\frac{1}{2}\,m\,(m-1)$ distinct such ratios ($m{=}\lfloor n/2 \rfloor$ as above).

C3 Then again, in an imaginatively iterative process — indeed, a quasi-fractal process — each subset in each Level can be regarded as a 'sub-level' which of course has its own Power set. And to each such 'sub-level' the previous two cataloguing tasks can again be applied. There are multitudes of combinatorial counts and ratios to be explored !

In a possibly more significant approach we might include the requirement that the subsets to be counted in C1 shall all be Prime Subsets. Such a **Prime Census Programme** would seem to be singularly difficult. It might prove to be impossible in any but a purely empirical/computational manner involving the testing for primality of each and every candidate Subset (*e.g.* by the use of possibly immense look-up tables of known prime numbers).

4.1 Practical computational barriers and evasions.

Utilizing numerical software these tasks are not difficult when only the lower Levels ($S_0 \ldots S_3$) are involved. It is impractical (impossible?) to do this in any meaningful way when we attempt to work with the last Level S_4. At this Level, where n is the order of 10^{38}, the quantity of ratios $\binom{n}{s} \div \binom{n}{r}$ is of the order of $\frac{1}{2}(10^{38})^2 \approx 10^{76}$ — the 'number of pairs of atoms in the universe', beyond even our `Ultimol` ! Even when the values of r and s are artificially restrained to be small of the order of, say, a few thousand, severe computational barriers arise. However, it can be shown, using *e.g.* analytic tools such as Stirling's approximation $n! = \sqrt{2\pi}\, n^{n+1/2}\, e^{-n}$ and Binomial approximations, that the ratio of counts at $[\star]$ can be reduced to this more useful approximation

$$\binom{n}{s} \div \binom{n}{r} \quad \approx \quad (1 + s - r) \times \sqrt{\frac{s}{r}} \times \frac{r^r}{s^s} \times n^{s-r} \qquad [\star\star]$$

With this, together with numerical software explorations, we can readily — if somewhat heuristically — convince ourselves that these prototypical ratios of the counts in $[\star]$ can vary over immense ranges as densely as one pleases.

But even without any such analytical assistance it becomes glaringly obvious that any *a priori* given physical non-dimensional scale constant, or whatever, could be matched, to any desired precision, simply by searching for the ratio of the counts of members in a suitable pair of subsets. In Level-Four there are subsets with as many members as you like, all the way from $1, 2, 3 \ldots$ up to the maximum, that huge number $2^{127} - 1$. We require no 'magic selection process' here. (Though we readily concede that the selection of Prime Subsets could prove to be far more difficult and elusive.)

There are two cases to consider.

(A) Whenever the physical non-dimensional scale constant is knowable only empirically as a finite decimal number, then the latter's representation as a vulgar fraction is enough ;

(B) Whenever the physical non-dimensional scale constant has arisen by a conceptual 'leap of deduction' and is identifiable with a 'real non-rational number' then the latter's representation by a Continued Fraction Convergent (a vulgar fraction) to any desire or permitted level of approximation is enough .

For two very simple Examples in Case (A) *we have* :

(i) there are subsets of Level $\mathbf{S}_4$ with : $s = 183615267261$ and $r = 100000000$ members, and with their ratio s/r precisely equal to 1836.15267261 — the ratio m_p/m_e ;

(ii) and there are two other subsets with : $s = 6851799983997$ and $r = 50000000000$ members, with their ratio s/r precisely equal to 137.03599967994 — the NIST value of the reciprocal of the Fine Structure Constant α . Precisely to every known decimal digit.

(Both of which raise the obvious doubt – *'So what ?'*)

And for two less simple Examples in Case (B) *we have* :

(iii) there are subsets of Level $\mathbf{S}_4$ with $s = 175568277047523$ and $r = 124145519261542$ members, with their ratio s/r differing by less than 10^{-28} from $\sqrt{2}$ — the incommensurable (irrational) 'length' of the diagonal of a square with sides of unit length.

(iv) there are subsets of Level $\mathbf{S}_4$ with: $s = 5371151992734$ and $r = 1709690779483$ members, with their ratio s/r differing by less than 10^{-24} from π — the incommensurable (transcendental) 'length' of the circumference of a circle with diameter of unit length.

It would seem to be an open question as to whether, from our present view point, there is a substantive (essential ?) difference between the last two 'non-dimensional scale constants', arising in the context of a physical observable universe, and any of those other 'non-dimensional recognizable real numbers' arising in a purely mathematical context[4] (*e.g.* the exponential constant $e = 2.718281828\ldots$ or the Euler-Mascheroni constant $\gamma = 0.577215664901\ldots$).

5 Historically previous invocations.

The idea of a hierarchy of sets each of which is the set of all subsets of the preceding set in the hierarchy is not new. As was pointed out to me (by Louis Gidney, 2008, in another private communication directing me to the

[4]Finch, Steven R., 2003, Mathematical Constants, CUP.

relevant Wikipedia web entries), the idea had certainly been described many years before the Classical Combinatorial Hierarchy had first been adumbrated by Frederick Parker-Rhodes in the 1960s.

As long ago as 1940 the very closely-related idea of *Gödel's constructible universe* had been proposed[5].

This was a slightly more elaborate formulation of a somewhat simpler system [6] introduced by von Neumann.

The latter, von Neumann's *Cumulative Hierarchy* (or *the von Neumann Universe*), is a set **V** constructed recursively, in almost the same way as I have constructed the much 'smaller' (finite) and hence simpler **PuCH**.

* We start with a set $\mathbf{V}_0$, the empty set $\emptyset$.

* For any set X we denote by $\mathcal{P}(X)$ the power-set of X.

* For ordinal numbers α and β, we construct the union of power-sets :

$$\mathbf{V}_{\alpha+1} \stackrel{\text{def}}{=} \bigcup_{\beta<\alpha} \mathcal{P}(V_\beta) \qquad \text{('cumulation')}.$$

* Finally, we construct the union of unions :

$$\mathbf{V} \stackrel{\text{def}}{=} \bigcup_{\alpha} V_\alpha$$

As the article on 'The von Neumann Universe' in Wikipedia says :-

> If ω is the set of natural numbers, then $\mathbf{V}_\omega$ is the set of 'hereditarily finite sets', which is a model of set theory without the axiom of infinity. $\mathbf{V}_{\omega+\omega}$ is the universe of "ordinary mathematics", which is a model of Zermelo set theory. If ω is an inaccessible cardinal, then $\mathbf{V}_\omega$ is a model of Zermelo-Fraenkel set theory itself [....]'.

The denial of the 'axiom of infinity' in $\mathbf{V}_\omega$ makes it of special interest to the finitary constructivist nature of **PuCH**, and the way it abhors infinity.

[5]Gödel, Kurt, 1940, Consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set theory. Princeton Univ.Press. For a modern account see *'The Hierarchy of Constructible Sets'* in Ch.13 in Thomas Jech, *Set Theory : 3rd Millennium Edition, revised and expanded.*, Springer-Verlag, 2003.

[6]Neumann, John von, The von Neumann universe, or The von Neumann hierarchy of sets, sometimes called The Cumulative Hierarchy. [Reference needed. But the date has to be before JvN's death in 1957. The **CH** was invented in 1964 many years later !]

We note two places where **PuCH** differs the 'von Neumann Universe' :

PuCH starts with a doubleton set $\{\ \emptyset, \quad \{\emptyset\}\ \}$, not the Empty Set itself.

PuCH uses regular-power-sets (power-sets with the Empty Set excluded) in place of the standard power sets used here.

5.1 Unions and Cumulations.

The 'von Neumann Universe' uses set unions. This was also a feature introduced with little explanation by Parker-Rhodes in his original version of the **CH**. Moreover, his 'Unions of Levels' invariably and somewhat inexplicably always excluded the shadowy 'Zero-th Level' with those two initializing 'seed' entities which were always signified by '0' and '1'.

Thus in the classical **CH** we have this Unionization :

LEVEL UNIONS	members sizes	Cumulative Sizes	Prime?
$\mathbf{U}_1 = \mathrm{S}_1$	3	3	Yes
$\mathbf{U}_2 = \mathrm{S}_1 \cup \mathrm{S}_2$	$3+7$	10	No
$\mathbf{U}_3 = \mathrm{S}_1 \cup \mathrm{S}_2 \cup \mathrm{S}_3$	$3+7+127$	137	Yes
$\mathbf{U}_4 = \mathrm{S}_1 \cup \mathrm{S}_2 \cup \mathrm{S}_3 \cup \mathrm{S}_4$	$3+7+127+\sim 10^{38}$	$\sim 10^{38}$	?

TABLE 3. **CH** LEVEL UNIONS

However, in the new **PuCH** we have this more inclusive and more rigorous Unionization :

LEVEL UNIONS	member sizes	Cumu. Sizes	Prime?
$\mathbf{W}_0 = \mathrm{S}_0$	2	2	Yes
$\mathbf{W}_1 = \mathrm{S}_0 \cup \mathrm{S}_1$	$2+3$	5	Yes
$\mathbf{W}_2 = \mathrm{S}_0 \cup \mathrm{S}_1 \cup \mathrm{S}_2$	$2+3+7$	12	No
$\mathbf{W}_3 = \mathrm{S}_0 \cup \mathrm{S}_1 \cup \mathrm{S}_2 \cup \mathrm{S}_3$	$2+3+7+127$	139	Yes
$\mathbf{W}_4 = \mathrm{S}_0 \cup \mathrm{S}_1 \cup \mathrm{S}_2 \cup \mathrm{S}_3 \cup \mathrm{S}_4$	$2+3+7+127+\sim 10^{38}$	$\sim 10^{38}$	?

TABLE 4. **PuCH** LEVEL UNIONS

Had Parker-Rhodes not excluded the 2-membered 'Zero-th Level' his cumulative counts would have not delivered the suggestive sequence 3, 10, 137, ..., but instead the sequence 2, 5, 12, 139, ... But this deliberate exclusion of the 2-membered 'Zero-th Level' has always raised doubt as to whether this was not really 'construction by hindsight'.

5.2 Heuristic musings on Prime Subsets.

By definition, a Prime Subset in **PuCH** is a set whose count size (cardinality) is a Prime Number.

Each of their five Levels S_0 (2), S_1 (3), S_2 (7), S_3 (127), and

S_4 (170141183460469231731687303715884105727), is a Prime Subset.

Every one of the myriads of Prime Subsets in **PuCH** is identified with a group (unique up to isomorphism), namely the cyclic group C_p where p is the count size of the Prime Subset — and hence can be visualized as a 'Necklace' (or Hoop[7]). Every (prime) Necklace with more than 7 beads can be partitioned (*cf.* Goldbach's Conjecture, which is verified for all prime numbers far greater than the size of our Level S_4) into the disjoint union of three smaller (prime) Necklaces in possibly very many different ways. These are not 'sub-Necklaces' of course, because a cyclic group of prime order has no non-trivial subgroups.

And all, except the seed Level S_0, not having even order, these (prime) Necklaces cannot even be seen as 'dihedral' — they do not have even 'flip-over' symmetry.

This is in stark contrast to every other non-prime subset in **PuCH**, many of which can be visualized as polyhedral and polytopic structures in many-dimensional spaces, exhibiting far more elaborate symmetries.

As 'necklaces', we can visualize the seed level base Level S_0 as a dumbbell necklace with two beads. Surrounding this is a S_1, a triangular necklace with 3 beads. This in turn is surrounded by a S_2, a heptagonal necklace with 7 beads. Which is surrounded by S_3 (127), a polygonal necklace with 127 beads. Then any attempt at visualization collapses: the top level S_4, surrounding all the others, is a very large necklace. If it were possible to imagine it with beads only a millimetre big then its diameter would be $\sim 10^{19}$ Light-Years. The 'Observable Universe' is thought to be a sphere of radius only $\sim 10^{10}$ Light-Years.

It is difficult to conceive of any systematic treatment that would do justice to the immensely elaborate networks of Prime and Non-prime Subsets ('Necklace' and 'Non-necklace' structures) pervading **PuCH**.

[7]See *e.g. A Combinatoric Bit-Hoop System*, John Amson, Keith Bowden, Proc. ANPA 2005 p.xxx-xxx.

ACKNOWLEDGEMENTS

I am grateful to Louis Gidney for his sustained and perceptive interest in these explorations and his many thoughtful observations, and to Keith Bowden for his willingness to read an earlier version of **PuCH** at the 2008 ANPA Conference in Cambridge for me when I was unable to be present to read it in person. It was Richard Smullyan's book *The Riddle of Scheherazade and Other Amazing Puzzles* that suggested my introductory riddle.

slant200 slant200

GOLDBACH PUCH AND CH

YET ANOTHER NEW COMBINATORIAL HIERARCHY

'WITHIN EVERYTHING IS THE SEED OF EVERYTHING'

JOHN AMSON

Anstruther, KY10 3DY, Fife, Scotland : john.amson@uwclub.net

[2011-Jan-10 : 1105 Ver : G2.F4]

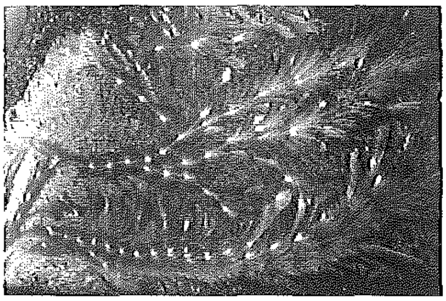


Figure 1: FROST FLOWERS CRYSTALLIZING ON A WINDOW ('BOTTOM-UP' GROWTH)

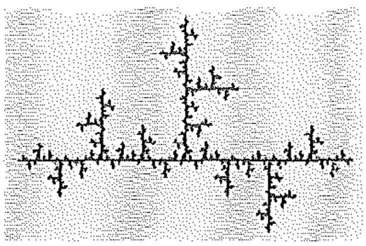


Figure 2: A GOLDBACH–TRIPLE DENDRITIC–FRACTAL STRUCTURE ('TOP-DOWN' GROWTH)

ABSTRACT

Continuing on the development of my Purely Combinatorial Hierarchy Set Structure called **PuCH**, use is made of Goldbach's Conjecture (valid for numbers far larger than any appearing in the subsets of **PuCH**) to introduce the novel idea of the '**PuCH** Goldbach Triple Partition Principle' : 'Every Odd Subset of every Level of **PuCH** higher than Level $\mathbf{S}_1$ can be partitioned into the disjoint union of three Prime Subsets.' Various illustrations are given, and an 'Iterative Goldbach Partitioning' structure is introduced into **PuCH**. The dual notion of an 'Iterative Goldbach Spreading' structure is then suggested for **PuCH**. The new structure is re-named **GoCH**. The fact that any Prime Subset in **GoCH** is identifiable with a 'Necklace' (*alias*'Hoop') as a cyclic prime group with a prime number of 'beads' suggests that **GoCH** is identifiable with a vast dendritic-like structure of chains of 'Triple-Prime-Necklaces'. It is suggested that unlike the classical **CH** (whose applications were apparently restricted solely to particle physics) this more-simply constructed set and number theoretic structure **GoCH** could have applications in biology and information-communication also.

This paper is a supplement to my paper
'PuCH' A Purely 'Combinatorial Combinatorial Hierarchy
of which a first version was submitted to the 2008 ANPA Conference in Cambridge, and a revised version appears in this current issue of the ANPA Proceedings 2010.
The section Numbering, *etc.*, continue from that previous paper.

---

6 Goldbach Triple Prime Subset Partition of PuCH Levels.

The previous section **5.1** (Unions and Cumulations) showed the 'Cumulative Counts' of the number of members in the successive Unions of the **PuCH** Levels. In this section we consider not 'Unionizations' but 'Partitions'. And because of the feature of both **CH** and **PuCH** that all their Levels $\mathbf{S}_0$, $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$, $\mathbf{S}_4$, have a Prime Number of members, we enquire into the extent that these Levels can be partitioned into the union of Prime Subsets (subsets whose count size is Prime).

An article in Wikipedia[1] tells us : *On 7 June, 1742, the German mathematician Christian Goldbach of originally Brandenburg, in Prussia, wrote a letter to Leonhard Euler (letter XLIII)[3] in which he proposed the conjecture:[. . . .]* — a modern version of which is:

EVERY ODD INTEGER GREATER THAN 5
CAN BE WRITTEN AS THE SUM OF THREE PRIMES.

We also read[2] that : *The conjecture that all odd numbers ≥ 9 are the sum of three odd primes is called the "weak" Goldbach conjecture.*
Vinogradov (1937ab, 1954) proved that every sufficiently large odd number is the sum of three primes (Nagell 1951, p. 66; Guy 1994), and
Estermann (1938) proved that almost all even numbers are the sums of two primes. Vinogradov's original "sufficiently large" $N \geq 3^{3^{15}} \sim e^{e^{16.573}} \sim 3.25\times10^{6846168}$ was subsequently reduced to $e^{e^{11.503}} \sim 3.33\times10^{43000}$ by Chen and Wang (1989).

The latter results confirm that all the Levels and their Unions have sizes ($\leq \sim10^{38}$) comfortably smaller than the currently known upper limit ($\leq \sim10^{43000}$) of the integers that satisfy Goldbach's Conjecture.

In other words, in **PuCH** we now have available a new principle :

> PuCH GOLDBACH TRIPLE
> PARTITION PRINCIPLE :
>
> Every Odd Subset of every Level of PuCH higher than Level S_1 can be partitioned into the disjoint union of three Prime Subsets.

[1] http://en.wikipedia.org/wiki/Goldbach_conjecture
and http://en.wikipedia.org/wiki/Goldbach's_weak_conjecture.

[2] http://mathworld.wolfram.com/GoldbachConjecture.html.

Example 1. Here is a sample of the Goldbach Triple partitioning of the prime number '137', the count size of the union $\mathbf{U}_3 = \mathbf{S}_1 \cup \mathbf{S}_2 \cup \mathbf{S}_3$ of three **PuCH** Levels. The first 54 partitions are into distinct primes and are followed by 8 more non-distinct partitions :

(1)	3 +	7 +	127 =	137
(2)	3 +	31 +	103 =	137
(3)	3 +	37 +	97 =	137
(4)	3 +	61 +	73 =	137
(5)	5 +	19 +	113 =	137
(6)	5 +	23 +	109 =	137
(7)	5 +	29 +	103 =	137
(8)	5 +	31 +	101 =	137
(9)	5 +	43 +	89 =	137
(10)	5 +	53 +	79 =	137
(11)	5 +	59 +	73 =	137
(12)	5 +	61 +	71 =	137
(13)	7 +	17 +	113 =	137
(14)	7 +	23 +	107 =	137
(15)	7 +	29 +	101 =	137
(16)	7 +	41 +	89 =	137
(17)	7 +	47 +	83 =	137
(18)	7 +	59 +	71 =	137
(19)	11 +	13 +	113 =	137
(20)	11 +	17 +	109 =	137
(21)	11 +	19 +	107 =	137
(22)	11 +	23 +	103 =	137
(23)	11 +	29 +	97 =	137
(24)	11 +	37 +	89 =	137
(25)	11 +	43 +	83 =	137
(26)	11 +	47 +	79 =	137
(27)	11 +	53 +	73 =	137
(28)	11 +	59 +	67 =	137
(29)	13 +	17 +	107 =	137
(30)	13 +	23 +	101 =	137
(31)	13 +	41 +	83 =	137
				...
cont.				
(32)	13 +	53 +	71 =	137
(33)	17 +	19 +	101 =	137
(34)	17 +	23 +	97 =	137
(35)	17 +	31 +	89 =	137
(36)	17 +	37 +	83 =	137
(37)	17 +	41 +	79 =	137
(38)	17 +	47 +	73 =	137
(39)	17 +	53 +	67 =	137
(40)	17 +	59 +	61 =	137
(41)	19 +	29 +	89 =	137
(42)	19 +	47 +	71 =	137
(43)	23 +	31 +	83 =	137
(44)	23 +	41 +	73 =	137
(45)	23 +	43 +	71 =	137
(46)	23 +	47 +	67 =	137
(47)	23 +	53 +	61 =	137
(48)	29 +	37 +	71 =	137
(49)	29 +	41 +	67 =	137
(50)	29 +	47 +	61 =	137
(51)	31 +	47 +	59 =	137
(52)	37 +	41 +	59 =	137
(53)	37 +	47 +	53 =	137
(54)	41 +	43 +	53 =	137
(55)	3 +	3 +	131 =	137
(56)	5 +	5 +	127 =	137
(57)	17 +	17 +	103 =	137
(58)	29 +	29 +	79 =	137
(59)	3 +	67 +	67 =	137
(60)	19 +	59 +	59 =	137
(61)	31 +	53 +	53 =	137
(62)	43 +	47 +	47 =	137

totals	928 +	2328 +	5238 =	8494
means	14.968 +	37.548 +	84.484 =	137

TABLE 5. GOLDBACH TRIPLE PARTITIONING OF PRIME NUMBER '137'

The distribution of how many times the Subsets with Prime summands 3, 5, ... occurred in those 54 uniquely different summations is given in the following Table 6 (which shows subsets with 'p' prime members and 'k[p]' counts of these subsets) and illustrated in Figure 3. This resembles a histogram

— the prime numbers are located along the x-axis, and the number of times they appear in Table 6 (frequency) is located up the y-axis.

p	2	3	5	7	11	13	17	19	
k[p]	-	4	8	7	10	5	11	5	
	23	29	31	37	41	43	47	53	
	10	7	5	6	7	4	8	7	
	59	61	67	71	73	79	83	89	
	6	5	4	6	5	3	5	5	
	97	101	103	107	109	113	127	131	137
	3	4	3	3	2	3	1	-	-

TABLE 6. TRIPLE 'DISTINCT' PRIME SUMMAND DISTRIBUTION FOR '137'

The prime 17 occurred the most (11 times), the primes 11 and 23 occurred the next most frequently (10 times each).

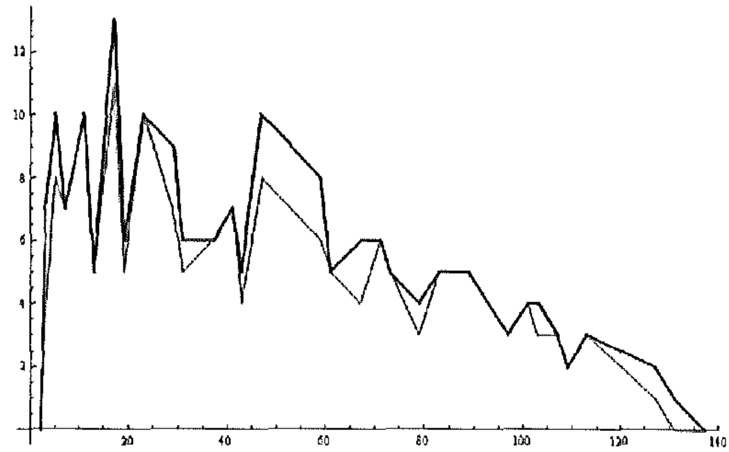


Figure 3: PRIME SUMMAND DISTRIBUTION FOR '137' 'DISTINCT' TRIPLES [FEINT:LOWER], AND INCLUDING 'NON-DISTINCT' TRIPLE [BOLD:UPPER]

Example 2. As another example, the Prime Subset with the Mersenne Prime $M_{(5)} = 8191$ number of members can be partitioned into three <u>distinct</u> prime subsets in 29,644 different ways, beginning with $3 + 17 + 8171 = 8191$ and ending with $2719+2731+2741 = 8191$. (I can supply the very long, full listing, if required, as a Text File). The way the prime summands are distributed is shown in Fig.4. Note the curious 'double banding concentrations', and the couple of dozen distinct 'peaks'.

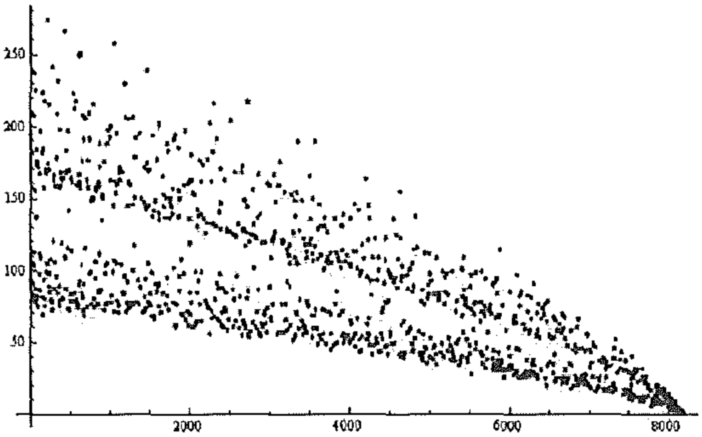


Figure 4: TRIPLE 'DISTINCT' PRIME SUMMAND DISTRIBUTION FOR THE MERSENNE PRIME SUBSET COUNT '8191' IN **PuCH** , THERE ARE 29,644 DIFFERENT TRIPLES.
[Primes along the x-axis; frequency of appearance up y-axis.]

6.1 Goldbach Partitioning into Mersenne Prime Subsets.

Despite the importance of Mersenne Prime Sets as the Levels of **PuCH**, the partitioning of a Prime Subset into three Mersenne Prime Subsets is surprisingly rare. Of the first 12 Mersenne Prime Subsets :

Level	MP-sets	n	members
S_1	$M_{(1)}$	$2^{\dagger}$	$= 3^{\dagger}$
S_2	$M_{(2)}$	$3^{\dagger}$	$= 7^{\dagger}$
	$M_{(3)}$	5	$= 31^{\dagger}$
S_3	$M_{(4)}$	$7^{\dagger}$	$= 127^{\dagger}$
	$M_{(5)}$	13	$= 8191$
	$M_{(6)}$	17	$= 131071$
	$M_{(7)}$	19	$= 524287$
	$M_{(8)}$	$31^{\dagger}$	$= 2147483647$
	$M_{(9)}$	61	$= 2305843009213693951$
	$M_{(10)}$	89	$= 2^{89} - 1 \sim 10^{28}$
	$M_{(11)}$	107	$= 2^{107} - 1 \sim 10^{34}$
S_4	$M_{(12)}$	$127^{\dagger}$	$= 2^{127} - 1 \sim 10^{38}$

TABLE 7. THE FIRST 12 MERSENNE PRIME SETS
(The 'Double Mersenne Primes' $M_{M_{(n)}}$ are marked with †.)

only the following six are currently known to be available :

$$
\begin{array}{ccccccc}
3 & + & 3 & + & 7 & = & 13 \\
3 & + & 3 & + & 31 & = & 37 \\
3 & + & 7 & + & 7 & = & 17 \\
3 & + & 7 & + & 31 & = & 41 \\
3 & + & 7 & + & 127 & = & 137 \\
3 & + & 127 & + & 127 & = & 257
\end{array}
$$

Even the disjoint union '3 + 3 + 127 = 133' is not a Prime Set.

A software census of all the $7\times7\times7 = 343$ triple unions that can be formed (and readily calculated) from the first 7 Mersenne Prime Sets shows that all but 18 of their 337 non-Prime unions have the very small common factor 3. For example, the disjoint union '$8191 + 131071 + 524287 = 663549 = 3 \times 29^2 \times 263$.

6.2 Iterative Goldbach Partitioning.

What is important about this new principle of partitioning Subsets in **PuCH** into the disjoint union of three Prime Subsets is that it is iterative.

Whenever we have partitioned any odd-sized Subset (*e.g.* Prime Subset) into a disjoint union of Prime Subsets it is obvious that each one of those three Prime Subsets may (if still sufficiently large enough) be itself partitioned into a further disjoint union of Prime Subsets and in many different ways (if still sufficiently large enough).

Each such iterated partitioning will eventually stop, and each will exhibit a simple or not-so-simple, perhaps a very highly complicated *quasi-Fractal* or *Dendritic Structure*[3] characterized as a sequence of prime numbers whose total sum is always the prime number count of the initial Prime Subset.

Here is a very simple (quite arbitrary) example :

- choose a 3-prime partition of 137 :

$$[13 + 17 + 107]$$

- choose a 1^{st}-SubTriple partition :

$$[\ \ [3 + 5 + 5] + [2 + 2 + 13] + [17 + 19 + 71]\ \ ]$$

- choose some 2^{nd}-SubTriple partitions :

$$[\ \ [3 + 5 + 5] + [2 + 2 + 13] + [\ \ [5 + 5 + 7] + [3 + 5 + 11] + [19 + 23 + 29]\ \ ]\ \ ]$$

- . . . and so on . . .

. . . . . . which we can also visualize as shown in Fig.5. (overleaf)

[3] See *e.g.* `http://en.wikipedia.org/wiki/Dendrite`

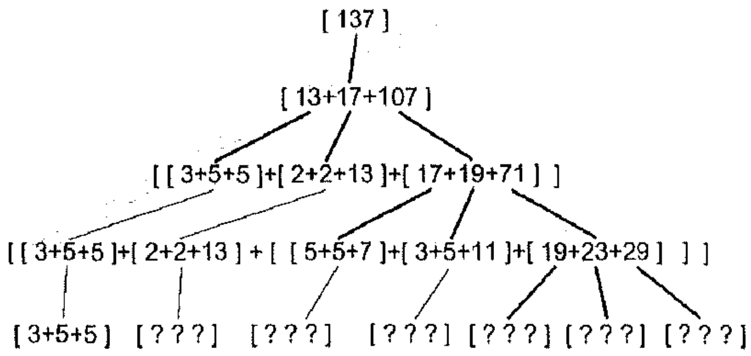


Figure 5: ITERATED TRIPLE PRIME PARTITION OF '137'

Clearly there are myriads of such 'dendritic 3-prime structures' latent within every Level of **PuCH** and threading down from higher Levels through all lower Levels.

It is a remarkably difficult Combinatorial Problem to find how many there are in **PuCH**, or even , say, in the union of just its first four Levels, excluding the huge top Level S_4 with its $\sim 10^{38}/2$ candidate 'starting nodes'...

And their relevance or significance is not yet apparent.

The analogy with communication (internet) and neural networks and biological systems is provocative, and a largely welcome shift of emphasis from the Particle / Quantum Physical milieu from which the **CH** originally sprang and by which it eventually became trapped, then enervated, and ultimately, after forty-six unproductive years, sadly sterile and self-serving.

In neurology, dendrites are the branched projections of a neuron that act to conduct the electrochemical stimulation received from other neural cells to the cell body, or soma, of the neuron from which the dendrites project.

In physical chemistry, a crystal dendrite is a crystal that develops with a typical multi-branching tree-like form. Dendritic crystallization forms a natural fractal pattern (see frontispiece).

6.3 Iterative Goldbach Spreading.

We have introduced the notion of 'top down' Goldbach Partitioning into **PuCH**. Prime Subsets within an existing hierarchical set structure are identified with partitioning into Goldbach Triples of Prime Subsets, a 'top down' process which is plainly finite, the partitioning terminating when the last candidate Prime Subset has become too small to be partitioned any further. The smallest non-partitionable Prime Subset has just 7 members ($2 + 2 + 3$) and is identifiable with the second Level $\mathbf{S}_2$ in **PuCH**.

There is of course the dual notion, directly related to the notion of 'process' and 'growth' — *Goldbach Spreading* — a 'bottom up' process within **PuCH**.

Here is a sketch of the algorithmic process involved :

Example 4.

G1 Start with that same non-partitionable Prime Subset $\mathbf{S}_2 \approx [2, 2, 3] : 7$, that smallest Goldbach Triple in **PuCH**. Label it **GT1**.

G2 Randomly select any one of the 3 Prime Subset components $\mathbf{F}_i$ of **GT1**, and select a new randomly chosen Goldbach Triple one of whose prime components is already the same as that chosen $\mathbf{F}_i$. Label it **GT2**. Form the 'nested tree' **GT1** $\triangle$ **GT2**.

E.g. Choose $\mathbf{F}_i = 3$ and choose $\mathbf{GT2} \approx [3, 7, 13] : 23$

we see that

$$\mathbf{GT1} \triangle \mathbf{GT2} \approx \begin{array}{l} [\,2,\ 2,\ 3\,] \\ \qquad\quad \downarrow \\ \qquad\ [\,3,\ 7,\ 13\,] \end{array}$$

G3 Repeat **G3** with **GT2** in place of **GT1**, and a new **GT3** in place of **GT2**. Form the 'nested tree' **GT1** $\triangle$ **GT2** $\triangle$ **GT3**.

E.g. Choose $\mathbf{F}_i = 7$ in $\mathbf{GT2} \approx [3, 7, 13]$ and $\mathbf{GT3} \approx [7, 23, 29] : 59$

we see that

$$\mathbf{GT1} \triangle \mathbf{GT2} \triangle \mathbf{GT32} \approx \begin{array}{l} [\,2,\ 2,\ 3\,] \\ \qquad\quad \downarrow \\ \qquad\ [\,3,\ 7,\ 13\,] \\ \qquad\qquad \downarrow \\ \qquad\qquad [\,7,\ 23,\ 29\,] \end{array}$$

Gn Continue in the same way, as in **G2** and **G3** with appropriate changes. Stop when no larger Goldbach Triple can be found in the last Level $\mathbf{S}_4$.

For any reader wishing to experiment with Goldbach Spreading in **PuCH** here is a short list of the first few prime numbers :

2	3	5	7	11	13	17	19	23	29
31	37	41	43	47	53	59	61	67	71
73	79	83	89	97	101	103	107	109	113
127	131	137	139	149	151	157	163	167	173
179	181	191	193	197	199	211	223	227	229
233	239	241	251	257	263	269	271	277	281
283	293	307	311	313	317	331	337	347	349
353	359	367	373	379	383	389	397	401	409

[Much larger lists can be found on the web : Google on "Prime Number Lists".]

Dendritic crystal growth is very common and illustrated by snowflake formation. In terms of 'structural configuration' the crystallization of ice crystals in 'Frost Flowers' on a window pane (see frontispiece) is an apt analogy of this Goldbach-Prime-Triple Spreading.

6.4 Temporary conclusion : Work–in–Progress

As I remarked in section §5.2 of the first part of this paper :

"Every one of the myriads of Prime Subsets in **PuCH** is identified with a group (unique up to isomorphism), namely the cyclic group $\mathbf{C}_p$ where p is the count size of the Prime Subset — and hence can be visualized as a 'Necklace' (or Hoop)".

Thus each Goldbach Triple in the above examples of a Goldbach Spreading or Partitioning may be visualized as a triple necklace, the union of three Prime Subset 'Necklaces'. [But not, NB, as a Borromean Ring or other kind of non-trivial knot.]

The resulting Goldbach Spreading in **PuCH** can thus be visualized as a growing chain of larger or smaller "Triple Prime Necklaces".

The form of such a 'bottom-up' spreading dendritic-like Necklace structure has visual similarities with the image of the 'Frost Flowers Crystallizing on a Window' in the Frontispiece of this paper.

By contrast the 'top-down' dendritic structure of Goldbach Partitioning has visual similarities with the image there of the 'Fractal Structure'.

I propose to give a name to this new system consisting of **PuCH** together with this iterative partitioning *via* 'Goldbach Partitioning' and 'Goldbach

Spreading' into Triple Prime subsets :

G O L D B A C H

abbreviated to **GoCH** (pronounced like "loch").

Of course there is no reason why the dendritic structure should always be in a monotonic direction : 'bottom-up' (spreading) $-v-$ 'top-down' (partitioning). Much more complex (perhaps chaotic ?) structures could evolve by permitting a change of direction at any node in one growth-direction to the other direction, the change-over node and change-over direction both being randomly chosen.

This raises the prospect of introducing 'reflexive' or 'self-referential' growth behaviour in which a 'Program Universe'-type clock-stepped process-control makes 'decisions' for us, based on the 'current state' of the evolving **GoCH** system.

There seems to be a huge range of possible, immensely elaborate dendritic down–growing, up–growing, Prime Subset Trees which could arise through the dual processes of Goldbach Partitioning and Goldbach Spreading in **GoCH**.

It is difficult to conceive of any systematic treatment that would do justice to this potentially hyper-complex Hierarchy and its emphasis on Prime Subsets.

All the Levels of **GoCH** and **PuCH** and Parker-Rhodes' original **CH** are Prime Sets. All their Levels have myriads of Prime Subsets (many of them Discriminately Closed Subsets in the 'old **CH**-speak').

What is about Prime Sets that appears suddenly to be so important for our study of such Hierarchies ?

TO BE CONTINUED . . . .

ACKNOWLEDGEMENT : It was Apostolos Doxiadis' rewarding novel

Uncle Petros and Goldbach's Conjecture (Faber and Faber, 2001)

that re-awakened my interest in triple prime number partitions. I recommend it as collateral reading matter.

MEANING OF NUMBERS

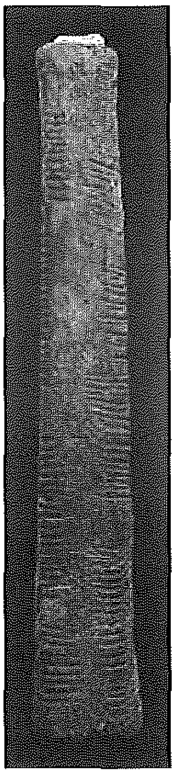
Anthony Blake

God made the integers, all else is the work of man. Leopold Kronecker

This is a brief survey of the meaning of number from ancient times leading to a summary of J. G. Bennett's scheme of 'multi-term systems'. It is a clarification and expansion of the talk given in 2010.

Prehistory

The *meaning* of 'number' is perhaps irreducibly bound up with our techniques of representation of number. It is hard therefore to imagine what number meant in ancient times, or even in recent historical times, before the prevalence of our taken for granted nomenclature with nine digits, zero, and place values. Certainly, people counted objects and events but probably by one to one mapping of them into marks. Whether ancient people then did calculations with these marks – turning marks into other marks – we do not know for certain.



The now famous Ishango Bone from Africa, dated 20,000 BCE, might show half of a lunar calendar or some of the earliest exercises in arithmetic and recording primes. If the former, the marks are only a reflection of external events and limited by them. If the latter, that would provide evidence of the way we think about a phenomenon becoming something we think about. As soon as we think about the way we think about something we are on the way to mathematics as a conscious discipline. Of course, the evolution of mathematical *language* takes time.

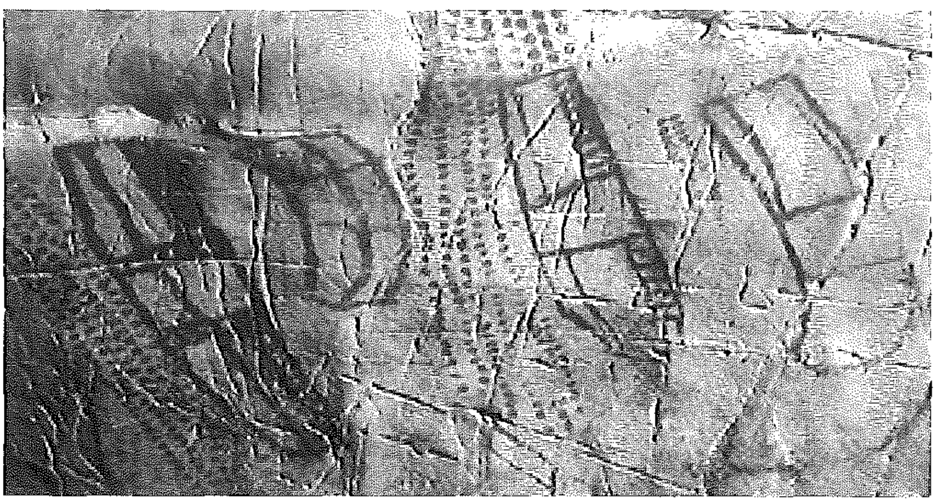
Let us bear in mind a simplistic *triad* – a term we will explain later – as follows:

IDEAS

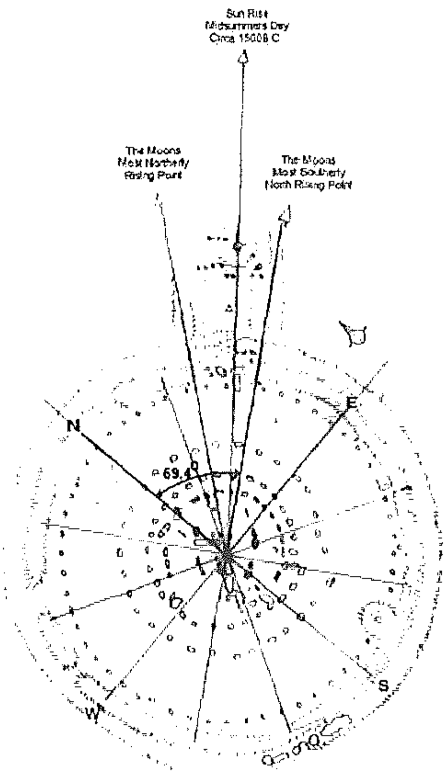
LANGUAGE

OBJECTS

As far as the historical record goes, the further back in time we go there are only objects to look at; but we suppose that some objects, such as the Ishango marks, might signify a rudimentary language. Alexander Marshack, who pioneered research into Palaeolithic notation, emphasised what he called the 'time-factored' structure of this notation. Ideas - on the 'other side' of language to objects (or marks) - are of course not visible to us but they are implicated in any kind of design or intentional production. The remarkable art of the Palaeolithic includes not only vivid and brilliant depictions of animals but also an elusive symbology of abstract forms still unread by us.



Reading is important. It is hard to imagine an oral mathematics extending beyond counting. Not only that, oral tradition is conservative (stories can be transmitted reasonably accurately for thousands of years) while writing and reading have a potential for innovation. This is simply because *we can read more than is written*. Thus we can look at an equation and ask: What does this mean? What does it tell us? This is because the *language* of mathematics is something active; it is operational.



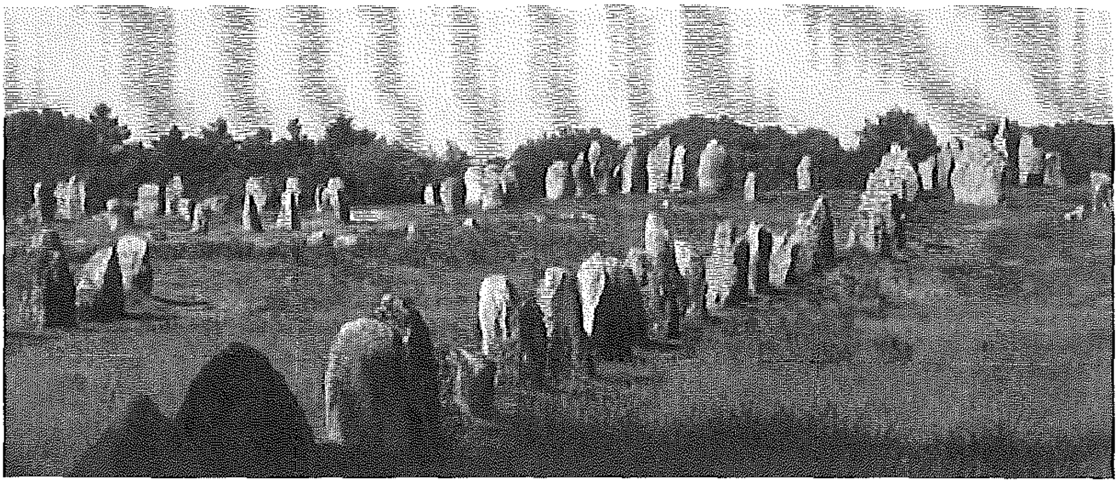
Old Plan Of Stonehenge

Twelve thousand or more years ago, at the end of the Ice Age, there is evidence that what has been called 'Megalithic' culture (dating six thousand years later) was already in operation (at Gobekli Teppe, Turkey). This remarkable site is situated at the very centre of what became known as the Fertile Crescent where agriculture and the Neolithic were born.

The Megalithic culture of Europe, which predates building of the pyramids in Egypt, has been the subject of controversy regarding its possible mathematical and astronomical knowledge. The still prevailing establishment attitude appears to belong to the world-view which regards rational science as a recent phenomenon and says that people earlier than us must have been less capable. Prehistory by definition lacks the documentation to support any idea that people could do astronomical calculations (beyond simple observations) so it is easy to deny that they did. Stonehenge for example is still largely classified as a ritual centre – religious rather than scientific – despite efforts by Fred Hoyle and others to show how it could have been constructed as a kind of 'computer'. Scientists and engineers are far more likely to appreciate ancient knowledge than people from humanistic backgrounds who are restricted to verbal language and lack mathematical literacy.

The 'form' of Megalithic monuments can be taken as ritualistic, an echo of the dream world one might say, or as intentional means of furthering calculation and observation, basic pursuits of the scientist: maybe both operated together without conflict. If we are going to build a calculating machine at great expense of labour and materials why not make it beautiful and attract tourists (pilgrims)?

Planting great stones in the landscape was a big step forward from making marks on bones. I like to think that this was a projection of simple patterns onto a large scale in such a way that the consequences of their conjunction could be made more visible. I hope it is not too quaint to cite the present day building of vast accelerators as a parallel (hopefully, we will eventually come to some more efficient way of pursuing this research). We go big in order to measure something small. In the case of Megalithic technology precise measurement of the sidereal year can be taken as an example.



Geometry was born in this epoch, literally as the 'measurement of earth' but in reflection of the skies. The key element was the right-angled triangle. We know that the 'Pythagoras theorem' was pragmatically known in ancient Babylon but suspect it was familiar long before that. I'd like to say that geometrical forms carry the sense of *ideas* and contain *relationships*. Circle, line, triangle, and square served to organise measurements. They remain evocative of the numbers 1, 2, 3, and 4 to this day.

The Indo-European words for numbers date back at least 5,000 years, probably at least twice that, and it has even been argued that Proto-Indo-European is continuous with the Palaeolithic. Certainly, literate communities had 'innumerable' words for numbers as best exemplified in the Sanskrit language of the Indian sub-continent, which was capable of embracing poetry and mathematics with equal facility. As is widely known, our written numbers came from this culture through the Arabs. So too, even more importantly, did our number *system* of place values and use of zero. Recent research suggests that language was and is important for generating our sense of numbers.

Music as Food for Thought

The much-abbreviated story (or mythology) of number we are telling brings us to the invention of writing around 5,000 years ago. This is writing in the sense of recording speech. It was transformed by invention of the alphabet around 1000 BCE, an invention that happened only once and has become global in its reach. The alphabet lent itself to equating letters with numbers but historical documents give us glimpses of a mode of thought in which the *gods were numbers*.

Let us not presume to know what the gods meant for ancient people. I believe we project onto them images of a 'more primitive' intellect, assuming a line of progress in time that positions us at its summit and seems to demand that we demean those who came before. The *Neter* of Egyptian thought were hardly capricious agencies but rather principles of understanding. I think we can speculate that the gods of all cultures were archetypes of explanation, even though the 'herd mind' has always turned them into personalised demons as for example in the reifications of 'terror' used by politicians in the USA today. Most early cultures had some form of sacrifice as appeasement of higher forces – even of human sacrifice on a horrific scale as in South America – but that can also be seen as a degeneration of previous insight into number and mathematics. Degeneracy and decline are common features of history. I remember being horrified to learn (from Pierre Duhem I believe) that mathematics in Egypt declined from 2000 to 1000 BC.

Numbers as principles of explanation raise the problem: which one is in charge? Some numbers are 'more equal' than others. The series 1, 2, 3, etc. continues without limit, but we can look for the appearance of limits on the one hand and the properties of the first numbers capable of generating numbers on the other. The Sumerian number system was sexagisimal, based on 60, and still exists in measurements of angles and times. In a more than metaphorical sense 60 = 1 *the authority.* The first five numbers are prime excepting 4 (= 2x2). We assume the Sumerians knew that doubling in music produces the same, the octave. Using an octave based on the ratio 30:60 the Sumerians could articulate the realm of the gods. Indeed, it seems there were:

60 Anu = 1 – Father of the Gods
50 Enlil = 5/6 – God on the Mountain – mankind's special guardian
40 Ea = 2/3 – God of the Sweet Waters – organises the Earth
30 Sin = 1/2 – the Moon

Only in this scale can we have the gods in the simple ratios 4:5:6, giving us the major and minor third. Anu became a transcendental principle that did not do anything, an idea that persisted well into Greek thought and can be found elsewhere. Enlil became important as associated with 5, the 'human number', and Ea as associated with 3 the 'divine number'.

Lesser principles were depicted in the next octave down, 15:30

20 Shamash – the Sun who 'judges the gods'
15 Ishtar – the Feminine

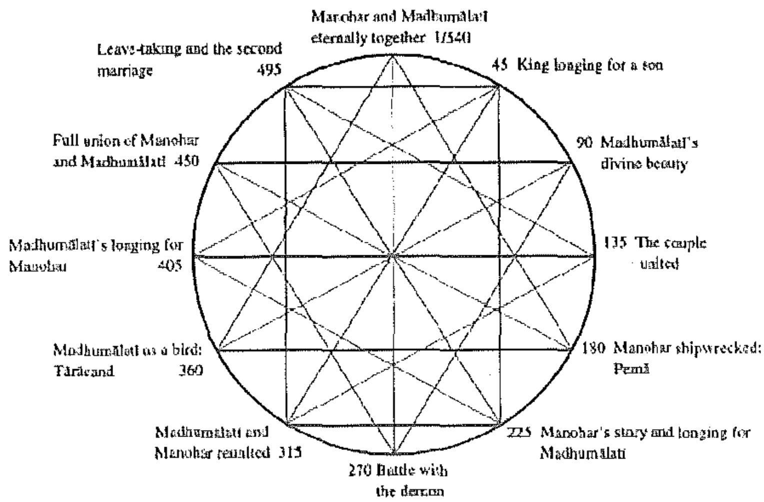
And below this were:

12 Nergal – God of the Underworld
10 Baal

It is intriguing to note that Baal became the top god, a precursor to the Pythagorean base-ten system (also known to the Sumerians) and eventually leading to the prominence of 1 and the rise of monotheistic religion ('There is but one God').

Those suggestions were made by the scholar Ernest McClain who claimed: "Their gods were assigned numbers that encoded the primary ratios of music, with the gods' functions corresponding to their numbers in acoustical theory." We can also see that the musical structure was used *to map elements of experience and imagination.* The very activity of making sense of things in this way provided new illustrative material essentially relating to 'thinking about how we think' or 'representing how we represent'. The numbers 5 and 3 were connected with the cycles of Venus with Mars. Greek mythology depicts the lame blacksmith (i.e. astronomer) Hephaestus, son of Hera and husband to Venus, casting a net over the couple caught *in flagrante* to signify the work of reconciling the two cycles. A net symbolised measurement or system.

It is possible that the geometric-numerical knowledge implemented in Megalithic times was transferred into *texts*. Certainly we should not presume that the earliest writers simply tossed off stories as these came into their heads. The usual assumption is that they recorded oral traditions. Interestingly, such oral traditions must have entailed *structure* because they relied on memory. The thread of mnemonic structure extended right into Renaissance and even modern times, as in the 'theatre of memory'. Stories recited by bards in the nineteenth century proved to be accurate renditions of narratives originating thousands of years ago. Repetition, contrast, and so on must have been the primary tools with the device called *chiasmus* following the form, e.g. ABBA, ABCB'A', or some other variant. *He knowingly led and we followed blindly* is an example of a very simple ABBA. Symmetrical form draws attention to the middle and scholars have said 'the meaning is in the middle'.



With writing the possibilities for chiasmus were greatly expanded and became more sophisticated, extending in fractal form over whole works such as the *Iliad.* Given that the readers of the culture would understand such structure it enabled checks on accuracy. Not only that, it afforded a geometrical view of texts. Above is a recent example (15th century) of a very sophisticated Sufi text from India which highlights the approach. It is also an example of the twelve-term structure that was a common form uniting the calendar, zodiac, astrology, music, and the construction of ritual centres.

McClain's thesis of musical structure introduces the prospect of yet other forms of relation. We must add that musical structures involve musical *scales* where, for example, we can have different numbers of notes in an octave and different ratios between them (Think of our standard major diatonic scale and Debussy's whole tone scale as examples). But, again, the twelve term structure stands out because all its intervals (from one note to the next) are the same. We leap ahead in time once more to Plato and the recent work of J. Kennedy on the composition of the *Dialogues.*

Using computers, Kennedy went back to the original form of the texts, which were in standard lines that could be counted, and discovered multiples of twelve. Dividing the texts accordingly, he scrutinised the content of each twelfth and concluded that this precisely echoed location in the text in terms of degree of harmony.

In a twelve-term scale 2, 3, 4, 6, 8, and 9 are harmonious while 1, 5, 7, 10, and 11 are not; and some more strongly than others. In the *Symposium* Socrates' discourse on the nature of Eros comes in the middle (6) and Diotima's inspiring and elevated discourse speaks of the Form of Beauty at 8 and the Form of the One at 9. At the earlier point 7 she starts with the myth of sex and seduction among the gods. According to Kennedy, the parallelism extends to quarter divisions or eighth tones with surprising precision. Also Plato explicitly speaks of music and the science of harmony in his text and even indicates the quarter division.

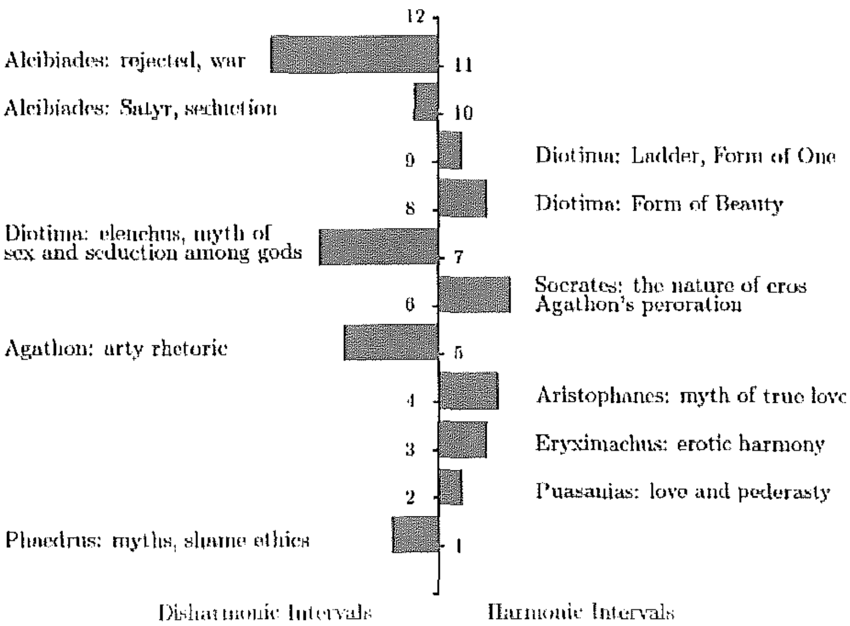
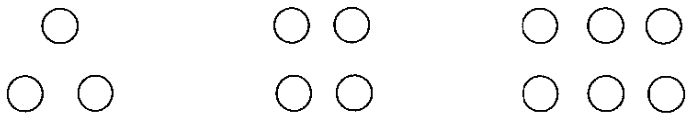


Figure 7: Regions near Harmonic Notes have Positive Themes, and Regions near Disharmonic Notes have Negative Themes

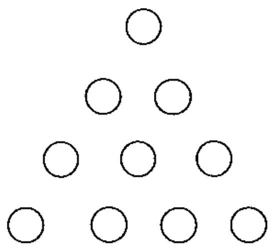
PYTHAGORAS

Pythagoras is reputed to have studied in Egypt and Babylon and assimilated much of the previous epoch's store of wisdom. He was the epitome of the union of mysticism and science which has echoed down the ages. It was with him that the sense of numbers in themselves having meaning reached an apogee. Not only was everything to be explained by number but it actually *consisted* of number.

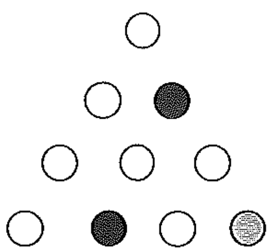
We have to take into account the prevalence of using stones or pebbles to represent units. A number could be laid out as a pattern of stones. This gave triangular, square, oblong, and other-shaped numbers.

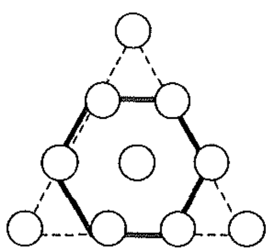


The number ten, represented in the *tetraktys* as the sum of the first four numbers, was supreme.

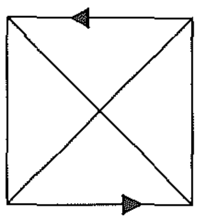


Units were arranged in shapes or had *form.* This preceded the Aristotelian combination of Form and Matter but with an atomistic emphasis. The Aristotelian 'matter' was inchoate whereas the Pythagorean was already articulate. Number then had two meanings: as the counting of units and as a shape or form which conveyed a sense of how the units were *organised.* This is very different from the idea of a set. For example, the *position* of a unit within the form would be significant. Geometry meets number. No records exist of Pythagoras exploring form in this way but it is plausible.





Alongside this playing with shapes we should remember verbal language in which ideas are first encountered. Perhaps verbal and mathematical expressions need to agree or be translatable into one another. In the later times of sacred scriptures this became a measure of authenticity. Numbers and calculations were used as 'proofs'.



One of the best-known Pythagorean formulations states *justice is a number squared.* A little reflection links this to the idea of doing to others as one is done to, or *reciprocity*. There is a relationship that reflects itself. The refinement: do unto others as you *would wish* to be done by is a more subtle form. A diagram such as the one here is suggestive but no more than that. It simply indicates the possibility of dealing with ideas as *forms of thought*.

The Pythagorean symbology of numbers is known through subsequent writers and appears to have corresponding features in many cultures. Anthropologist Levy-Bruhl later asserted: "Each number has its own individual physiognomy, a kind of mystic atmosphere, a 'field of action' peculiar to itself". Pythagoras distinguished odd and even numbers as masculine and feminine, possibly carrying on the tradition of associating odd months of the year with the gods. In terms of early Greek ideas the odd numbers and the gods were linked to *limit* and the even to the *unlimited.* The number 2 as the archetypal feminine took various guises including that of breeding (times 2) and breaking of wholeness or strife. These are the sort of associations numbers had with verbal ideas:

One is *nous* or mind (original)

Two is *doxa* or opinion (divided) **feminine**

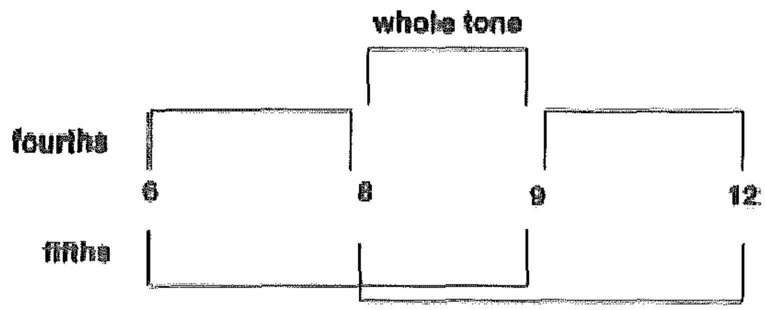
Three is a number of the whole (beginning, middle, and end) **divine**

Four is justice as equal times equal (the 'whole nature of numbers' in the *tetraktys*)

Five is marriage (union of first odd and even numbers) **human**

Six is a *perfect number* because the sum of its factors is also six.

Seven is *kairos,* opportunity, or the 'virgin' number (Athena) because it cannot be produced from any previous number or produce any other number (within the ten). Seven is particularly striking because it is harder to make a picture of it than of six, for example. Hence it has greater *numinosity.*



Pythagoras had great interest in the structure of the musical octaves and the laws of harmony. The first four integers provide a coherent structure for the octave (as shown above).

The first four numbers, as Pythagoras seems to have discerned, engage us in archetypal (mental) operations. Understanding these operations may well be the way to understand understanding.

Pythagoras brought about a transformation in the ancient knowledge of his time, stretching back millennia, rendering it into a form that was able to survive the great changes taking place in human life. This ancient material persists up to the present day even though it has long been superseded in the development of mathematics. In modern mathematics, *number has no content*, as illustrated by the derivation of numbers from the empty set. In ancient times number was a way of linking the human to the cosmic, of finding the proportion as in music.

ABJAB

The mapping of numbers onto letters – in Greek, Jewish, and Islamic cultures – produces a kind of rewrite system. The alphabetical sequence we know as A, B, C, D, etc has strangely endured since its origins as in *alpha, beta, gamma, delta,* etc.

Alpha = 1

Beta = 2

Gamma = 3

Delta = 4 etc.

Doing sums with words made out of letters, each of which had a numerical value, is called *Gematria*, in Islam *abjab.* A word is defined by the sum of its letters. The suggestion is that words that have the same number express the same idea, even though their origins might be quite different. *Abjab* is the alphabet of Arabic. Right from the start of writing down the Koran, scholars looked for hidden meanings in the text that might serve to rationalise its apparently arbitrary use of letters and terms, mainly to reinforce the belief that the Koran was of 'divine' origin.

Allah has the number 66. He has 99 Names. Adding them we get 165 which is the number of the declaration of faith *La ilaha illa Allah* and also of the Divine Name *as-Samad* the Eternal Sustainer.

The *Science of Letters,* the *'ilm al-huruf,* concerned both the numerical value of the letters of the alphabet and their *shape.* The first letter *alef* is said to give rise to all other letters and all the letters together map the universe.

The Gematria of the Jewish Kabala is more widely known in the West than *abjab* but has much the same nature and purpose. The word might stem from the third Greek letter *gamma* or the word *geometria*. It is fairly certain the system came from the Greeks, and not entirely fanciful to see some parallels with particle physics. There are twenty or so fundamental 'particles' which form into larger particles (words or phrases); whenever words have the same number they can be 'transposed' into each other. So we have some kind of conservation laws and rules of transformation.

The critical idea is that *things of the same number have the same meaning.*

THEOLOGY

"These are plenty complicated enough for now — why don't we save the Trinity stuff for later?"

If in ancient times numbers were gods, then what became the number of God when the monotheistic religions came to prominence? The obvious answer that God = 1 is incomplete because we have to understand what this oneness amounts to. For different religions God had different values:

Zoroastrianism	2	
Buddhism	0	
Judaism	1	
Hinduism	many?	
Christianity	3	and so on

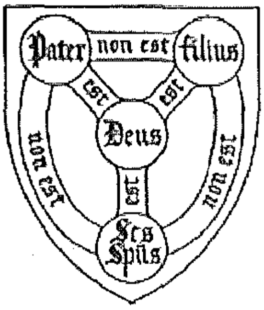
For example, Muslims would hate Christianity because of its blasphemy of 3 gods.

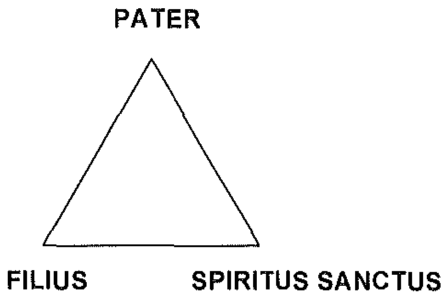
The word God indicates whatever we mean by the active principle of the universe, including ourselves, as source, origin, power, intelligence, etc. This has been treated anthropomorphically, naturalistically, abstractly, psychologically, and so on. Every worldview has a version even when 'God' is not part of its vocabulary. Dealing with the concept of the 'One God' was probably instrumental in developing abstract reasoning.

We owe most of our Western civilization to the dual influences of the Greeks and Jews (who both managed to assimilate what they could of other cultures in the Mediterranean and the Middle East) and we can think of mathematical abstract reasoning on the one hand and inspired narratives and writings on the other as their two major contributions. It is significant that just as the Church became a repressive and destructive force in relation to high ancient culture it carried within itself seeds of a very active form of mentation, hidden as it were in abstruse theological doctrine.

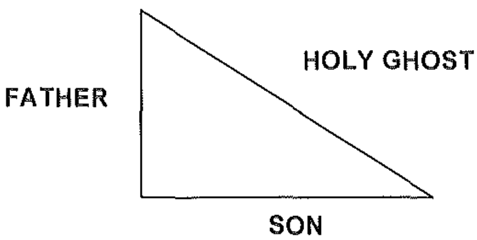
Violent discussions of doctrine began with the question of the duality of Christ as both God and man. How was this possible? Dealing with this paradox raged on, producing various 'heresies' such as those in which either the human or divine nature of Christ was denied. It would be wrong to dismiss these contentions as parochial or misguided; this type of paradox is universal and crucial for creative insight. The formula *One Person of Two Natures* spoke of an *hypostatic* or *mystical union*. This kind of union evokes a special sense of number, one which does not divide but *is* in a stronger sense than mere existence. It exhibits two levels of language.

The formulation of the *Trinity* as *One Nature in Three Persons* is a striking reflection or complement of the hypostatic union. The Trinity is not an 'impossible union' but a fundamental dynamism – the workings of God within the nature of God. It makes *relationship* primary. This was represented by various forms, usually triangular. The medieval image below shows the perspective of Western Catholicism in that it supports the doctrine that the Holy Ghost proceeds from the Father *and* the Son, in contrast to the view of the Eastern Church that it came from the Father only as sketched in the other diagram.





We can revert to the archetypal right-angled triangle to give a reasonable picture of the doctrine of the Western Church highlighting the Holy Ghost as Reconciliation. The three were ascribed such operations as: Creation, Redemption, and Perfection (a clockwise motion in the triangle can be imagined):



The apparent sequence or progression from 2 to 3 in Christianity has led some recent thinkers such as Carl Jung to anticipate the emergence of the 4, in particular to include the significance of the Mother. The divinity of Mary was only acknowledged by the Church in 1925, after a thousand years of controversy, so there is still some way to go.

Whatever the number of God may be, there would be a duality of terms revolving around the implicit equation:

$$N = 1$$

The equation implies a distinction between compresence and coalescence, two distinct types of togetherness (which we will speak of later). It has contrasting interpretations:

(a) Can be found to be true whatever the value of N

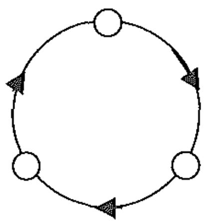
(b) Is true for only one value of N (dogma)

We need to find an N clever enough to generate the whole universe. This was proposed in Taoism in such terms as: *Out of the one comes forth the two, out of the two comes forth the three, and out of the three comes forth all the ten thousand.* There is a kind of internal progression leading to an N which generates externally. Thus Meister Eckhart saw the Trinity come out of the Godhead. Behind the duality of Ahura Mazda and Ahriman in Zoroastrianism is Zurvan, infinite time.

In crude terms then religion splits the one to enable a view of reality which makes sense, or something we can reason with. Theologies wrestle with forms of reflexivity to reach a form that can be mapped onto experience. These forms might be enigmatic as in Spencer Brown's "Only nothing could be sensitive enough to nothing to produce something", which we include here under theology just to mention in passing the hitherto ignored role of zero.

In terms of ordinary senses of meaning, the articulation of God follows the quest for what is active, original, self-caused, etc. That raises the question of how anything can happen as in the well-known puzzle of the omnipotence of God in an imperfect world. The active principle must be able to oppose itself or *limit itself.*

. The image here might help us to grasp how this could be seen. The single active principle (God) is shown as a *rotation* through three points. Three steps equal one cycle. 3 = 1.

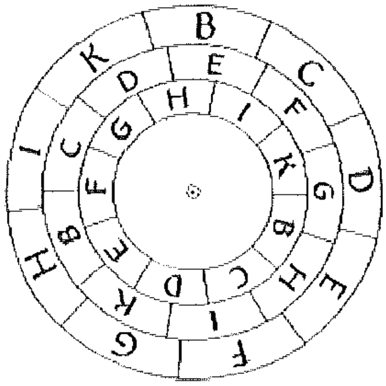


Spinning, rotating, iterating, and so on are ways of conceptualising our N = 1. In the theology of the Eastern Church God is one in his *Essence* and many in his *Energies.* What we call the active principle is the ultimate and irreducible *reason why.* When Schopenhauer as a secular philosopher wrote his *World as Will and Idea,* will was presented as blind and had to be given form and meaning by ideas. But ancient theologians, and perhaps some modern thinkers too, persist in seeking reason within will itself. In this way of looking at the history of ideas, what is called *faith* has the character of 1 whereas *reason* has the character of N. Faith is the intensive aspect of will and reason is its extensive aspect. This is to see *Will as the prime substance of number* in this stream of human enquiry. All notions that God is a *being* who manifests in action are confused. The N points of a N = 1 are recurrences in the act of will. Historically, Will came to prominence in recent times whereas in Greece Being was predominant. If Being is associated with the number 1, then Will might be associated with the number 0. In Hinduism it is associated with ignorance or not-knowing.

ARS COMBINATORIA

Ars combinatoria, the 'art of combinations', was the name given by Leibniz to the technique, invented by Raymond Lull in 1275, of generating new ideas by making combinations of existing ones. That led Leibniz to speculations about a universal calculus of ideas.

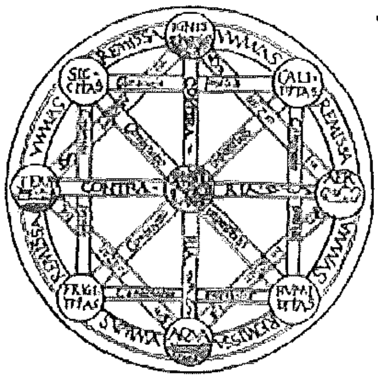
Such speculation may have been inspired by contact with Islamic scholars. As a Semitic language, Arabic has words largely based on triliterals, combinations of three consonants which act as primary roots of meaning with variations produced by different vowels, etc. For example, the root KTB appears in all words to do with *writing,* such as *kitab* for 'book'. Other examples are more subtle. TRQ includes TaRQ, the sound of a musical instrument, and TaRiQa, the mystic path in Sufism. The number of possible triliterals must run to several hundred thousand and thus holds out the prospect of discovering new thoughts. Lull does not appear to have delved into the possibilities of Arabic and Hebrew but adapted the notion of combinations for his own purposes. Circles of divine attributes and virtues could be rotated relative to each other to throw up combinations, often as a basis for making sermons. Lull's 'fourth wheel' shows how he could generate his own kind of triliterals through rotating the circles of nine elements:



The system uses a mechanical operation on given data to generate new information. This leads to the mechanisation of thought (and to the GIGO phenomenon!), supposing that we have information in a form which can be treated this way. It anticipates the question of whether ideas can be rendered into numbers with the concept of number taken broadly.

If ideas can be numbers – have 'extension' - then they can be added together, etc. Any phenomenon of an additive nature can be made by addition of smaller units; but ideas are not really like that. We have to treat numbers and ideas as of different natures. Yet the urge to find correspondences between them remains as strong as ever.

Arithmetic and grammar govern number and words respectively. We can make numbers from other numbers and we make words (in alphabetical language) from letters. When we treat certain numbers as elemental they can be combined, in a way analogous to language, to create structures. That is an artifice and perhaps no more than a metaphor. Down through the ages many great minds have sought the holy grail of being able to combine ideas so as to make new ideas. None more so than Gottleib Leibniz in his search for a *characteristica universalis* or philosophical algebra. Here is one of his reasoning diagrams.

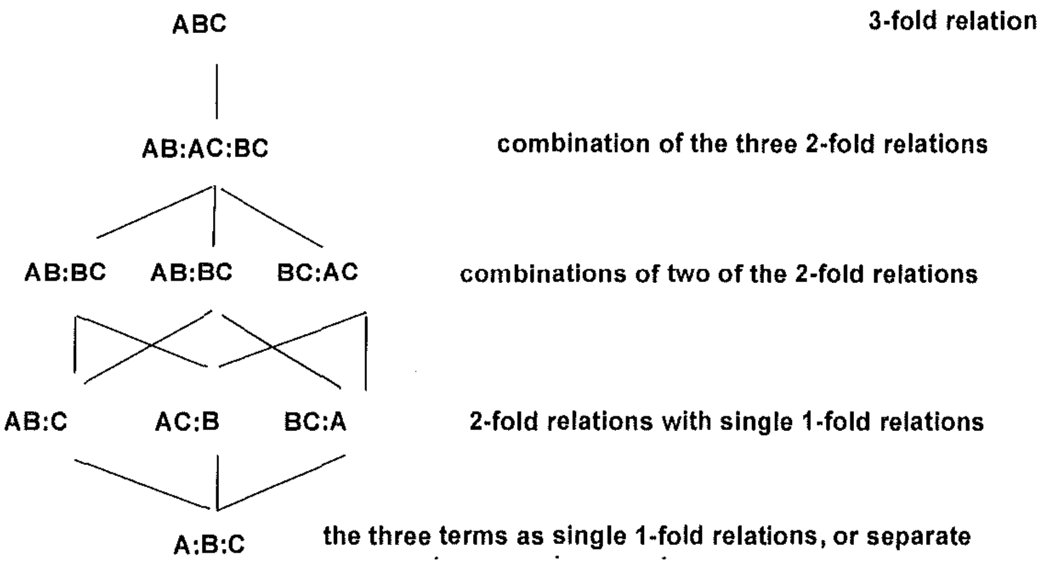


He described the *characteristic universalis* as:

"... a kind of *general algebra* in which all truths of reason would be reduced to a kind of calculus. At the same time, this would be a kind of universal language or writing, though infinitely different from all such languages which have thus far been proposed; for the characters and the words themselves would direct the mind, and the errors — excepting those of fact — would only be calculation mistakes. It would be very difficult to form or invent this language or characteristic, but very easy to learn it without any dictionaries" (letter to Nicolas Remond, 10 January 1714)

This Holy Grail has proved elusive but nevertheless has provided valuable stimulus. Also, whether we regard this as discovery or invention, it is linked to the *creative power of mathematics.* Number is studied for having structure and as revealing structure it seems to underlie the construction of physical reality. We encounter this in technology because we want all the parts of a machine or system to work together and act as one whole. If a device or construction works then all its parts conspire to one end. This can be perfected to various degrees: all machines involve waste or production of entropy – the complement of serving a purpose – and we can find ways of minimising this. The synergy of working together to one end, or acting as a whole, was called *coalescence* by John Bennett - to distinguish this from the state of *compresence* in which the constituent elements have not 'fused' but still act separately. Technology has long been sidelined in the history of ideas but offers a basis for understanding in the realm of practical action where we experience a progressive development of finding out what works. The realisation that every machine we can make produces waste reveals the intrinsic limitations of any 'perfect' scheme.

Distinguishing coalescence from compresence in a scheme of combinations gives us a hierarchy of degrees of 'inner togetherness', as in this example involving threefoldness.



Thus we can have three 'terms' or elements but they only partially attain their ideal coalescence.

The combinatorial hierarchy of interest to ANPA proceeds by taking simple elements with a dyadic operator such that two distinguished elements generate a third, leading to a concatenation with four levels associated with the numbers 3, 7, 127, and $2^{127} - 1$. Remarkably, the first two numbers stand out in the traditions of ancient number reaching into modern times through occult or esoteric channels. The four levels also remind us of the Pythagorean *tetraktys*, suggesting some archetypal origin.

MODERN TIMES

The work of the American philosopher C. S. Peirce is increasingly being appreciated as people 'catch up' with his insights. He appears to have anticipated much of Spencer Brown, for example, in his entitative diagrams. He was also aware of the importance of the *qualitative* significance of numbers as *categories of understanding.*

> "Perhaps I might begin by noticing how different numbers have found their champions. Two was extolled by Peter Ramus, Four by Pythagoras, Five by Thomas Browne, and so on. For my part, I am a determined foe of no innocent number; I respect and esteem them all in their several ways; but I am forced to confess to a leaning to the number three in philosophy. In fact, I make so much use of three-fold divisions in my speculations that it seems best to commence by making a slight preliminary study of the conceptions upon which all the rest must rest. I mean no more than the ideas of First, Second, Third - ideas so broad that they may be looked upon rather as moods or tones of thought, than as definite notions, but which have very great significance for all that. Viewed as numerals, to be applied to what objects we like, they are indeed thin skeletons of thought, if not mere words . . ..
>
> First and Second, Agent and Patient, Yes and No, are categories which enable us to roughly describe the facts of experience, and they satisfy the mind for a very long time. But at last they are found inadequate, and the Third is the conception which is then called for. The Third is that

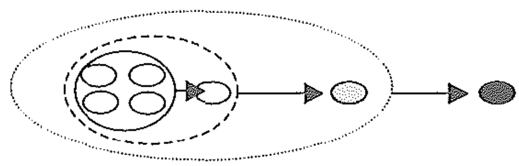
> which bridges over the chasm between the absolute first and last, and brings them into relationship. . ..
>
> . . . I believe that if my suggestions are followed out, the reader will grant that One, Two, Three, are more than mere count words like "eeny, meeny, mony, mi" but carry vast, though vague ideas. . .."
>
> 'A Guess at the Riddle', *The Essential Peirce*, Vol. I.

In this remarkable passage Peirce not only clearly and straightforwardly states that numbers carry 'vast though vague' ideas but admits his personal preference for three without claiming for it universal truth. He had reflected on his own practice and discovered the pattern of threeness governed what he did. In his semiotics, for instance, signs consist of a triad: sign, object, and interpretant. The third element, the interpretant, is an understanding of the sign-object relation. His work is permeated by a sense of relatedness and he was conscious that this required at least three terms. Peirce then influenced Bertrand Russell who in his *Principles of Mathematics* extended the principle of numbers as categories of understanding to link *order* with four terms.

The question arises whether the series of numbers interpreted as categories can be extended beyond three or four and, if so, how far could this be taken? Russell's co-writer of the *Principia,* Alfred North Whitehead, approached a similar question in his *Process and Reality* by his concepts of 'conjunction' and 'disjunction'.

""Together" presupposes the notions "creativity", "many", "one", "identity" and "diversity". The ultimate metaphysical principle is the advance from disjunction to conjunction, creating a novel entity other than the entities given in disjunction. The novel entity is at once the togetherness of the "many" which it finds, and is also one among the disjunctive "many" which it leaves ... The many become one, and are increased by one. In their natures, entities are disjunctively "many" in process of passage into conjunctive unity. This category of the Ultimate replaces Aristotle's category of "primary substance"."

This open-ended principle gives no special meaning to the number of entities brought into conjunction. It does however suggest that the 'last' term in any given creative step will embody the synergy of the other terms or somehow 'embrace' them - as for example Peirce's Third 'bridges over the chasm between the absolute first and last' and unites them. Whitehead's notion might be pictured as here:



It has the character of an organic and evolutionary worldview. For example, biological evolution seems to follow a sequence in which many elements of an earlier stage are united in a new element. The emergence of a novelty leads to its subsequent integration. Ideas of progressive action of this kind can be found in ancient lore, as in the shamanic image of the shaman firing an arrow on which he stands to fire another arrow, and so on.

Peirce's example has a form such as: U(a,b) = (a,b,u) where U is Unification, and u is the expression of this operation as a synergic term in a higher system.

The thought has the form ((((((((( )))))))))

General ideas like these show nothing of any special character or meaning in number-ideas like threeness. However Peirce, like many others before and since, showed a preference that implied only a few number-ideas were needed, the rest being derivative.

One of the most remarkable mathematicians to appear in the early twentieth century was Ramanujan, renowned for number theory. He claimed these insights came from his family's deity, Nagiri-Lakshmi, consort of Vishnu. Just recently his work on partition theory has been revived and a formula found for calculating p(n), partition number for any n. P(n) is the number of ways we can produce a number n from constituent numbers, e.g.:

4 = 3 + 1 = 2 + 2 = 2 + 1 + 1 = 1 + 1 + 1 + 1 p(4) = 5

Partition numbers divisible by 5, 7, and 11 do not show the 'fractal' properties possessed by partition numbers divisible by other numbers. Ramanujan said that 5, 7, and 11 have 'simple properties' before the explanation was found. What is significant for us is the notion of a few *special numbers* having the most simple properties. To paraphrase George Orwell, all integers are simple but some are more simple than others. This simplicity was a focus of ancient mystical and religious concerns. Number theory may give us a basis for distinguishing integers but this has a long pedigree.

Like Peirce, most of us would assume that the smallest integers, particular 1, 2, 3, and 4, would be the most simple and capable of generating the interesting features of all the remaining integers. However there is a qualitative argument that suggests otherwise, deriving from the vague general notion of 'being interesting', an indication that what is found interesting has something that no other interesting thing has. Imagine we go through the integers and come to one that is not interesting. It then has the interest of being the first integer after the last interesting one, and is thus itself interesting.

Throughout the ages great interest has been shown in the *prime numbers,* assuming them to be of similar character. Pythagoras distinguished *odd* and *even* numbers as male and female. We have mentioned his idea of a *perfect* number such as 6 (also a product of the first two even and odd). In addition their 'shape' – square, triangular, and so on – distinguished integers.

LIMITS TO NUMBER-IDEAS

The prospect of an unlimited series of numbers, each uniquely interesting, is daunting. It is also not something we would easily tolerate if we were to do any science or mathematics. Like the Pythagorean Greeks we are on the side of limit. There is a restraining factor of diminishing returns, necessitating a return to basic psychology. We distinguish things by difference but what sort of difference does this have to be? As Gregory Bateson said (in *Towards an Ecology of Mind*): information is difference that makes a difference. In the series of integers 1, 2, 3, 4, 5, 6, ... elementary maths gives us at least two measures of difference.

One is the arithmetical difference (N + 1) – N = 1 which is the same in every case.

Another is (N + 1)/N or the interval (as in a musical scale) which decreases with N.

We generate the series: 2/1, 3/2, 4/3, 5/4, 6/5 ....12/11 .....By the time we get to 12/11 it is harder to register difference. We don't have the *power* to do so, the means of amplification. We would need to zoom in at a greater magnitude. The greater the value of N, the more energy is needed to do so. That is why I believe that large N cannot enter human situations without the generation of high energies that are the psycho-social equivalents of those produced in physics by particle accelerators.

The situation is analogous with the representation of number in different number-bases. In a number base of 5 we count 10 as five, 11 as six, etc. Any number higher than four can only be represented using 0, 1, 2, 3, 4. Eleven is 21 in base 5. Let us call this use of place notation 'extrinsic' in contrast to the 'intrinsic' nature of the integers one to four.

Those who restrict the number of intrinsic factors – or primordial, simple, generative, etc. integers - to just a few act out a metaphysical faith that most scientists would seek to emulate. Contra Hamlet's "There are more things in heaven and earth than are dreamt of in your philosophy", they seek the virtue of the least number of numbers: those that shall rule over all. Its proponents deny anything more is needed. Some say that two is enough, others like Peirce that three does it all, and yet others such as Arthur Young swear by four. Traditional cultures frequently embrace 7 and 12 but these are regarded as primitive and unscientific. We note they are the numbers of astrology.

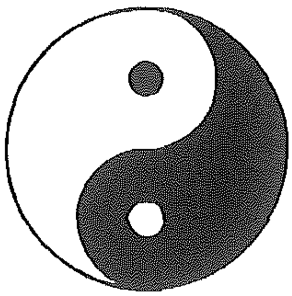
The metaphysical significance of number-ideas should be noted. A prime example can be found in Wolfgang Pauli's thinking on the history of science. He claimed that physics *made a wrong turn* in the 17^{th} century by suppressing fourness, exemplified in alchemy, in favour of threeness, exemplified in mathematical physics, a view that linked him with the work of Carl Jung. It also links him with the 'literary' stream of ideas stemming from Giambattista Vico and Francis Bacon down to James Joyce and taken up by Marshall McLuhan and his son. This stream recognises the significance of human artefacts, a 'second-order kind of reality' one might say (put in that way to invoke an intrinsic sense of the tetrad.)

We dimly glimpse some restraining or limiting factors in the adoption of number-ideas. One is due to simply fatigue or limitations of energy. Another is analogous to the limitations we have in discriminating colours. A third is reluctance to entertain number-ideas which cannot be codified or mastered in some mathematical scheme; they will be relegated to superstition. [From *superstitio* 'standing over' (in amazement) or a 'hold over' from previous beliefs]. Finally, there well might be deep-rooted and maybe unconscious psychic elements at work.

VARIETIES OF THE BEAST

Let us look at some varieties of expression of number-ideas. When expressions arising at different times, and in different disciplines or areas of experience, converge we might suppose some 'intrinsic' agreement. After all, millions of people in cultures all over the world spanning thousands of years have been playing with numbers and one might expect some major attractors of meaning to have emerged.

We can begin with Twoness, with its obvious linkage to sexuality. In cultural terms this is widespread in the Taoist *Yin* and *Yang.* The symbol below has long been treasured as revealing how opposites are also contained in one another.



The emblem was used by Niels Bohr who adopted a dyadic *complementary* approach to everything, including philosophical and social problems. In other thinkers, by sleight of hand, the form of the dyad as 0/1 gives the four states: 00, 01, 10, 11 which have been used to create a fourfold metaphysics.

Keeping to fairly modern times, we compare the threeness of Peirce with those of psychoanalyst Jacques Lacan, philosopher of science Karl Popper, and pioneer of dialogue Patrick de Mare. Each of these requires its own explanatory essay so the table can only offer the merest glimpse.

LACAN	*POPPER*	*PEIRCE*	*DE MARE*
The Real	World 1 – things	Oneness–thing itself	Speech
The Imaginary	World 2 – experiences	Twoness – otherness	Mind
The Symbolic	World 3 – meanings	Threeness – logic	Dialogue

When we add the Christian *Father – Son – Holy Ghost*, comparisons become even richer. We cannot say that the different expressions are the 'same' or that they are 'different'. They exhibit what we shall look at later as characterising the *monad* or one-term system: *diversity in unity and unity in diversity.*

Once we start looking for threes or *triads* we find they are ubiquitous. In Newtonian physics there are the three laws of motion, just as there are the three of thermodynamics (first, second, zeroth). Celtic culture is riddled with triads. The Samkhya system of India has three *gunas* (literally 'ropes') making up the workings of Nature (*prakriti*).

NEWTON	Inertia	Force	Reciprocity
SAMKHYA	Tamas	Rajas	Sattva
TRINITY	Father	Son	Holy Ghost

Even this brief survey shows that the comparisons tend to be *interesting.* We want to suggest at this stage only that they provide a method for generating metaphors.

We have mentioned Jung a few times as the arch-proponent of quaternity or fourness. His four function model is well known:

INTUITION

FEELING **THINKING**

SENSATION

But all his work, especially that concerning alchemy, is permeated with tetradic forms.

The quaternary is archetypal for Native North American culture, starting with the teaching of the Four Directions: North, South, East, and West. When we look at this culture's models of *mind* we find intriguing links:

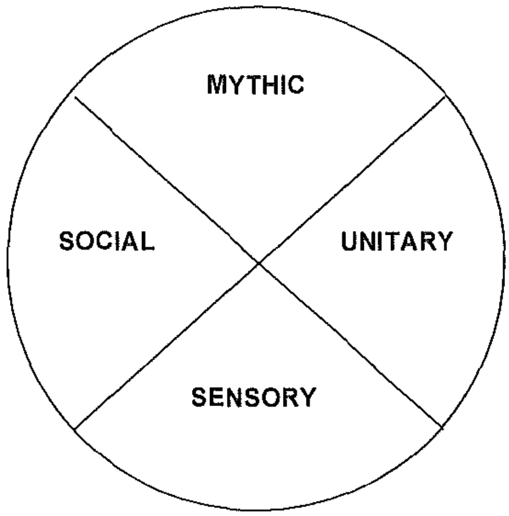
NORTH
Wisdom

WEST
Looks Afar

EAST
Looks Within

SOUTH
Innocence

Native Americans have long thought about fourness! Modern Western thinkers have made similar discoveries, especially when considering perspectives or viewpoints obviously allied to metaphors of spatial direction. Here is a representation by management consultant William McWhinney, based on the ideas of sociologist Lawrence LeShan.



But what can we make of the example of fourness given us by David Deutsch? In his book *Fabric of Reality* he goes against the grain of seeking a single unitary theory and advocates four types of theory to be used in conjunction.

QUANTUM MECHANICS	Many Worlds
EPISTEMOLOGY	Popperian
COMPUTATION	Turing and Quantum Computing
EVOLUTION	Uniting Gene and Meme

Arranging them in the following way may enable a better comparison with our other examples:

EVOLUTION

EPISTEMOLOGY COMPUTATION

QUANTUM MECHANICS

Or the fourfold scheme of Heidegger on the 'limitation of being':

OUGHT

↑

BECOMING ← *BEING* → THINKING

↓

APPEARANCE

Let alone tetrads composed by engineer Arthur Young such as this one based on time-derivatives of space (the top term 'control' is Young's interpretation of the third derivative and close in meaning to 'will'):

CONTROL

ACCELERATION POSITION

TRANSLATION

In passing we should note the techniques required of *arranging* the four terms of each example in such a way that we indicate their best *correspondences*. We suppose that the same order of terms applies in each case so that we can map them from one to the other. There are 24 ways of arranging four terms in a quaternary. With 'good' arrangements we can develop new insights by exchanging information between different examples.

We will conclude our scan of some examples of fourness with Scotus Erigena whose ideas influenced Wolfgang Pauli. In a scheme reminiscent of the Greek elements as combinations of hot-cold and wet-dry he proposed:

NONCREATED + CREATIVE

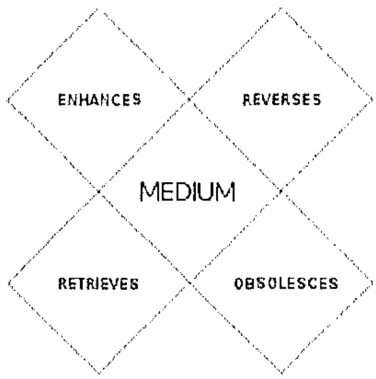
CREATIVE + CREATED

CREATED + NONCREATIVE

NONCREATIVE + NONCREATED

The fourth term is maybe Plato's missing guest that fails to turn up in the triadic *Timeaus.* It is of the form *Neither A nor Not-A* reminiscent of the Shaffer stroke that presaged Spencer Brown's calculus of distinctions but has also surfaced in many other ways in the twentieth century. Theologically it goes beyond the Trinity, a move to be seen throughout the centuries to Jung and beyond.

Finally, just as we had Peirce extolling threeness as the ultimate idea, Eric McLuhan (son of Marshal McLuhan) describes the tetrad as "the single biggest intellectual discovery not only of our time, but of at least the last couple of centuries". Marshall McLuhan's original tetrad is shown here:



Higher number-ideas are not so frequent and, so far as we know, none have the kind of universalist claims accorded the smaller ones. Nevertheless there are varieties of the beast and comparisons are challenging and stimulating. What can one make of Kenneth Burke's pentad with regard to the drama of human situations as compared with the Islamic *akham,* the fivefold categorisation of acts for a Muslim? What could such apparently disparate schemes have in common? Are they like phenomena stemming from identical causes (or *reasons*)? Or are they more like two 'orthogonal' views of the same thing?

BURKE'S MOTIVES

1. **Act:** What happened? What is the action? What is going on? What action; what thoughts?
2. **Scene:** Where is the act happening? What is the background situation?
3. **Agent:** Who is involved in the action? What are their roles?
4. **Agency:** How do the agents act? By what means do they act?
5. **Purpose:** Why do the agents act? What do they want?

ISLAMIC ACTS

1. **Wajib**, obligatory; also known as: *fard*, *rukn*
2. **Mustahabb / Sunnah**, recommended, also known as *fadilah*, *mandub*
3. **Mubah**, neither obligatory nor recommended (neutral)
4. **Makruh**, abominable (abstaining is recommended)
5. **Haraam**, prohibited (abstaining is obligatory)

Evidently there is a wide field of research seeking to correlate diverse versions of number ideas, and to discover what might underlie them or enable us to discriminate between better and worse interpretations.

SYSTEMATICS

Systematics was the rather prosaic name given by John Bennett to his synthesis of number-ideas. He used the term 'system' to emphasise the 'working together' of elements contained in a number (his initial definition uses 'set'):

A system is a set of independent but mutually relevant terms

The key idea is that of *mutual relevance.*

A system as a whole has a *systemic attribute*

Its terms adopt different *characters* though all are of the same type or *designation.*

SYSTEMIC ATTRIBUTE

TERM DESIGNATION

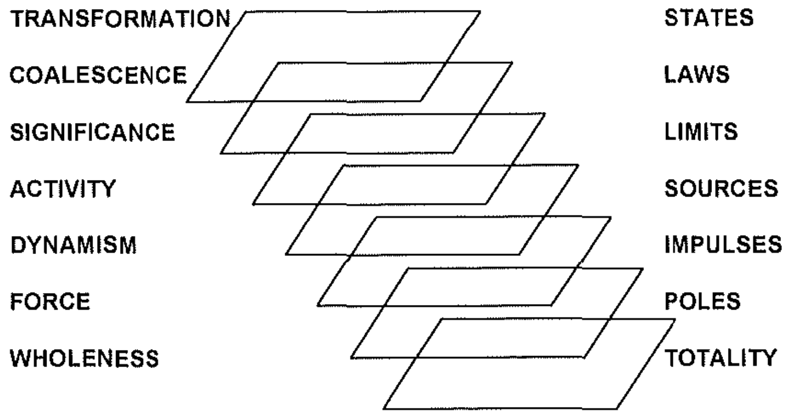
TERM CHARACTERS

For example, for the triad or three term system we have:

DYNAMISM

IMPULSES: Affirmative, Receptive, Reconciling

One of the most striking aspects of Bennett's scheme is that each of the number-systems is granted its own *view of reality.* It is as if reality were sliced in different ways, each valid in its own terms. We cannot just add on another term to get from one system [N] to the next [N + 1] because the 'next' term will be of a different nature. In the sketch the left hand column gives systemic attributes and the right term-designations (for the first seven systems):



The acceptability or otherwise of this nomenclature is not important here. They simply serve to illustrate how we could distinguish between different systems and terms. The systems were given Greek names: Monad, Dyad, Triad, and so on. The nature of the Monad was particularly important since it was not that of a simple object or singular principle but rather the undifferentiated wholeness of something of interest but not yet understood, rather like Peirce's Oneness or Lacan's the Real. The terms of most systems can be more like qualities than entities or things.

Bennett's systematics can be called a *radical pluralism* and is in spirit democratic. It proposes that every system is *compatible* with every other but one system cannot replace another. In supposing this, systematics was open to the prospect of an infinite series of systems because it contained no rules whereby the series could be terminated. Only pragmatic considerations restrained the scheme to the first twelve systems.

Different N-term systems address different aspects of reality. As far as we know, Bennett was unusual in adopting this view, even though it was implicit in the Pythagorean scheme of things. Many proponents of number systems have ended up believing not only that they have identified the ultimate or final system but also that there are no other aspects to be discovered not already subsumed under their favoured system.

Bennett argued that people think and argue within different number systems without realising this and are therefore often at cross purposes. To be able to communicate with another person we need to be working in the same system. If we are thinking in terms of change (dynamism) we need to look for threes, but if we seek what is significant we need to explore fiveness.

Articulating a series of N-term systems produces a kind of narrative sequence, as it were a story-line of *progress* (the sequence continues further). It is not difficult to see a picture of *evolution* in this sequence, or something like that.

TRANSFORMATION

COALESCENCE

SIGNIFICANCE

ACTIVITY

DYNAMISM

FORCE

WHOLENESS

Bennett made much of his proposal that there are no entities or beings until the fifth stage of the pentad. The series of systems might indicate the passing from possibility into actual existence. In this view, the universe is 'seeking' to bring into existence ever more complex or interesting forms.

Multi-term systems are mutually relevant to each other just like the terms within any system. A series as an ordered sequence is one way in which this can be represented. It is also projected that whole systems can act as terms within a *structure*. For example, there could be a triad of tetrads.

Whole dictionaries of synonyms for systemic attributes could be compiled. The meaning Bennett gives his attributes requires extensive discourse. The words have no magical relation to 'their' numbers. They are class concepts with diverse members. Once systemic attributes are defined, designations of terms have to conform.

All this would be next to impossible to spell out working from first principles – and we are not as yet sure first principles exist at all – but Bennett had recourse to an obvious method: finding good exemplars or *paradigms* of the systems and generalising from them. For example a 'good' exemplar would be one that has stood the test of time or originated with a respected authority. Here we cite his use of the Aristotelian tetrad of causes: material, efficient, final, and formal.

FORMAL

EFFICIENT **FINAL**

MATERIAL

He linked this (relevant to the activity of science) to:

IDEAL

PRACTICAL THEORETICAL

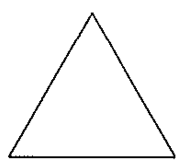
ACTUAL

Both of which emphasise the tetrad as an intersection of two axes – motivational and instrumental. However there are other forms of the tetrad.

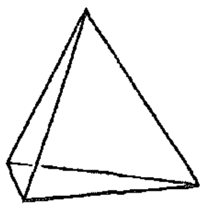
We need to look into the ways in which a system can take different forms

PARTITIONS AND THE SHAPES OF SYSTEMS

A triad can be looked at as [2 + 1], a tetrad as [3 + 1], a pentad as [4 + 1], and so on, conforming to Whitehead's view of successive conjunctions. The form of the triad lends itself to treating the third term as mediating the other two and being somewhat apart from them (impartial) or above them (higher). These relative judgments come from the very form or *shape* we give the system. Thus the partition [2 +1] suggests the third term as a *mediation* between the other two.

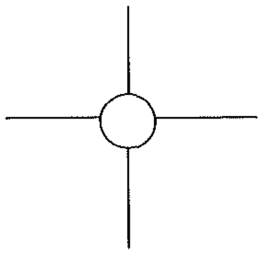


The tetrad as [3 +1] is pictured as a tetrahedron.

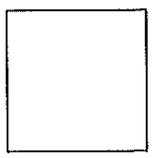


Such an image lends itself to considering the fourth term as a *unification* of the other three.

The pentad as [4 +1] is most easily seen as an *emergence* from the tetrad:

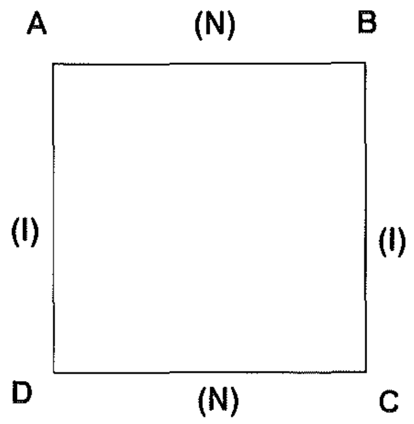


The cross form of the tetrad corresponds to the partition 4 = 2 + 2 but so does the square:



The square will encourage us to look for *parallelisms.*

Conceptually the partition also suggests a complementarity of two kinds of opposition, call them 'negation' (N) and 'inversion' (I).



This maps into such versions as: A, not-A, A and not-A, neither A nor not-A.

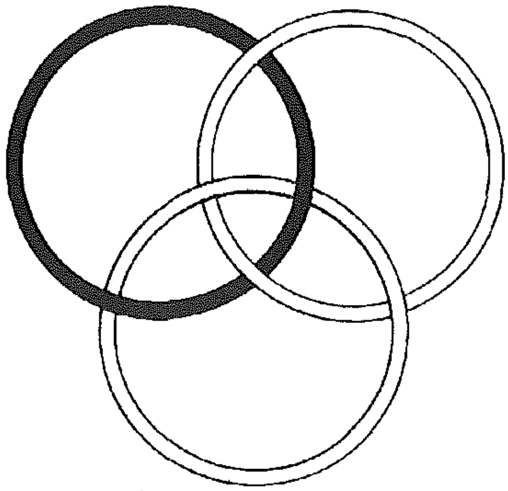
These examples suffice to indicate that, for a given N-term system, we can find a number of different ways of representation that will strongly affect our interpretation of its meaning. Partition theory gives us some of the possibilities but not all, not least because it does not include the multiplicative options (such as 6 = 2 x 3). There is also the variation caused by having to represent systems in two-dimensions, giving rise to different images of them.

THE HOLY GHOST OF SYSTEMS

N-term systems or number-ideas do not come within the accepted scope of number theory, though mathematicians such as Ian Stewart and Rudy Rucker have played with such ideas. As we have seen, only the most elementary operations such as addition or the most elementary geometry have been called into play. It is possible to speculate that, given time, some elements of recent mathematics will find their way into explicating number ideas. But very little accommodates to the view that every integer signifies a distinctive way of understanding the world and this can be treated dismissively as an archaic left-over from more primitive stages of mathematical development, remaining now only in the world of superstition. All we can suggest is that it may not be so.

Bennett's definition of a multi-term system rests on the concept of *mutual relevance.* This is obviously associated with *meaning.* It is logical in character but far from classical logic. Indeed, part of the impetus for Bennett's thinking was to find non-dyadic logics far from true-false dichotomy and also vistas beyond right and wrong. The phraseology makes clear that he was looking for an equality of status between the terms in a system. More subtly, he avoided having to introduce anything *between* the terms or even talking about one of them acting on another – which obviously would have led to an infinite regression.

Nearly all our thinking is predicated on one thing acting on another, or binary operations. This is so entrenched that most people find it impossible not to think in such terms. A genuinely threefold action is rarely countenanced, though maybe it has emerged in what are called *hyperstructures*; combinations of Brunnian rings linked together in such a way that cutting any of them breaks the whole. They are related to the well-known Borromean rings sometimes used to symbolise the Trinity.



Dealing with the interactions between pairs of things suits our normal world- view of separate agents acting locally. In contrast, the idea of mutual relevance can appear mystical and holistic. It evokes a paradigm of people coming together in a state of mutual understanding. Bennett himself was much influenced by the Christian concept of the *communion of saints.* On a less mystical level, we can consider the significance of co-operation in human enterprise. Especially in recent times, people have sought ways of designing or understanding groups containing different perspectives or ways of thinking that in conjunction can 'see the whole'. It has long been thought that only when both men and women are involved and taken equally seriously can something be understood.

This leads us to reflect on the nature of the terms of a system. To be able to be mutually relevant they need to be of the same kind – *homoousias* is the word used in relation to the Persons of the Trinity. This entails that when we move to another system, the nature of the terms must correspond. It is not 'correct' simply to 'add on' something else.

The history of art affords us a glimpse of what the logic of multi-term systems is like. One of the postulates of systematics that was implied but not highlighted in our discussion is: *the emergence/creation/realisation of a new term changes the meaning of all the terms.* T. S Eliot wrote:

> "What happens when a new work of art is created is something that happens simultaneously to all the works of art which preceded it. The existing monuments form an ideal order among themselves, which is modified by the introduction of the new (the really new) work of art among them. The existing order is complete before the new work arrives; for order to persist after the supervention of novelty, the whole existing order must be, if ever so slightly, altered; and so the relations, proportions, values of each work of art toward the whole are readjusted; and this is conformity between the old and the new. Whoever has approved this idea of order, of the form of European, of English literature will not find it preposterous that the past should be altered by the present as much as the present is directed by the past." (Eliot, *Selected Essays,* 15)

Bennett's metaphysics proposed that multi-term systems are forms of *will*, an idea previously mentioned here. However such a prospect exceeds our brief. In spiritual psychology *love is the* *medium of will.* Suffice it to say that the ordinary concept of will as in 'will-power' is grossly inadequate, as is our concept of *love* as an emotion between people.

When we see human groupings through the eyes of the systems it seems obvious that there is always a tendency towards dyads. When three people work together the likelihood is that two of them will be on one side and the third on the other, following the old saying 'two's company but three's a crowd'. In real life, when someone seeks to intervene in a dispute between two people he is likely to be rejected by both. It would seem that if we looked at groupings which did realise higher-term systems, then they would occur with decreasing frequency the greater the value of N. The series of systems then is akin to a heuristic or inspirational structure calling upon our creativity and resolve.

For every multi-term system there is a type of mutual relevance to be discovered and recognised. The experience or *taste* is primary since it can bring the intelligence of a given system into play without elaborate considerations. Though we do find a great deal of rank superstition in relation to number-ideas – particularly in China – there is some shared instinct about reality and human action that keeps us open to new possibilities. It is of their very nature that they address both the mental and the material equally.

This may remind us of David Bohm's triad of Matter, Mind and Meaning and of the version with which we began: IDEAS-LANGUAGE-OBJECTS.

Placing Language in the middle of the hierarchical or 'vertical' form conveys a strong sense of mediation:

IDEAS

LANGUAGE

OBJECTS

But it could be placed otherwise, even at the 'top' to signify that Language presages Mind, rather as Giambattista Vico, followed by James Joyce were inclined to suppose. In Bennett's systematics such a higher language is conjectured in his N-term or multi-term systems.

REFERENCES

Bennett, J. G.: 'The Structure of the World' in *The Dramatic Universe Vol. 3*, Bennett Books, 1990

Bennett, J. G. & Blake, Anthony: *Systematics – A New Technique in Thinking*: *http://www.toutley.demon.co.uk/systematicsbooklet.htm*

Blake, Anthony: *The Supreme Art of Dialogue*, DuVersity, 2008

(ed.) *The Bohm-Bennett Correspondence 1962-4*, DuVersity, 1998

Bohr, Niels: *The Philosophical Writings of Niels Bohr, Vol. IV Causality and Complementarity*, Ox Bow Press, 1998

Bohm, David: *Unfolding Meaning*, Routledge, 1987

Bonner, Anthony: *Doctor Illuminatus – A Raymond Lull Reader*, Princeton, 1994

Brooks, Michael: 'Quirky New Matter via Mathematics', *New Scientist*, January 8, 2011

Burke, Kenneth: *A Grammar of Motives*, Berkeley, 1969

Burkert, Walter: *Lore and Science in Ancient Pythagoreanism* Harvard, 1974

Deutsch, David, *The Fabric of Reality - towards a theory of everything*, Penguin, 1998

Douglas, Mary: *Thinking in Circles*, Yale, 2010

Eco, Umberto: *The Search for the Perfect Language*, Wiley-Blackwell, 1995

Eckhart, Meister: *Meister Eckhart, from Whom God Hid Nothing: Sermons, Writings and Sayings*, Shambhala, 2005

Eliot, T. S.: *Selected Essays*, Faber and Faber, 1999

Heidegger, Martin: *An Introduction to Metaphysics,* Yale, 1976

Ifrah, Georges: *The Universal History of Numbers*, Harvill, 1998

Kanigel, Robert:*The Man Who Knew Infinity: Life of the Genius Ramanujan*, Abacus, 1992

Kennedy, Jay: 'New Research on Plato and Pythagoras' http://personalpages.manchester.ac.uk/staff/jay.kennedy/

Laurikainen, K. V.: *Beyond the Atom: The Philosophical Thought of Wolfgang Pauli*, Springer-Verlag, 1988

Lossky, Vladimir: *Mystical Theology of the Eastern Church,* SVS, 1997

Marshak, Alexander: *Roots of Civilisation*, Weidenfeld & Nicholson, 1972

McClain, Ernest: *The Myth of Ivariance*, Hays, 1990

McWhinney, Will: *Paths of Change,* Sage, 1992

Peirce, C. S.: *The Essential Peirce, Vol. 1*, Wiley, 1992

Pensinger, William & Nha Trang: *The Moon of Hoa Binh*, Autopoy, 1995

Russell, Bertrand: *Principles of Mathematics, Forgotten Books, 2010*

Shah, Idries: *The Sufis*, Octagon, 1999

Thom, Alexander: *Megalithic Sites in Britain,* Oxford, 1967

Weightman, Simon: *Manjhan Madhumalati, an Indian Sufi Romance*, Oxford, 2001

Vico, Giambattista: *The New Science,* Cornell, 1984

Whitehead, A. N.: *Process and Reality*, MacMillan, 1979

Digital Text Representation Expression and Content

Dino Buzzetti
University of Bologna

Processable data and text representation

What is a digital text representation essentially helpful for? In computer science, digital data, that is to say information represented in whatsoever binary format, are essentially functional to specific forms of processing, and it can be assumed that the same obtains in computer applications to humanities. Text in digital form is essentially data to be processed. In my opinion, as I have written elswhere, 'the true rationale of a genuine digital edition consists precisely in taking advantage of the digital form of representation to improve our critical engagement with the text through effective computational processing.'[1] But how?

First of all, a clarification on the notion of text in a digital environment is here in order. From a computer scientist's point of view, the text is conceived of as 'information coded as characters or sequences of characters' and not as 'literary material as originally written by an author.'[2] The text is then defined simply as a data type for the manipulation of alphanumeric symbols arranged in a linear order. This notion of the text is obviously discrepant from a literary critic's understanding of the text. What cannot be overlooked, from this point of view, is the semiotic nature of the text, that consists not only in a sequence of graphic symbols, but also in what these symbols mean for a writer and a reader. Accordingly, it has been maintained that 'the text is not a physical reality at all but a concept-limit [*Grenzbegriff*]'; and, if 'the nature of the text is not material,' we may say that 'the text is, therefore, only' and 'always an image.'[3]

There is something remarkable about the nature of the text to be gathered here. If the text cannot but be an image of something, which is not of a physical nature, the text is made of two components. According to the stucturalist linguist Louis Hjelmslev, 'the sign,' or for that matter what we refer here to as an image of the text, 'is an entity generated by the

---

[1] D. Buzzetti, *Digital Editions and Text Processing*, in M. Deegan and K. Sutherland (eds.), *Text Editing, Print, and the Digital World*, Aldershot, Ashgate, 2009, pp. 45-62, p. 46.

[2] A. C. Day, *Text Processing*, Cambridge, Cambridge University Press, 1984, p. 1.

[3] C. Segre, *Introduction to the Analysis of the Literary Text* [1999], Engl. transl. by John Meddemmen, Bloomington, Ind., Indiana University Press, 1988, pp. 301, 315.

connection between an expression and a content.'[4] *Expression* and *content*, then, are the two basic components of the text. Broadly speaking, 'the expression plane refers to the material aspect of the linguistic sign, the content plane to the semantic aspect, there not necessarily being a one-to-one correspondence between both aspects of the linguistic sign.'[5] Apparently, this is due to the fact that the material aspect, or the image of the text, is always an approximation to the semantic aspect, whose complete characterization cannot be exhausted by any of its possible images. It is an intrinsic property of natural language, that there are many ways of expressing the same content, just as there are many ways of assigning a content to the same expression. The relationship between expression and content, therefore, is not a one-to-one relationship, being rather a one-to-many compensation relationship, obeying a kind of indetermination principle: if you fix the expression, its content remains undetermined, and vice versa, if you fix the content, its expression results equally indeterminate. Accordingly, *synonymy* and *polysemy*, that 'are relations between [linguistic] form and meaning,' or in other words relations between expression and content, can be defined as follows: 'synonymy: more than one form having the same meaning,' and 'polysemy: the same form having more than one meaning.'[6] The relation between expression and content is then, essentially, an indetermination relationship between the two basic components of the text.

All this has a bearing on the adequacy of the kind of processing we may apply to a digital representation of the text. On the one hand, processing text in the literary or ordinary sense cannot consist in processing mere strings of characters, for one would process only the image, or the expression, and not the content of the text. On the other hand, the approach proposed by the practitioners of the so-called hard artificial intelligence, grounded as it is on formalization, seems to overlook the basic indeterminacy of the relationship between expression and content in ordinary language. Formalization amounts to ensure a direct correspondence between syntax and semantics. According to Donald Davidson, to formalize or to 'give the logical form' to a sentence is 'to describe it in terms that bring it within the scope of a semantic theory.'[7] And more explicitly, John Haugeland advocates the following 'Formalist's Motto: "You take care of the syntax and the semantics will take care of

[4] L. Hjelmslev, *Prolegomena to a Theory of Language* [1943], Engl. transl. by Francis J. Whitfield, Madison, Wis., University of Wisconsin Press, 1963, p. 47.

[5] H. Bussmann, *Routledge Dictionary of Language and Linguistics*, translated and edited by Gregory Trauth and Kerstin Kazzazi, London, Routledge, 1996, p. 245.

[6] G. N. Leech, *Semantics: The Study of Meaning*, Harmondsworth, Penguin Books, 1974, pp. 101-102.

[7] D. Davidson, Essays on Actions and Events, Oxford, Oxford University Press, 2001, p. 144.

itself".'[8] This assumption is based on the adoption of the Physical Symbol System Hypothesis (PSSH), first formulated by Newell and Simon in their famous Turing Award paper, which states that 'a physical symbol system has the necessary and sufficient means for intelligent action.'[9] This hypothesis implies that physical symbol systems, such as digital computers, 'when we provide them with the appropriate symbol-processing programs, will be capable of intelligent action,' such as the assignment of semantic interpretations.[10] Because the operations of a physical symbol system are formal, and because its symbols directly designate content, the formalist's motto supports the belief that formal symbol manipulations preserve meaning. But again such an approach endorses the fallacy that text processing can focus on the formal or syntactic properties of symbol systems, that is to say on the expression of the text, and not worry about processing its content, through the assignment of suitable semantic interpretations.

The presence of an intrinsic indetermination relationship between the two fundamental constituents of language and text has significant consequences. If there is no one-to-one correspondence between syntax and semantics, or in other words between each and every element in the structure of the expression and each and every element in the structure of its content, the relationship between the two components can only obtain between their respective structures taken as a whole. As a result, a variation in either structure, affects the other in its entirety and produces a gestaltic shift. Variation, in a text, is thus essentially governed by an holistic principle. Textual holism impinges on the relation between language and what it means, be it purely conceptual – its *sense* – and so confined to sole meanings, or be it actually existing – its *reference* – and so concerned with observational evidence. Holism and indetermination tie in with each other: Quine's well-known theses of translational indeterminacy,[11] epistemological holism,[12] and ontological relativity,[13] all

---

8 J. Haugeland, *Artificial Intelligence: The very idea* [1985], Cambridge, Mass., MIT Press, 1989, p. 118.

9 A. Newell and H. A. Simon, 'Computer Science as Empirical Inquiry: Symbols and Search,' *Communications of the ACM*, 19:3 (1976), 113-126, p. 116.

10 N. J. Nilsson, 'The Physical Symbol System Hypothesis: Status and prospects,' in Max Lungarella, Fumiya Iida, Josh Bongard (eds.), *50 Years of Artificial Intelligence: Essays dedicated to the 50th Anniversary of Artificial Intelligence* (Lecture Notes in Computer Science, 4850), Berlin, Springer, 2007, 9-17, p. 9.

11 Cf. W. V. Quine, *Word and Object*, [Cambridge, Mass.], Technology Press of the Massachusetts Institute of Technology, [1960]. Quine sometimes refers to this kind of indeterminacy as *holophrastic indeterminacy*, and the claim, which 'involves the whole language,' is that 'there is more than one correct method of translating sentences where the two translations differ not merely in the meanings attributed to the sub-sentential parts of speech but also in the net import of the whole sentence' (P. Hylton, 'Willard Van Orman Quine,' in *Stanford Encyclopedia of Philosophy*, URL = http://plato.stanford.edu/entries/quine, accessed 26.02.2011).

have to do with the relation between language and what it means, and all confirm the point. Textual indeterminacy, then, implies that language and text cannot be restrained to a unique assignment of syntax and semantics. Language and text, therefore, are to be conceived of as essentially mobile and dynamical systems.

A textual and literary critic, Jerome McGann, describes this constitutive feature of the text in a paradoxical way: 'Text is not self-identical.'[14] Textual ambiguity comes about precisely because 'text is dynamic and mobile and textual structures are essentially indeterminate,' so that 'neither the *expression* nor the *content* of a text are given once and for all.'[15] According to McGann, a text is endowed with 'perceptual features,'[16] that relate to its expression, and a *Gestalt* shift in the perception of formal textual patterns opens 'doors of perception' towards 'new' interpretational 'opportunities and points of view.'(288) So, if it can be said, that 'the invariant rule of the textual condition' is 'variation,'[17] this is clearly due to textual indetermination and holism. The two basic textual components, the expression and the content, behave very much in the same way as do organic wholes in biology. The biologist William Morton Wheeler states that in biological organisms 'the whole is not merely a sum, or resultant, but also an emergent novelty, or creative synthesis,' and describes the 'unique qualitative character of organic wholes' as 'due to the peculiar non-additive relations or interactions among their parts.'[18] In a living system,

---

12 Cf. Id., 'Two Dogmas of Empiricism,' in *The Philosophical Review*, 60:1 (1951), 20-4. 'We could sum up this thesis – sometimes called epistemological holism – in two points: (1) no knowledge is a priori and immune to empirical refutation, and no knowledge is completely theory-independent; (2) in cases of conflict between theory and observations we cannot summon certain statements in isolation; the whole system of beliefs, or large parts thereof, must stand to trial' (S. Bem and H. Looren de Jong, *Theoretical Issues in Psychology : An introduction*, London, SAGE, 2006², p. 68).

13 Cf. Id., *Ontological Relativity and Other Essays*, New York and London, Columbia University Press, 1969. According to Quine, 'while it is possible to verify or falsify whole theories, it is not possible to verify or falsify individual statements' (Entry 'W.V. Quine,' in *Wikipedia*, URL = http://en.wikipedia.org/wiki/ W._V._Quine, accessed 26.02.2011), for 'there is more than one way of translating sentences' and 'the various versions differ in the reference that they attribute to parts of the sentence but not in the overall net import that they attribute to the sentence as a whole' (P. Hylton, 'Willard Van Orman Quine,' cit.).

14 D. Buzzetti and J. McGann, 'Critical Editing in a Digital Horizon,' in L. Burnard, K. O'Brien O'Keeffe, and J. Unsworth (eds.), *Electronic Textual Editing*, New York, The Modern Language Association of America, 2006, 53-73, p. 64. For a more thorough discussion of this assertion, see J. McGann, *Radiant Textuality: Literature after the World Wide Web*, New York, Palgrave, 2001, especially chapter 5 and the Appendix to chapter 6.

15 Ibid.

16 J. McGann, 'Visible and Invisible Books: Hermetic Images in N-Dimensional Space,' in *New Literary History*, 32:2 (2001), 283-300, p. 297.

17 Id., *The Textual Condition*, Princeton University Press, 1991, p. 185.

18 W. M. Wheeler, 'Emergent Evolution and the Social,' in *Science*, 64:1662 (1926), 433-440, p. 443.

the 'relations' that 'determine the dynamics of interactions and transformations it may undergo' is what another biologist, Francisco Varela, calls its 'organization.'[19] Organization acts as an informational holistic principle of 'mutual interconnection' (102) and as Norbert Wiener would say, organization has to be though of as '"information," not matter or energy.'[20] Moreover, a living system, can be described as 'autopoietic,' in as much as it 'generates and specifies its own organization,'[21] and since in a living system 'what makes it a unity with identity and individuality' is its own 'invariant organization,'(26) that is, its intrinsic, self-defining and self-regulating information content, autopoietic systems 'are unities because, and only because, of their specific autopoietic organization.'(15) Also these features of a biological system can be observed in a text. The relations the organization consists of are described by Varela as 'co-dependent,'(xv) and in McGann's opinion, textual artefacts – that is to say, 'print and manuscript encoding systems' and 'technologies' – are 'organized under a horizon of co-dependent relations.'[22] So we can say that 'like biological forms and all living systems, not least of all language itself, textuality is a condition that codes (or simulates) what are known as autopoietic systems,'[23] and we can think of literary texts precisely as 'paradigms of those interactive and feedback mechanisms that Humberto Maturana and Francisco Varela have studied as, and called, autopoiesis.'[24] Of his book, *The Textual Conditition*, McGann writes that it 'attempts to sketch a materialist hermeneutics,' in as much as it 'considers text as autopoietic mechanisms operating as self-generating feedback systems that cannot be separated from those who manipulate and use them.' To be more precise, textual autopoiesis 'functions through a pair of interrelated textual embodiments,' namely a content and its expression, that 'we can study as systems of linguistic and bibliographic codings.'(15) Finding a suitable computational model for complex textual phenomena of this kind is thus the obvious challenge faced by scholars aiming at a digital representation of the text up to befitting standards of critical enquiry.

Available technologies

A digital representation of the two basic components of the text, the expression and the content, has already been implemented, but through

---

19 F. J. Varela, *Principles of Biological Autonomy*, New York, North Holland, 1979, p. 9.

20 N. Wiener, *Cybernetics*, 2nd ed., Cambridge, Mass., MIT Press 1961, p. 132.

21 Varela, *Principles*, p. 13.

22 J. McGann, 'Marking Texts ofMany Dimensions,' in S. Schreibman, R. Siemens and J. Unsworth (eds.), *A Companion to Digital Humanities*, Maiden, Mass., Blackwell Publishing, 2004, 198-217, p. 200.

23 Id., 'Texts in N-Dimensions and Interpretation in a New Key,' in *Text Technology*, 12:2 (2003), pp. 1-18, p. 7.

24 Id., *The Textual Condition*, p. 11.

different techniques for each component and with a different approach. The primary base for the processing of textual data was laid by the introduction, back in the 1960s, of character codes, such as ASCII and EBCDIC, proposed and developed by the U.S. computer manufacturing companies. Character codes were soon extended to handle special characters for other European languages and non-Latin alphabets. But the mere sequence of encoded characters is not enough to represent all the information contained in a manuscript or in a printed page. This need led to the development and use of *markup languages*, first introduced to provide instructions as to how a printed document should look.[25] The use of markup languages was then extended to provide not only display instructions, but also meaning or semantics to words or phrases, or to provide processing instructions. What made that possible, was the introduction of 'generic' markup languages, such as Generalized Markup Language (GML) developed at IBM, which later became an ISO standard as Standard Generalized Markup Language (SGML).[26] This last so-called 'generic' or 'descriptive' language is the parent of current languages now generally used, such as HyperText Markup Language (HTML) and eXtensible Markup Language (XML). HTML is a markup language designed to allow links between documents and enriched, after the introduction of graphical interfaces, to serve the visualization of documents in a browser. XML is a more powerful markup language to provide structure and meaning within documents, and to enable the exchange of data. As a matter of fact, SGML and XML are not properly markup languages, but they have been described as metalanguages, because they provide the rules to define particular markup languages or set of tags applicable to a given purpose. HTML, or, to give another example, the *Guidelines for Electronic Text Encoding and Interchange* of the Text Encoding Initiative (TEI),[27] a language introduced for the

---

25 *Markup* consists of codes or tags attached to given strings of characters to describe their properties.

26 'Historically, electronic manuscripts contained control codes or macros that caused the document to be formatted in a particular way ("specific coding"). In contrast, generic coding, which began in the late 1960s, uses descriptive tags (for example, "heading", rather than "format-17"). Many credit the start of the generic coding movement to a presentation made by William Tunnicliffe,' of the Graphic Communications Association (GCA), to a meeting of the Canadian Government Printing Office in September 1967, entitled 'The separation of information content of documents from their format' (Ch. F. Goldfarb, *The SGML Handbook*, New York, Oxford University Press, 1990, 'Appendix A: A brief history of the development of SGML, p. 567). The 'content of documents,' made, as it is, of coded characters, is not to be confused with the *content* of the *text*, for a *document* is an 'image' or the *expression* of the text, that is made of graphic signs or, for that matter, of their digital representations.

27 'The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics' (URL = http://www.tei-c.org/, accessed 05.03.2011). The earlier releases of the Guidelines, P1, P2 and P3, were

processing of literary texts, are properly called languages or SGML/XML applications. SGML is a generalized data representation language and all descriptive markup languages are suited to the representation of data structures such as a marked up string of characters. They are applicable, therefore, to the representation of the *expression* of the text.

To represent the information *content* of a text, either data modelling languages or semantic annotation languages can be used. A data modelling language is 'a mathematical formalism with a notation for describing [(a)] data structures and [(b)] a set of operations used to manipulate and validate that data.'[28] The two elements, (a) and (b), constitute a data model, and its definition requires the specification of both a syntax and a semantics: 'the syntax, or notation, may be given formally in a grammar,' and the semantics is needed 'to refer to the properties of objects within a data model [...] and in particular to the effect (behaviour or abstract meaning) of operations on those objects.'[29] Data modelling languages, then, provide a semantics for structured data.[30] Notable examples of data modelling languages are the so-called entity-relationship-attribute (ERA) diagrams and the Unified Modeling Language (UML). Semantic annotation languages, on the other hand, have been developed, in particular, in the framework of the Semantic Web initiative to describe the information content of Web documents and resources. They are used to provide a semantics for semistructured data.[31] Notable examples are the Resource Description Framework (RDF) and the Web Ontology Language (OWL), a family of knowledge representation languages for authoring ontologies, characterised by formal semantics and RDF descriptions – or, in a more technical jargon, serialisations.

How far can we get by means of these languages in the construction of a dynamic model for text representation? Let us examine some examples. In a medieval manuscript, the following proposition might be found:

> Animal currere si homo currit est necessarium.[32]

But how is it to be interpreted? We may understand it *in sensu composito* and, by adding a comma, we can rewrite it in this form:

---

issued as SGML applications, whereas the last two, P4 and P5, have been issued as XML applications.

[28] 'Data model,' in D. Howe, *FOLDOC: Free On-Line Dictionary of Computing*, URL = http://foldoc.org/ data+model, accessed 05.03.2011.

[29] M. L. Brodie, 'Axiomatic Definitions for Data Model Semantics,' in *Information Systems*, 7:2 (1982), 183-197, p. 184.

[30] Data are usually described as *structured*, when their structure and meaning is formally defined by a data model.

[31] Although so-called *semistructured* data may have some structure, they lack a formal data model.

[32] Cf. Ricardus Sophista, *Abstractiones*, URL = http://www.hs-augsburg.de/~harsch/Chronologia/Lspost13/ RicardusSophista/ric_abst.html, accessed 05.03.2011.

Animal currere si homo currit, est necessarium.

Or else we may construe it *in senso diviso*, and then we would rewrite it in this way:

Animal currere, si homo currit, est necessarium.

In the first case, the proposition would have to be taken as true, whereas, in the second case, it would have to be taken as false. But what does a modern editor actually do by adding punctuation in his or her transcription? Does she make explicit what is implicitly there, and assume that the sense of the proposition is already given and objectively recognizable, or does she add punctuation marks to convey an external and truly metalinguistic description of the structure of the sentence, that she takes for objective as it stands, devoid of punctuation marks? As it stands, the sentence is open to different interpretations, that we may signal with diverse punctuation marks. Textual variants are the result of the indeterminacy of interpretation. On the contrary, variant interpretations are the result of textual indetermination. Any new act of writing, any variation in the expression of the text, just as a new punctuation mark, requires a variation in the content or in the interpretation of the text. In the same way, any new act of reading, any variation in the information content of the text, just as a new interpretation of a given phrase, requires a variation in the expression of the text. Again, as it has been convincingly argued by Jerome McGann, 'we may usefully regard all criticism and interpretation as deformance.'[33]

Another example, taken from a recent bestseller on punctuation, 'a marvellous punctuation-fan joke about a panda who "eats, shoots and leaves",' allows us to examine these textual phenomena in more detail.[34] Here again, the presence of a comma is decisive in determining the meaning of the entire sentence. By removing the comma, the whole meaning of the sentence undergoes a restructuring shift: the two verbs (V) 'shoots' and 'leaves' are now seen as two nouns (N) that express the object of the predicate 'eats'. It is not only the whole comprehensive meaning of the compound sentence, as in the former example, that gets changed, but here the meaning and grammatical status of individual words is recast just as well.

Now, as it as been maintained, 'punctuation is not simply part of our writing system,' for it is also 'a type of document markup that may vary and be replaced by other types of markup.'[35] So, for the sake of simplicity,

---

33 J. McGann and L. Samuels, 'Deformance and Interpretation,' in *New Literary History*, 30:1 (1999), 25-56, p. 46.

34 L. Truss, *Eats Shoots & Leaves: The zero tolerance approach to punctuation*, London, Profile Books, 2003, p. 2.

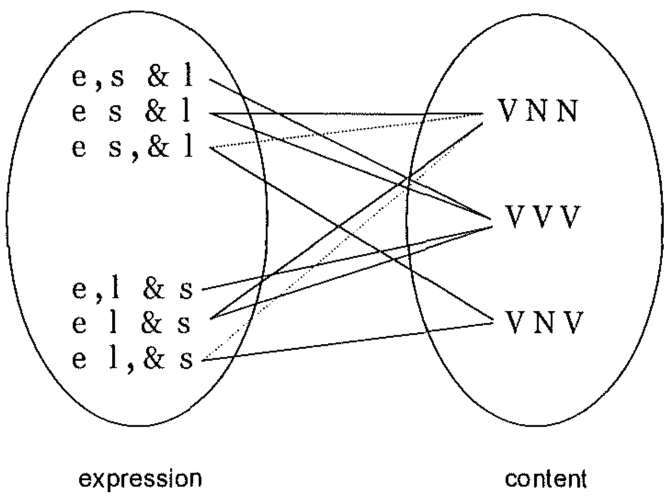
35 . H. Coombs, A. H. Renear and S. J. DeRose, 'Markup systems and the future of scholarly text processing,' in *Communications of the ACM*, 30:11 (1987), 933-47, p. 940.

we may reason on punctuation instead of markup and use this example to consider the relationship between the structure of the document – or the expression of the text – and the structure of its information content. It is the markup, or punctuation, that assigns a structure to the document. Markup is introduced precisely for the sake of 'rendering "digital stuff" as structured information,' so as to transform unstructured string of characters – mere 'digital stuff ' – into stuctural units 'accessible as information within intelligent computing environments.'[36] Punctuation then, or markup for that matter, signals possible alternatives between different structural arrangements, as it can be seen from our example:

	Eats , shoots and leaves
Semantic structure	V V V
Markup	Eats <**add**> , </**add**> shoots and leaves

	Eats shoots and leaves
Semantic structure	V N N
Markup	Eats <**del**> , </**del**> shoots and leaves

By way of illustration, we can also outline a (partial) mapping between two sets of possible structural configurations of both the expression and the content of the text (Fig. 1). Here again we can see that punctuation, or markup, acts as a special textual marker for the semantic structure of the text.



— Fig. 1 —

A dynamic model of the interactions between the two kinds of structure, the structure of the expression and the structure of the content, should be

[36] R. Cover, N. Duncan, and D. T. Barnard, 'The Progress of SGML (Standard Generalized Markup Language): Extracts from a Comprehensive Bibliography,' in *Literary and Linguistic Computing*, 6:3 (1991), 197-209, pp. 197-98.

capable of accounting for the procedures of both textual and literary criticism. An editor tries to reduce to a coherent unity the several documents and variant representations through which the text was handed down to her, whereas an interpreter proceeds from a single, coherent form of text representation to several interpretative structures, all compatible with it. In both cases, our dynamic model should take into account a one-to-many mapping: from several textual variants to their unique 'logical sum,'[37] or, vice versa, from a single 'logical sum' of compatible interpretations to their several and distinct representations. In the first case, we have to consider an ambiguous content as an indeterminate whole compatible with different forms of expression; in the other, we have to consider a number of different interpretations, or assignments of information content, all compatible with an ambiguous expression taken as an indeterminate whole.

From these observations, we may draw some further conclusions about markup and see that we may take advantage precisely of its intrinsic features to contrive a conceptual model of textual dynamics. In the first place, we can see that through punctuation, or markup, a variant interpretation can be transformed into a textual variant and vice versa.
We can also elicit that the same indetermination relationship that occurs between the expression and the content of the text is to be found between the set of possible markup structures and the set of their possible semantic representations. As it has been clearly pointed out, 'the same markup can convey different meanings in different contexts,' just as 'markup can communicate the same meaning in different ways using very different syntax.'[38] This means that the outcome of applying 'mapping rules,' from 'syntactic relations' between markup elements into semantic relations between the elements of an 'object level domain,' (6) would amount either to 'the re-tagging of documents with richer markup,' or to a new semantic description or serialisation 'in the form of RDF or a topic map' (8).[39]

But there is another feature of the markup that is crucial for the contrivance of a dynamic model of text representation in digital form. Markup exhibits a distinctive duality, that it shares again with punctuation and, more generally, with all diacritical marks. For, as it has been observed, the markup is 'simultaneously embedded and separable' from the text, and we can say that it 'is part of the text and yet it is distinct' from

---

[37] Cf. M. Thaller, 'Historical Information Science: Is There Such a Thing? New Comments on an Old Idea,' in T. Orlandi (ed.), *Discipline umanistiche e informatica*, Roma, Accademia Nazionale dei Lincei, 1993, 51–86, p. 64.

[38] D. Dubin and D.J. Birnbaum, 'Interpretation beyond markup,' presented at *Extreme Markup Languages 2004*, p. 1.

[39] Topic Maps 'is an ISO standard for describing knowledge structures and associating them with information resources' (V. Lombardi, *Metadata Glossary*, URL = http://www.noisebetweenstations.com/personal/essays/metadata_glossary/metadata_glossary.html, accessed 10.03.2011).

it.[40] Accordingly, the markup has been described either as an external 'technique for representing structure,' (3) or as that very 'structure' (4) embedded in the text. So the markup can both exhibit and describe a structural feature of the text, and 'it can perform both functions only by changing its logical status' and commuting between object-language and metalanguage.[41] Markup shares this property with diacritical signs. Markup tags are then, in fact, diacritics and like all diacritical marks can be considered as an expression of the 'reflexive metalinguistic nature'[42] of natural language, the capability that all natural languages – and texts – possess of saying something about themselves. A diacritical mark or phrase, a punctuation mark or a markup construct, is an expression of the object language, as to its expression, just as it is an expression of a metalanguage, as to its content. Diacritics can be viewed as part of the text and as separate from the text, and markup has actually been described in both ways. On the one side, it has been maintained that 'the markup is not part of the text or content of the expression, but tells us something about it,'[43] or in other words that the markup belongs to a metalanguage. On the other side, it has been acknowledged that the markup is 'constitutive of the text it characterizes,' albeit with a reservation, for this 'recognition' raises 'new puzzles about just what markup really is, and in particular, when it is about a text and when it is part of a text . . . and when, and how, it may sometimes be both.'[44] But it can indeed be both, without posing any problem, for diacritics and markup are essentially ambiguous.

And it is precisely this kind of *diacritical ambiguity* possessed by markup that can be exploited to devise a dynamic model of the text. Such a model can be expounded through a diagram, a kind of multidimensional matrix, whose elements are connected by a series of operations. The resulting process is a kind of loop. But let us examine it in detail.

---

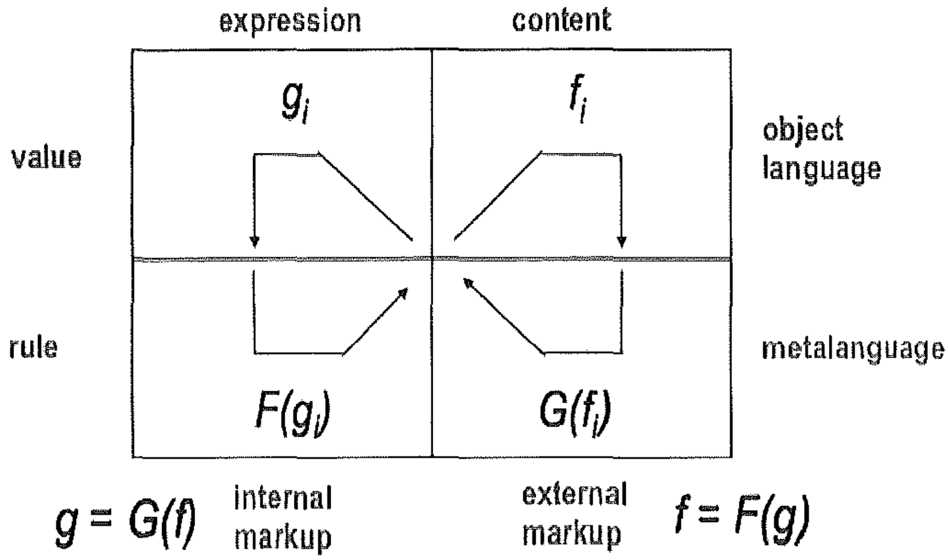
40 D. R. Raymond, F. W. Tompa and D. Wood, 'Markup Reconsidered,' paper presented at the First International Workshop on Principles of Document Processing, Washington DC, 22-23 October 1992, URL = http://softbase.uwaterloo.ca/~drraymon/papers/markup.ps, accessed 10.03.2011, p. 3.

41 Buzzetti and McGann, 'Critical Editing in a Digital Horizon,' p. 63.

42 Cf. T. De Mauro, *Minisemantica dei linguaggi non verbali e delle lingue*, Bari, Laterza, 1982), pp. 93-4, and Id., *Prima lezione sul linguaggio*, Bari, Laterza, 2002, pp. 89 and 91-93.

43 Coombs et al., 'Markup systems,' p. 934.

44 A. Renear, 'The descriptive/procedural distinction is flawed,' in *Markup Languages: Theory & Practice*, 2:4 (2001), 411-420, p. 419.



– Fig. 2 –

The structural elements of the *expression* of the text, represented by embedded or *internal markup*, are here arranged in the first column ot the table. In the second column we find structural elements of the *content* of the text, as described by data modelling or semantic description languages, that can also be regarded as a form of *markup*, albeit *external*. Now, one and the same internal markup construct can be seen either as belonging to the *object language* of the text, in as much as it is a structural element of its expression, or else as a representation of that very element, separate from the text and belonging to a *metalanguage*. These two aspects of a markup construct can be severed, and the operation that converts the one into the other is a logical move, that rests on the assumption that the 'meaning of the markup' is 'the set of inferences about the document that are licensed by the markup.'[45] Accordingly, this move posits a markup construct as an inference-licence. If so, we can place it in the lower part of the first column and regard it, to recall Gilbert Ryle's famous description, as an 'inference-ticket,' or a *rule*-statement 'to move from asserting factual statements to asserting other factual statements'[46] – in our case, to infer from a statement about an observed textual property, to a statement about a property of its content. That content property, in its turn, expressed in a semantic annotation language, can be placed in the upper compartment of the second column as the *value* of the operation prompted by the instruction found in the lower compartment of the first column. All this

[45] C. M. Sperberg-McQueen, C. Huitfeldt and A. Renear, 'Meaning and Interpretation of Markup,' in *Markup Languages: Theory & Practice*, 2:3 (2000), 215-34, p. 231.

[46] G. Ryle, *The Concept of Mind*, London, Hutchinson's University Library, 1949 , p. 121.

means that markup can have both 'descriptive' and 'performative' force,[47] and what has just been said about markup constructs, or the structural elements of the expression of the text, applies also to semantic annotation constructs, or the stuctural elements of its content. We can therefore posit a semantic description as a rule, place it in the lower part of the second column, and move from it to the value of the operation it commands, ending up again with a property of the expression, in the upper part of the first column. And so the cycle is complete.

A plausible logical explanation can actually be provided for the whole series of operations represented by this conceptual model. As Ryle reminds us, inference-licences 'belong to a different and more sophisticated level of discourse from that [...] to which belong the statements of the facts that satisfy them.'[48] In other words, rule-statements are to be seen as 'second-order object-language statements,' or statements based on a second-order form of predication, that 'are equivalent to first-order metalinguistic statements,' or statements based on an ordinary form of predication.[49] So the logical import of markup expressions understood as rules, or inference-licences, is different from the logical import of markup expressions construed as factual statements about observed textual properties. And this ambivalence of markup expressions is paralleled by the 'double sense' acquired by 'the only explicit symbol' of Spencer Brown's calculus of indications, the mark of distinction, since 'on the one hand it represents' an operation, namely 'the act of distinction,' and 'on the other hand it is a value,' namely 'the content of a distinction,' or the result of an operation.[50] The mark, then, can be a mark of ambiguity, for in a sense 'there can be no' real 'separation between distinctions and acts of distinguishing.'[51] The separation can be only formal and this kind of ambiguity may also be seen as an instance of the notorious *distinctio formalis a parte rei*, Duns Scotus has more than often been blamed for.[52] Be it as it may, the analogy shall actually be of further use.

Can we now derive from this model, merely conceptual, a formal mathematical model of textual dynamics? This is the challenge that a digital tepresentation of the text apparently has to face. But before turning to this question, a clarification on another technological issue is here in

---

47 Renear, 'The descriptive/procedural distinction,' p. 419.

48 Ryle, *The Concept of Mind*, p. 121.

49 Buzzetti, *Digital Editions and Text Processing*, p. 57.

50 Varela, *Principles*, p. 111. Cf. G. Spencer Brown, *Laws of Form*, London, Allen & Unwin, 1969.

51 L. H. Kauffman and F. J. Varela, 'Form dynamics,' in *Journal of Social and Biological Systems*, 3:2 (1980), 171-206, p. 205.

52 For a formal reconstruction of Scotus' formal distinction, see D. P. Henry, *Medieval Logic and Metaphysics: A modern introduction*, London, Hutchinson University Library, [1972], pp. 88-95.

order. The adoption of SGML, or XML, as 'a standard for encoding textual data,'[53] has had the doubtful consequence of suggesting an 'OHCO structure,' i.e. 'an ordered hierarchy of content objects,' as the 'basic model of the text.' (6) A 'content object' is a portion of a 'document' (5) that contains or is contained within other content objects, or portions of the document, and that forms with them a 'hierarchy' of containment relations, the smallest elements of which are 'ordered' in succession, in the sequence of characters that form the document. The shortcomings of the OHCO model are quite obvious. By equating the structure of the text with the structure of its expression, it engenders a confusion between the document and the text. Moreover, it does not take into account the 'limitations'[54] of current 'SGML-based markup systems, that cannot handle "overlapping hierarchies",'[55] that is to say features that do not nest within other features. Structuring a poem by verse would not allow for a concurrent structure by grammatical units spanning over different verses.

A solution to this problem has been proposed through the introduction of the so-called Multi-Version Document (MVD) model. According to Desmond Schmidt, who has devised it, 'overlap is a serious problem in the encoding of cultural heritage texts,' and a problem that cannot be solved through markup, for it resides in 'the technical limitations of embedded markup itself.'[56] An MVD model represents 'all the versions of a work' as 'a directed graph, with one start node and one end-node,' (350) as shown in Fig. 3 (a). This

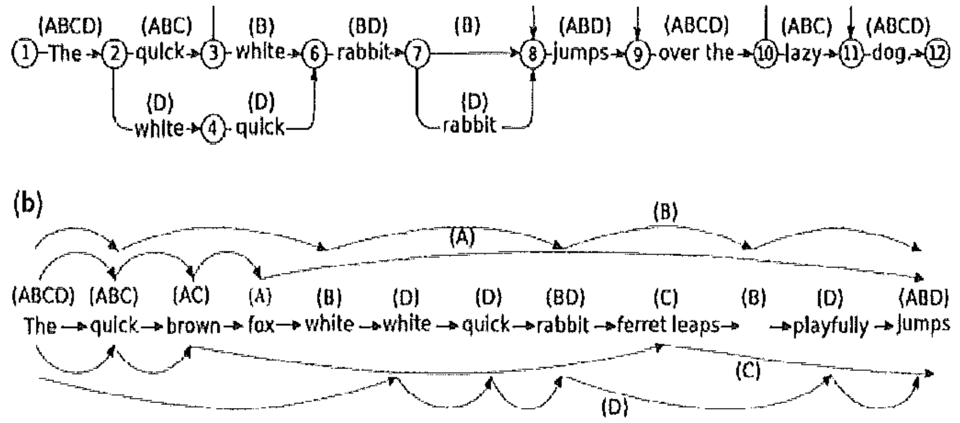
---

[53] S. J. DeRose, D. G. Durand, E. Mylonas, and A. H. Renear, "What Is Text, Really?," in Journal of Computing in Higher Education, 1:2 (1990), 3-26, p. 18.

[54] W. Piez, 'Form and Format: Towards a Semiotics of Digital Text Encoding,' in *Digital Humanities 2007: Conference Abstracts*, 153-57, p. 156. According to the author, current markup systems 'can gracefully handle only a single organizational hierarchy at a time' (Ibid.). But see note 53 below for a partial amendment of this view.

[55] A. Witt, 'Multiple Hierarchies: New aspects of an old solution,' *Extreme Markup Languages 2004*, p. 4. The author points out, though, that 'different structures often can be expressed within one hierarchy,' when 'none of [the] elements belonging to [...] different tag sets overlap': so, 'in practice,' it is possible to find also 'multi-hierarchically' structured texts compatible with 'SGML-based markup systems,' although 'from a formal point of view' they can only 'allow for the representation of exactly one hierarchy' (p. 1). The 'problem of overlapping hierarchies' was first discussed in a paper by D. Barnard, R. Hayter, M. Karababa, G. Logan, and J. McFadden, 'SGML-based Markup for Literary Texts: Two problems and some solutions,' in *Computers and the Humanities*, 22:4 (1988), 265-276.

[56] D. Schmidt, 'The Inadequacy of Embedded Markup for Cultural Heritage Texts,' in *Literary and Linguistic Computing*, 25:3 (2010), 337-356, p. 348.



– Fig. 3 –

'variant graph,' as it has also been called, comprises 'a direct analogue' of 'each type of editing operation: deletion, insertion, replacement and transposition.'[57] Alternatively, an MVD model 'can be serialized as a list of paired values, each consisting of a fragment of text and a set of versions to which that fragment belongs,'[58] as shown in Fig. 3 (b). An MVD graph actually works as the 'logical sum' of all the versions of a text and since 'XML documents are simply text files'[59] all different semantic descriptions of a document produced by an XML semantic annotation language can be represented through an MVD graph. A mapping between markup structural elements and semantic description structural elements becomes thus a mapping between elements of two isomorphic graphs. It remains to enquire whether this circumstance can help in finding a mathematical function to represent the relations between the textual and the interpretational variants of a text.

A mathematical model ?

A formal model for dynamic text representation should account for two essential aspects of textual mobility, namely self-reflexivity and indetermination. In what follows, only a few hints shall be surmised on how these two basic issues could possibly be approached.

---

[57] D. Schmidt, R. Colomb, A data structure for representing multi-version texts online, in International Journal of Human-Computer Studies, 67:6 (2009), 497-514, p. 503.

[58] Schmidt, 'The Inadequacy of Embedded Markup,' p. 350.

[59] K. Williams [et al.], *Professional XML databases*, Wrox Press, Birmingham, 2000, p. 2.

Textual self-reflexivity originates from the interconnexion between the expression and the content of the text. The relation of mutual dependence between the expression and the content is undetermined and conveys mobility and dynamism to textual structures, 'which may become stable as definite structural forms either of the expression or of the content, and may, reciprocally, determine the instability either of the corresponding content or of the corresponding expression.'[60] To one and the same arrangement of the expression many possible content assignments may be related, and vice versa. The text, then, is not identical to itself, because of its self-reflexive instability. Jerome McGann derives this very 'law of non-identity'

$$A = A \Leftrightarrow A \neq A \qquad [1]$$

from the distinction produced by the primary partition of the text into expression and content, as expounded formally by means of the 'form of distinction' introduced by George Spencer Brown.[61] Through self-reflexivity the text determines its own internal organization and can be analysed by means of the logic of autopoietic systems, as proposed by Lou Kauffman's and Francisco Varela's extension of Spencer Brown's calculus of indications.[62]

Because of the indeterminacy of the relationship between expression and content, the set of internal relations between the constituent parts of the text remains mostly implicit, and the structure of the text may be defined as the 'set of latent relations'[63] among its structural elements. The text can then be assumed as an indetrminate whole, and applied to it, Spencer Brown's primary partition can be provided by the distinction of its primary subunits, expression and content.

---

60 D. Buzzetti, 'Diacritical Ambiguity and Markup,' in Id., G. Pancaldi, and H. Short (eds.), *Augmenting Comprehension: Digital tools and the history of ideas*, London-Oxford, Office for Humanities Communication, 2004, 175-188, p. 180. In this article the following considerations on textual self-reflexivity are developed in greater detail.

61 'Texts and their field spaces,' such as the expression and the content of the text, 'are autopoietic scenes of co-dependent emergence. As such, their primal state is dynamic and has been best characterized by G. Spencer Brown's *Laws of Form* (1969), where "the form of distinction" – the act of making indications by drawing a distinction – is taken as "given" and primal (1). This means that the elementary law is not the law of identity but the law of non-identity (so that we must say that "*a* equals *a* if and only if *a* does not equal *a*"). Identities emerge as distinctions are drawn and redrawn, and the acts of drawing out distinctions emerge as co-dependent responses to the field identities that the form of distinction calls to attention' (J. McGann, 'Marking Texts of Many Dimensions,' in S. Schreibman, R. G. Siemens, J. M. Unsworth (eds.), *A companion to digital humanities*, Malden, Mass., Blackwell Publishing, 2004, 198-217, p. 212).

62 See Kauffman and Varela, 'Form dynamics,' cit.

63 C. Segre, *Introduction to the Analysis of the Literary Text* (1985), transl. by J. Meddemmen, Bloomington Ind., Indiana University Press, 1988, p. 44.

The structural instability of any image or representation of the text is the immediate outcome of this distinction. Expression and content constitute two subsystems of the whole textual system; but once considered as separate units, they constitute two new distinct and related wholes. By analysing the expression, we determine its structure in relation to its integral whole. However, for any given and self-identical structure of the expression we can have several ways of analysing its content. By equating the identity of the text with the partial unit that consists in its mere expression, the other partial unit that consists in its content remains undetermined. A symmetrical and similar phenomenon occurs by equating the identity of the text with the other partial unit that consists in its content. The phenomena of instability and indeterminacy of the structure of the text come about when we reduce the integral identity of the text to the identity of one of its two partial subunits.

The 'indication' of the expression, or its representation in Spencer Brown's calculus, and the 'indication' of the content presuppose their distinction, produced by the primary partition operating upon the whole of the text. The 'indication' of the expression makes it a subunit of the text identical to itself and determines its structure. The determination and the identity of the expression with itself is expressed formally by the law of idempotence of the expression with respect to its representation, as specified by Spencer Brown's first axiom, or 'law of calling,'[64] and 'form of condensation'(5):

The 'indication' of the content makes it a subunit of the text identical to itself, and determines its structure. In the same way, the determination and the identity of the content with itself is expressed formally by the law of idempotence of the content with respect to its representation. But how can the identity of the text with itself, expressed as the idempotence of its partial subunits with respect to their representation, both depend on and at the same time be cancelled by the primary distinction that defines them?

It can be shown that the 'law of non-identity' [1] – or law of compensation between determination and indetermination of the expression and the content of the text, as the case may be – presupposes and implies an endomorphism (f) between the structural constituents of the text:[65]

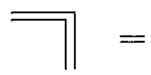
64 Spencer Brown, *Laws of Form*, p. 1.

65 See D. Buzzetti, 'Digital Representation and the Text Model,' in *New Literary History*, 33:1 (2002), 61-88, pp. 82-84.

$$(A = A \Leftrightarrow A \neq A) \Leftrightarrow A \xrightarrow{f} A \qquad [2]$$

This endomorphism maps elements of the content into elements of the expression, or conversely, elements of the expression into elements of the content. The compensation between the reciprocal determination and indetermination of the expression and the content of the text is represented by the inversion of the domain and co-domain of the endomorphism and shows itself explicitly in the ambiguity of markup constructs and the oscillation between their dual logical function, declarative and performative as it may be.

The relationship between the structure – or the logical form – of the expression of the text and the structure – or the semantic model – of its content can in turn be considered an example of Spencer Brown's second axiom, or 'law of crossing,'[66] (Spencer-Brown 1969: 2), and 'form of cancellation' (5):



For, the reference of the structural articulation of one of the two textual subunits to the structural totality of the other cancels its identity with itself and brings about its indeterminacy. In conclusion, then, the identity of the text with itself is posited by the primary partition between expression and content, and is cancelled by the crossing from one subunit to the other, which revokes the separate identity of each determinate partial unit and reintegrates the indeterminate totality of the text. The text can be considered and described, in brief, only as a holistic unit.

As the extended calculus of indications may account for the self-reflexive features of the text, another mathematical approach may account for its indeterminacy.[67] But the convergence of the two possible points of view should not be altogether ruled out.

As it has been pointed out, markup constructs are a kind of diacritical expressions that oscillate between their dual logical function, and can be seen both as a metalinguistic representation of a textual structure, and as an objectivised object-language textual structure themselves. From this

[66] Spencer Brown, *Laws of Form*, p. 2.

[67] The following considerations have already been exposed, in the essence, in D. Buzzetti, *Text, Science, and Technology: Construing text as a system*, in G. Castellani, V. Fortunati, E. Lamberti and C. Franceschi (eds.), *Biocomplexity: At the cutting edge of physics, systems biology and humanities*, Bologna, Bononia University Press, 2008 (Quaderni del Centro Interdipartimentale "L. Galvani," 1), pp. 295-320.

point of view, a 'principle of representation-theoretical self-duality' can apply to markup conceived of both as the structure, and as a representation of the structure of the text. For it is precisely the 'identification [...] between structures and the collection of all representations of the structure' that is 'expressed in the principle of self-duality' as introduced by Shahn Majid.[68]

According to Majid, 'an evaluation *f(x)* can also be read *x(f)*, where *f* is an element of a dual structure.' This kind of self-duality holds in a pure mathematical language, where 'such an "observer-observed" reverse interpretation of the mathematical structure can always be forced,' but 'will the dual interpretation also describe physics?'[69] or for that matter a text, viewed as an information carrier and physical device? In physics, as Majid has shown, Hopf algebras, one of the simplest self-dual 'categories,' or types of mathematical structures, can provide 'models in which quantum mechanics and gravity are unified into one mathematical structure.'[70] Likewise, in a text, a diacritical sign or markup element of the expression can be seen as a representation of the structure of the content, just as a structural unit of the content can be seen as a representation of the structure of the expression. A restructuring operation from an expression unity to compatible content assignments can be easily reversed, and markup elements, either internal or external, can be seen both as signs, or values, as well as instructions, or operations. In physics, self-duality implies that a theory 'should admit a "polarisation" into two halves each of which is the set of representations of the other,' so that we 'should be able to reverse interpretations.'[71] And that is precisely how we can construe the polarisation between the expression and the content of the text.

The analogy with self-dual physical systems can be assumed as a starting point for a formal description of textual phenomena and the construction of what McGann has called 'quantum poetics.'[72] In quantum mechanics, observables or 'coordinates like x, p,' that describe the position and momentum of particles, 'become operators $\mathbf{x}$, $\mathbf{p}$' that 'do not commute,' so that '$\mathbf{xp}$ no longer equals $\mathbf{px}$.' The non-commutativity of position and momentum coordinates

---

68 S. Majid, 'Principle of Representation-theoretic Self-duality,' in *Physics Essays*, 4:3 (1991), 395-405, p. 396. On Majid's philosophy of physics, see M. Heller, 'Algebraic Self-Duality as the "Ultimate Explanation",' in *Foundations of Science*, 9:4 (2004), 369-385.

69 Id., *Foundations of Quantum Group Theory*, Cambridge, Cambridge University Press, p. 293.

70 Id., 'Principle of Representation-theoretic Self-duality,' p. 402.

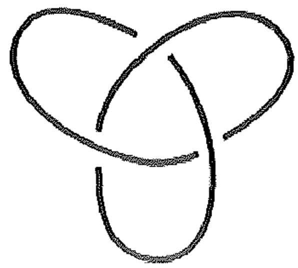
71 Id., 'Quantum Groups and Noncommutative Geometry,' in *Journal of Mathematical Physics*, 41:6 (2000), 3892-3942. Also at URL = http://arxiv.org/PS_cache/hep-th/pdf/0006/0006167v1.pdf, accessed 19.03.2011, p. 61.

72 Cf. McGann, 'Visible and Invisible Books,' p. 297.

> has the interpretation that it matters which you measure first, **x** or **p**, and this in turn is related to the famous Heisenberg uncertainty principle, that you cannot measure both of them accurately at the same time.[73]

Likewise, in a text, markup elements behave on the one side as the observable representations of structural units of its expression or its content and, on the other, as operators that rearrange textual units and produce a restructuring of the expression or the content of the text. In a text, 'a structural shift changes the entire network of internal relations and affects the whole range of textual units.' This kind of '*Gestalt* leap,' is then a 'discrete' and 'discontinuous' phenomenon.[74] Similarly, in physics, 'non-commutativity leads to a kind of "finite difference" or discretization,' which is a 'general feature' of physical self-dual structures.[75] The shift from an object-language to a metalinguistic interpretation of a diacritical mark can be seen as a shift from a classical to a quantum interpretation of the textual condition.

A structural shift introduces a temporal dimension. To take that into account, the braided structure of Fig. 2 should be extended: it should comprise a third dimension besides expression and content to represent perceptual restructuring operations.[76] The result would be a trefoil knot structure (Fig. 4), whose 'invariant,' or defining characteristic, can be described in terms of a non-commutative geometrical structure such as a quantum group.[77]

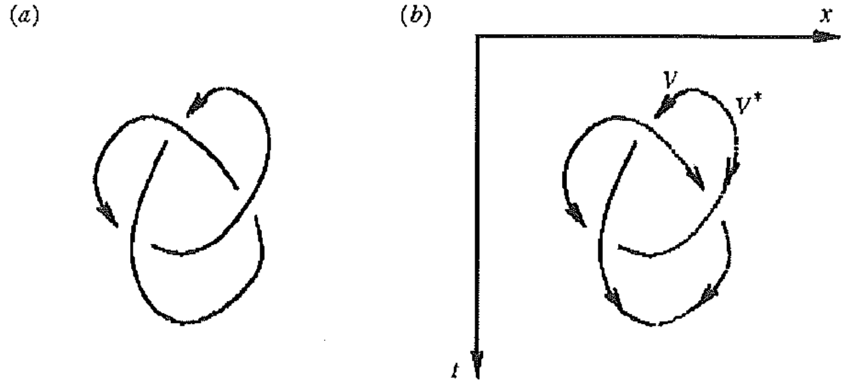


– Fig. 4 –

According to Majid, the time dimension could be introduced as shown in Fig. 5 (b), (99) where the vertical axis is interpreted as time and the knot

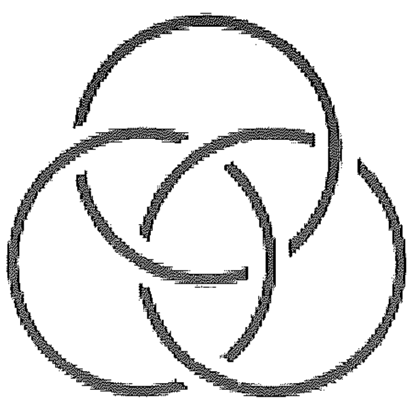
[73] S. Majid, 'Non-commutative Geometry and Quantum Groups,' in *Philosophical Transactions of the Royal Society of London*, Series A, 358:1765 (2000), 89-109, p. 90.
[74] Buzzetti, *Text, Science, and Technology*, p. 312.
[75] Majid, 'Non-commutative Geometry and Quantum Groups,' p. 91.
[76] According to Charles Sanders Peirce, a sign always involves 'thirdness.' As he writes, 'thirdness is the triadic relation existing between a sign, its object, and the interpreting thought, itself a sign, considered as constituting the mode of being of a sign.' (*A Letter to Lady Welby*, CP 8.331-332, 1904).
[77] See Majid, 'Non-commutative Geometry and Quantum Groups,' pp. 98ff.

as describing the trajectories of self-dual elements *V* and *V** flowing down the page.



– Fig. 5 –

These last considerations are purely tentative and are meant only to suggest an evocative line of research. In this respect, it may be interesting to note that psychoanalysts find it necessary to expose the kind of discourse that constitutes their analytic practice expressly through an interlacing of the Real, the Symbolic and the Imaginary[78] as represented by a structure akin to the trefoil knot, known as the Borromean link – or rings, or knot (Fig. 6):



– Fig. 6 –

'The Borromean knot, is defined as the way in which we *imagine* the *real* effect of the *symbolic*.'[79] Could we say that a textual structure lives its life

[78] Cf. J. Lacan, 'Au-delà du "principe de réalité",'(1936) in *Écrits*, Paris, Seuil, 1966, pp. 73-92.
[79] Ph. Julien, *Pour Lire Jacques Lacan*, 2 e éd., Paris, E.P.E.L., 1990, p. 221.

precisely in its enacting an analytic practice? which, after all, is as language based as a text is? But perhaps more in line with the thrust of our argument is to recognise that structures like the trefoil knot or the Borromean rings, as shown in Fig. 4 to 6, 'are topological diagrams, not geometrical representations.'[80] As such, they provide, to use Maturana and Varela's terminology, the *organisation* of an information carrying physical *structure* of sorts – be it a biological organism, a semiotic system, or a molecule, as in the use of 'DNA components' to forge molecular Borromean rings.[81]

So far we have only ventured to provide some data to build a mathematical model for a dynamic text representation, but we have not even attempted to solve the problem. Whether that would be feasible along the lines of reasoning that we have tried to suggest is left to further, more mature considerations.

---

80 Chengde Mao, Weiqiong Sun, and N. C. Seeman, 'Assembly of Borromean Rings from DNA,' in *Nature*, 386 (1997), 137-138, p. 137.

81 Ibid.

The Digital Universe: Gravity and other loose ends

Geoffrey Constable

106 Kings Road
Berkhamsted
Herts HP4 3BP

Abstract
The concept of a Digital Universe, and the deductions that flow from it, are tested against three groups of gravitational anomaly. Further deductions indicate that the existence of an ether needs deeper consideration.

1 Introduction

This paper relies extensively upon previous papers presented to ANPA, in particular upon the following.

Paper A 'A Digital Universe' (ANPA 30) which argues that

- The Universe is a digital system
- All variables are quantised in one way or another
- All particles with half-integral spin have probabilistic structures that, in the case of the electron, extend to the limits of the universe
- The probability of such a particle occurring at a distance 'x' from its mean position is proportional to $1/x$.

Paper B 'An investigation, arising from the theory of quantised variables, into inter-particle forces' (ANPA 18) which argues that

- One half-integral spin particle will act on an another such to create a force
- If both particles are stationary, the force obeys the laws of electrostatics
- If the particles are in relative rotation, the force obeys the laws of electro-magnetism
- Each particle is at the centre of its own universe (i.e. that the boundary of a universe is defined by the limit of an observer's perception, rather than by geographical limits).
- The force of gravity is caused by the universes of two adjacent particles not overlapping, the resulting end effects giving rise to a weak attractive force - gravity.

- The law of gravity thus derived approximates closely to an inverse square law at laboratory distances but can vary significantly from such a law at greater distances.
- Gravity is caused by a particle in the nucleus that has a mass equal to half that of the pion- i.e. a pion quark..
- Consequently, gravity is a comparatively short-range force that cannot operate over distances greater than the radius of the universe divided by approx 137 (that this factor resembles the fine structure constant may have meaning or may be mere coincidence).

Paper 'C' 'The theory of Quantised Variables: an investigation into the magnetic moments of elementary particles' (ANPA 19) which argues that

- The magnetic moments of the proton and neutron arise because the nucleus of each particle consists of six pions, each pion orbiting at half the speed of light.

The object of the present paper is to test some of the deductions summarised above against recently observed gravitational anomalies. Such a test leads on to whether present-day beliefs concerning the existence – or otherwise – of an ether need to be re-visited.

2 Gravitational Anomalies

2.1 Anomalies observed during full and partial eclipses of the sun

Dr Maurice Allais, a distinguished French scholar, was the first to note strange gravitational effects associated with solar eclipses. Although primarily an expert – and Nobel Laureate – in economics, Dr Allais was also a keen scientist with a particular interest in the behaviour of the Foucault Pendulum. Coincidentally, in 1954, a partial solar eclipse occurred during an extended period of observation. During this extended period Dr Allais measured the azimuth of his pendulum every few minutes.

At the beginning of the period, the azimuth rotated slowly and at a constant rate, consistent with the rotation of the Earth. However, at the start of the eclipse, the azimuth rotated backwards, and at a considerably faster rate than before.

As the eclipse continued, the rotation decreased and then reversed so that, by the end of the eclipse, the azimuth had regained its normal trajectory and rotated slowly as before.

This azimuth excursion is illustrated by the graph in Fig 1, which also shows the results of another experiment undertaken in 1956, 150 meters underground.

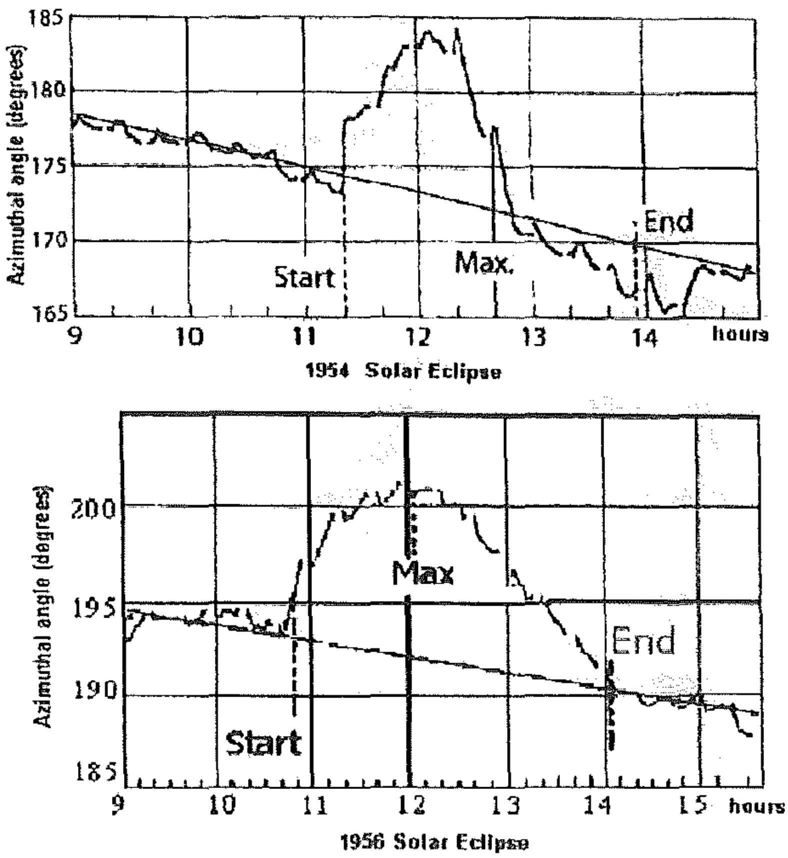


Fig 1 Two graphs showing Foucault Pendulum experiments by Dr Allais during solar eclipses.

Although not widely publicised, these experiments aroused considerable interest and led to a series further experiments, with varying results. The following summary has been taken from a schedule prepared by Dr Xavier Amador, as published in the proceedings of the 6th Mexican school on gravitation and mathematical physics.

1954.

Tests were conducted by Tomaschek in Shetland. The instrument used was a gravimeter, not a Foucault pendulum, and a nil result was obtained.

1961, 1965.

The next test using a Foucault pendulum was undertaken by Jeverdan in Romania. A positive result was recorded. However, a later test in Trieste gave a nil result.

1970

A widely-reported experiment was performed by Saxl and Allen at Harvard. The apparatus was a torsion pendulum designed so that the period changed with variation in gravity. The onset of a solar eclipse caused the period to increase from 29.57 to 29.58 seconds – a result that was above the normal level of noise and was, therefore, held to be significant. Curiously though, after the eclipse, the period did not immediately revert to its pre-eclipse value but remained briefly at its higher value.

1987 – 1990

A team at Imperial College investigated whether the frequencies of various atomic clocks could be affected by solar eclipses, and obtained 'positive' results.

1990

Kuusela conducted an experiment in Finland. He used a torsion balance, and obtained no significant results.

1991 – 1994

Kuusela joined forces with Savrov, van Ruymbeke, and Mena Jara, to test the effects of full and partial eclipses in Mexico and Brazil using both a torsion pendulum and a gravimeter. The results were 'ambiguous'.

1995

Misra and Rao obtained 'positive' results in India, using a gravimeter.

1997

A notable experiment, in the presence of a full eclipse, was conducted in the Mohe province in North-east China. The researchers were Wang and Yang who used a 'very sensitive' gravimeter. Their remarkable, positive, results are illustrated in Fig 2

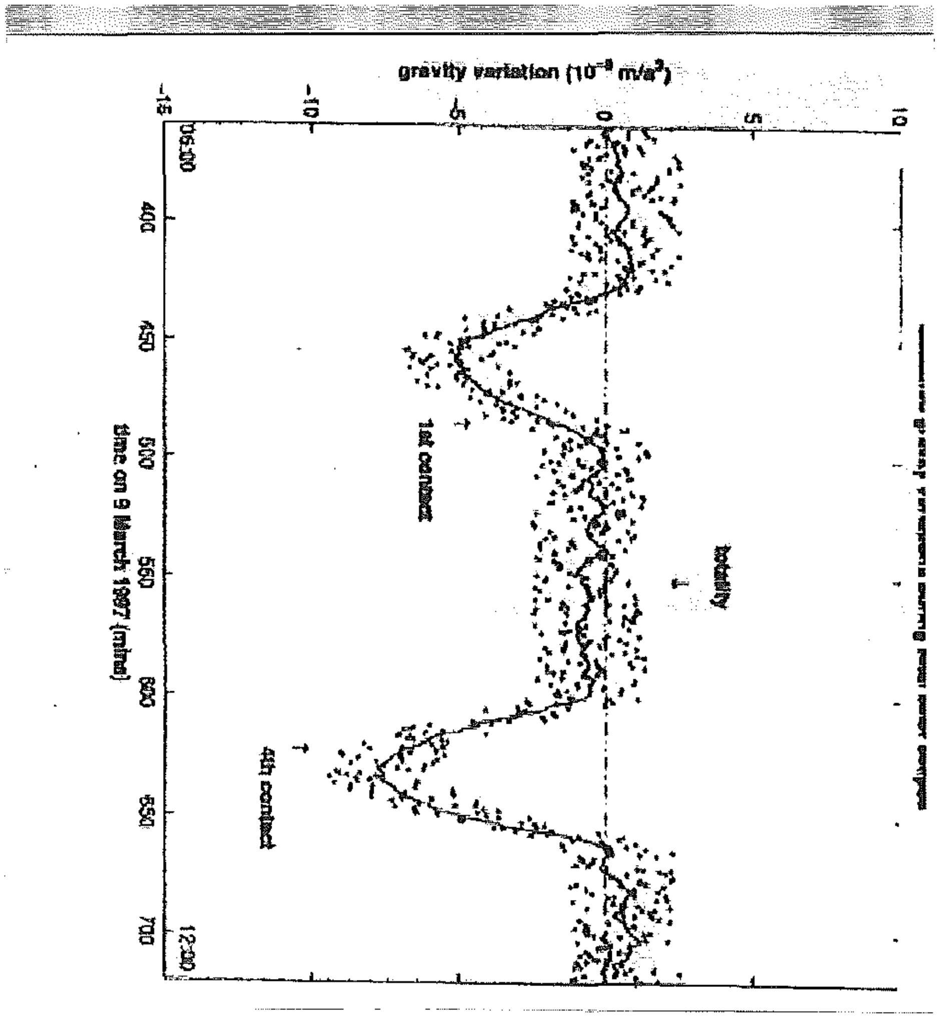


Fig 2 Graph showing gravimeter readings by Wang and Yang during solar eclipse.

2001

Another full eclipse was tested in Zambia by Wang and Tang using the same gravimeter. The results were 'positive'.

2002

The same researchers repeated the experiment in Australia with a further full eclipse. The results were again 'positive'

2004

Olenici used a Foucault Pendulum in Kuching, Borneo, and obtained results that were 'positive'.

2005

Tests by Goodley in Columbia using a Foucault Pendulum were 'inconclusive'.

2006

Kuusela used a magnetic gravimeter in Turkey and obtained results that were 'positive'.

In the light of this summary, it seems fair to conclude that:

a) gravity on Earth is in some way affected by a solar eclipse;
b) this phenomenon is not explained by present-day science; and
c) due perhaps to (b) above, the phenomenon is somewhat unpredictable, which might explain why some experiments produced results that are inconclusive or anomalous.

In accordance with the object of this paper, it is reasonable to assess whether the deductions listed in (1) can explain any of the gravitational anomalies summarised above.

In paper 'B' it was deduced that gravity arises from the extended, probabilistic, structures of particles with half-integral spin. To illustrate this point, although such a particle may have a mean position on the Sun, it will spend some of its time – admittedly not very much – here on Earth.

During a solar eclipse, the relative locations of the Sun, the Moon and Earth are as illustrated in Fig 3. When the particle referred to above is adjacent to the Moon, it will be attracted towards the Moon by gravity and so will its path

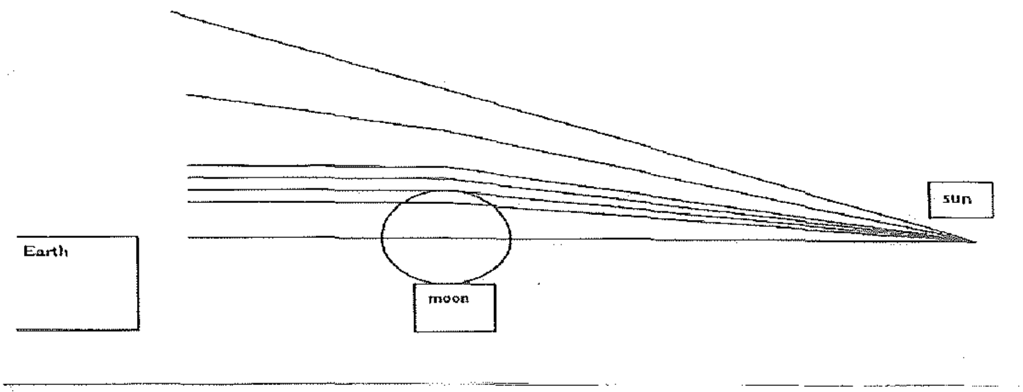


Fig 3 Diagram showing distortion of solar gravity by Moon causing a ring of enhanced solar gravity to be projected on to the surface of the Earth.

However, if the particle passes through the centre of the Moon, its path will not be deflected. (Measurement shows that the Moon does not shield the Earth from solar gravity).

In consequence, the Moon acts as a gravitational lens, focussing the Sun's gravity into a ring of enhanced solar gravity, which is then projected on to the surface of the Earth.

(A qualitative explanation, such as this, is easy to construct. It is more difficult to express such an effect in quantitative terms, principally because the properties of a 'quantum jelly', as created by a cloud of probabilistic particles, are as yet unclear. In particular, the existence of any restoring force that might determine the scale of the bending of particle paths has yet to be established. Hence, a qualitative explanation will have to suffice for the time-being.)

Immediately, such an explanation starts to fit the experimental facts. Fig 2 shows the results of an experiment involving a full eclipse of the sun. The twin peaks come from the gravimeter bisecting the predicted ring of enhanced solar gravity. Furthermore, these peaks show reductions in gravity as measured, precisely what is to be expected from the introduction of solar gravity which is opposite in sense to (and at the Earth's surface much smaller than) gravity arising from the Earth's mass.

Very few solar eclipses occur on the equator at noon. Consequently, projecting the ring of enhanced solar gravity on to the surface of the Earth will cause distortion of the ring, due both to variation in the local time of day and the latitude. Such a ring could become a quasi-ellipse or even -hyperbola.

Thus, predicting the precise areas where enhanced solar gravity from a full solar eclipse is to be observed is less than straightforward. (If the eclipse is partial, further complications are to be expected). Such distortion could well account for some of the reported difficulties in gaining 'positive' experimental results. In particular, the 'high level' that continued briefly after the eclipse as reported by Saxl and Allen might be explained by such distortion.

Returning to Fig 2, the first peak occurred at ca 7 am while the second occurred at ca 11 am. The first peak is somewhat smaller than the second - precisely what one would expect as the angle of incidence at 7 am would have been less, and the cosine effect correspondingly greater, than that when the sun was near its zenith at 11 am.

At first sight, the excursions in Foucault Pendulum azimuth illustrated in Fig 1 seem anomalous. Why have the twin peaks shown in Fig 2 now become single peaks? Fortunately, the explanation is not hard to find. For the Fig 1 experiments, the eclipse was partial, not full. Hence the ring of enhanced solar gravity would not have been bisected, the contact being closer to tangential. In such a case, a single but extended peak would have resulted – precisely what is shown in both graphs.

What remains unclear is why in both of the Fig 1 experiments the azimuth of the Foucault Pendulum was affected to such a large degree. However, it is known that the azimuth of such a pendulum is sensitive to horizontal forces and it is speculated that the horizontal component of the enhanced solar gravity may account in some measure for the scale of the observed effect.

It is known that NASA (1999) had plans to coordinate a world-wide series of experiments to take this subject further. It was intended that some experiments should employ a line of sensors across the track of an ellipse in order to counter the distortion of enhanced gravity location mentioned earlier. Furthermore, it was hoped to measure horizontal as well as vertical variations in local gravity.

The findings from such a programme should advance this intriguing subject and help to resolve present uncertainties. In the meantime, the theory (as described above) is tested further against other anomalies as follows.

2.2 Anomalies where gravity runs 'uphill'

This section relies heavily upon a paper 'Anomalies of gravitation in Italy and Poland' written by Grazyna Fosar and Franz Bludorf, and helpfully published on the Internet in English.

The first anomaly is located on the via dei Laghi by Lago Albano, some 30 kilometres South-east of Rome. This road, adjacent to the lake, runs uphill, but round or cylindrical objects such as empty bottles roll up it in a very puzzling manner.

The two authors, who are scientists, made a series of simple but telling observations that can be summarised thus:

- The length of the anomaly is some 200m.
- The anomaly ends before the crest of the hill.
- A bottle rolls uphill, but diagonally, off the centre of the road.
- The anomaly pulsates; sometimes it is present and sometimes not.
- An electric field test and a Geiger Counter test yielded nil results.
- Although the road certainly ran uphill, a test with a spirit level indicated it ran 5^0 downhill.
- A simple pendulum with a natural period of 0.9629s as measured at Berlin had an increase of period at the anomaly of 0.015s, a change of 3.2% (both periods measured over 100 swings). Normal maximum deviation in Earth's gravity is one part in 10^4.

Thus, the gravitational anomaly has both vertical and horizontal components (as shown by the pendulum and spirit level tests). The pulsating nature suggests the cause of the anomaly is dynamic, rather than static, in nature.

The second anomaly is located at the health resort of Karpacz Gorny (Brückenberg) in the mountains of South-west Poland. The observations of the two authors are summarised thus:

- Once more, bottles roll uphill.
- A large Volvo plus passengers, luggage, and a heavy dog coasts uphill for 400m.
- GPS readings at start and finish confirm the gradient is uphill and thereby refute any assertions concerning 'optical illusions'.
- Pendulum swings show vertical gravitation decreased by 4%.
- No pulsation.

Further enquiries revealed that this anomaly is widely known in Poland and is a national tourist attraction. Also, recent satellite measurements indicate large deposits of very hot water at a depth of 2,000 m.

As Lago Albano marks the site of an extinct volcano, the two authors propose that the presence of water and volcanic activity are significant common factors, and may in some way explain the observed anomalies.

Can the theory proposed in (2.1) provide further explanation? Maybe so. Water is much less massive than rock, so substantial nearby quantities of subterranean water might account for a local decrease in Earth-caused gravity, such as was observed at both locations. But a decrease in vertical gravity on its own is insufficient because on its own it would be unlikely to cause bottles to roll uphill.

How, then, might a horizontal component of the gravity vector arise? At a vertical boundary between water and rock, gravity will be 'bent' or distorted in a manner similar to that described in (2.1) above. The sense of the distortion will be opposite to that caused by the mass of the eclipsing Moon, but a horizontal component of gravity will still result. Given suitable geometry for the water/rock interface, such a horizontal component might be sufficient to roll bottles uphill.

Whether the presence of volcanism is relevant to such distortion is unknown. On the other hand, volcanic activity might well, in conjunction with the fluidity of water, account for the pulsating feature as noted at Lago Albano.

2.3 Gravitational anomaly related to Pioneer 10 and Pioneer 11 space probes

During 1972/3, Pioneer 10 and Pioneer 11 were launched in opposite directions to examine deep space. Pioneer 10 has ceased to function, although its position can still be determined. Pioneer still functions well and is now 13 billion km from Earth.

The distances from Earth of both probes are less than those predicted (using the law of gravity) by some 400,000 km – roughly the distance from Earth to the Moon. Other spacecraft have exhibited similar errors. Although the distances involved were much smaller, the scale of error was comparable to that observed with Pioneer 10 and 11.

Prolonged study at NASA has not found any system errors to explain such a discrepancy, although the existence of such errors cannot be entirely discounted. Indeed, new sources of possible system error are still proposed occasionally by third parties.

Therefore, the possibility that the cause may lie in some subtle error in the present law of gravity has to be considered. It has been calculated that the positional error 'is equivalent' to adding a constant gravitational acceleration towards the Sun of

$$8.74 \pm 1.33 \times 10^{-10} \ {}^{m}\!/_{\sec^2}$$

though it is not stated specifically that such acceleration has been proven to be constant.

As the two probes fly in opposite directions and are short of calculated range by similar amounts, it is unlikely that the errors are caused by some factor beyond the Solar System. The assumption has to be that it is Sun-based gravity that is in some way responsible.

The conventional law of gravity is a simple inverse square law and does not allow for a term of gravity (no matter how small) that is independent of the distance of separation.

The fine adjustments to the law of gravity referred to in (1) above all decrease with distance, as one would expect. In fact, the thought that the Sun causes a constant gravitational acceleration on every particle throughout the entire universe irrespective of distance seems to be an absurdity; not least because such an acceleration would arise from every other star, no matter how distant. Unless the universe were entirely homogeneous (which it is not – it's somewhat lumpy instead) there would be a net acceleration in some direction and both Pioneer probes would have identical errors in the same direction, not in opposite directions.

So we need to know more. Is it possible that the positional errors of the pioneer probes are caused by a distance-dependent gravitational term and, if so, what is the appropriate law that governs such a term? If, for example, it were determined that the acceleration in question varies, say, reciprocally with distance, progress might be possible.

3 Is there an ether?

3.1 The need for an ether

One implication from the deductions listed in (1) is that action at a distance can be accounted for by interaction between the extended, probabilistic, structures of various sub-atomic particles. If this is so, there is no need for electric, magnetic, or gravitational fields. The concept of a field becomes no more than a convenient scientific fiction that, in most circumstances, can be used to facilitate calculation of the forces of electrostatics, magnetism, and gravity.

This creates a problem in that the propagation of light has been explained by Maxwell's equations, and such equations rely upon the existence of fields. If fields do not exist, how is light to be propagated? The concept of an 'ether' may have to be revived.

3.2 Does the ether consist of virtual particles?

Conventional science already recognises the existence of 'quantum soup' and 'zero point energy', whereby virtual particles can leap in and out of existence, even in a vacuum. Such a concept is not far removed from the idea that some particles have extended, probabilistic structures so that, for example, an electron can momentarily (i.e. virtually) exist at a location far removed from is mean position, and then disappear to reappear somewhere else. We explore the idea that such virtual particles form the ether we seek.

3.3 The datum for angular velocity

The universe has many particles, so space everywhere is well-populated with virtual particles, as described above. In addition to enabling action at a distance, what other functions do they serve?

It is well-known that linear velocity is relative but angular velocity is absolute. A passenger in a windowless space ship cannot do an experiment within the limits of the ship to measure speed, but can always do a simple test with an accelerometer to measure rate of rotation. Experiments confirm that very distant, 'stationary' stars form the most accurate datum from which angular velocity can be measured – i.e. what exists at the periphery of our universe defines what is zero rotation here on Earth.

The implication of this is that, in some way, the distant elements of the universe are present here on Earth to enable, by means as yet undetermined, an instant sensing of angular velocity to be made. This implication aligns closely with the concept of space being populated with virtual versions of distant particles as described above.

Furthermore, as a simple calculation using the probability information given in (1) will show, it is the most distant particles that provide the greatest density of virtual particles. Such a finding aligns well with the fact that it is the most distant parts of the universe that act as the true datum from which angular velocity can be measured.

Hence it seems reasonable to ask the question, 'Does the Earthly presence of virtual particles, whose mean positions lie at the extremities of the universe, provide the datum for the absolute measurement of angular velocity?

If so, we have the important implication is that such sensing by virtual particles would have to be sensitive to angular velocity but insensitive to linear velocity.

3.4 Gravitational Lensing

A hint as to what is going on is provided by the phenomenon of 'Gravitational Lensing'. As is well-known in astronomy, if a distant star 'A' is obscured by a nearer star 'B', it is sometimes possible to see A as a halo surrounding B. The light from A is bent by the mass of B, according (at present) to the general theory of relativity, and forms the ring of light – the halo – mentioned above.

Such an effect is analogous, but not identical, to the theory proposed in (2.1) whereby the Sun's gravity is deflected during a solar eclipse by the Moon to form a ring of enhanced solar gravity on the surface of the Earth.

3.5 Are such effects related?

This poses the question, 'Is it possible that these are not two separate effects but are one and the same?' In other words, could distortion (caused by the mass of the Moon or Star B) of the cloud of virtual particles in space account (perhaps through two different mechanisms) for both the bending of light and the bending of gravity? If such a suggestion seems plausible, it is but a short step to propose that the cloud of virtual particles is the medium (i.e. the ether) by which light and other forms of electro-magnetic radiation can propagate.

3.6 The ether wind problem

Before such a proposal can be taken seriously there is a significant objection to be overcome. Earlier beliefs in the existence of an ether were destroyed by the celebrated Michelson/Morley experiment. It was argued that the motion of the Earth should create an 'Ether Wind' leading to measurable differences between the paths of light along and across the Earth's path. Famously, the experiment showed no such differences, and hence no ether wind and no ether.

Such a deduction failed to take into account the possibility that the ether might have special properties by being insensitive to linear velocity and responding only to other variables such as angular velocity. It is doubtful whether, in days gone by, such an unusual ether could have been contemplated. However, the arguments above plus the deductions listed in (1) now require such a possibility to be considered.

3.7 Can linear velocity be measured by observing a cloud of virtual particles?

The deduction (see 2.1) that the universe of any observer is defined by the limit of that observer's perception implies, given isotropy, that any observer is always at the centre of his/her individual universe. Hence, the observed density of virtual particles is unlikely to vary with motion, so linear velocity cannot be determined by measuring rate of change of particle density.

Furthermore, virtual particles flash very quickly in and out of existence, arguably much too quickly in a digital universe to measure linear velocity using any one virtual particle as a datum. Hence, it seems likely that an ether consisting a cloud of virtual particles would be such that measuring linear velocity relative to the ether would not be possible.

3.8 Can angular velocity be measured by observing a cloud of virtual particles?

The answer to this question is ,'Yes'. To the rotating observer in a cloud of virtual particles, the particles appear to possess angular velocity. Such particles, although virtual, possess half-integral spin, and thus - see deductions in (1) – possess charge. The observer is thus enclosed by orbiting charges, the existence of which could be detected immediately by reading a compass.

(Of course, the mean position of such virtual particles would be very remote. However, this would not affect the response of the compass. Returning to conventional science, the strength of a magnetic field inside a conducting coil does not vary with the coil's radius, even if that radius happens to be that of the universe.

However, a similar argument does not invalidate the conclusion given in (3.7) above. Linear velocity in such circumstances could be held to produce a linear electric current, which would also produce a magnetic field. However, in this case, the field strength does vary according to the separation between current and observer. As such separation would also, in this case, approximate to the radius of the universe, the ability to measure linear velocity by such means seems improbable.)

3.9 The relevance of measuring rotation

As a final indication that all this may be on the right lines, recall the comments (3.3) above on measuring angular velocity. We know that absolute linear velocity cannot be measured while absolute angular velocity can. This, too, implies that angular velocity, but not linear velocity, can be sensed by reference to a cloud of virtual particles.

4 Conclusions

Some of the gravitational anomalies examined are consistent with the earlier deductions that arise from the concept of a digital universe and, therefore, support of such a concept. However, the gravitational anomaly related to Pioneer 10 and 11 is not explained by such deductions, although access to further experimental data might enable such a conclusion to be reversed.

The idea that the presence of a cloud of virtual particles might function as an ether for the transmission of light is appealing but as yet unproven.

Further Thoughts on a New Type of Jet Engine

Grenville J. Croll
mail@grenvillecroll.com

We change the scale of our original design, discuss existing related technology and introduce the notion of cavitation rather than pressurisation. We find a way of separating hot exhaust plasma from the fuel

1 Introduction

Our basic engineering concept [Croll, 2009] is to force Deuterated Helium II through a small nozzle. The sole assumption behind the engineering concept is that the Gödelian undecidability implicit in the monoidal prespace encoding of reality, per Kondo's model [Croll, 2006], provides an unprovable opportunity for a D-D fusion process to take place.

2 Engineering Scale

Should a D-D fusion process take place, energy output would be extremely high. We are therefore motivated to reduce the scale of our nozzle in order to increase the surface area to volume ratio of any cavity that contains such a reaction and to reduce the volume of the fuel in the cavity. Scale reductions have a further benefit in that so called quantum effects are more pronounced at reduced scales. We seek the Gödelian undecideability implicit in small scale physical processes.

3 Existing Technology

By coincidence there already exists a class of device which, save the fusion burn and the low temperature, are manufactured in some volume. These devices presently comprise an array of several hundred closely spaced holes and operate at frequencies of several kilohertz. The cavity array is usually of the order of one square centimetre. Each cavity operates independently under either piezoelectric or thermal control. Each cavity processes about one picolitre of low viscocity fluid (with a specific gravity of approximately 1.0) per cycle. The reliability of these devices is very high with a MTBF of in excess of 1000 hours. One of these devices was used to print copies of the various versions of this paper. There are hundreds of patents on these devices and dozens of manufacturers.

Regarding our own Jet Engine design, we can now envisage a cavity array rather than a single nozzle as originally proposed. Given knowledge of the materials likely to be used in the manufacture of such a device, we can back calculate the dimensions and properties of a cavity array that gives us a flight capable Jet Engine. For example, a one hundred square centimetre cavity array comprising one million nozzles ejecting one nano litre of 1.0sg fluid at one kilohertz provides a fluid throughput of one litre per second.

The Nozzle Array – the Blast Wall – can now be separated from the Fuel Supply by a Non Conducting Part (NCP) containing as many tubules as the Blast Wall contains nozzles. The superfluidity of Helium 2 suggests that the NCP can be quite long without compromising fuel flow as long as it remains cold.

The pumping system is envisaged to be a piezoelectric array integrated into the forward facing part of the NCP. If fusion is assisted by pulsed operation and low level laser ignition as previously proposed, the nozzles and pumps can be addressed and timed individually, to control fuel and coolant flow, the thermal profile of any burn across the Blast Wall and the sound qualities (Bach, Mozart, Pink Floyd etc) of the operating engine. The Blast Wall may be surrounded by further coolant flow such as steam.

4 Cavitation versus pressurisation

The effects of Fluid Cavitation such as is seen around ships propellers is well known. Less well known and more controversial is the allegation that cavitation can be used as a non-conventional means of energy creation [Griggs, 1987]. The effect has been studied recently under controlled conditions, however it is not a popular thread of research due to the difficulties of gaining funding and the likelihood of the collapse of one's conventional scientific reputation. Cavitation is therefore of interest for two reasons. Firstly we are seeking any means for fusion to take place and secondly we note the likelihood that pressurising Helium II will warm it up, potentially above the lambda temperature, an effect that is not desired, due to the loss of superfluidity. Cavitation, acting in the reverse sense of pressurisation may cause the fuel and dopants to spontaneously change form in the manner sought.

5 Fusion Turbine

The concept of the Blast Wall at the size envisaged suggests that we have conceived a Fusion Turbine. The Blast Wall, coolant and containment correspond to the combustion container of a conventional Jet Engine. The requirement to pump high volumes of coolant around a Blast Wall and its containment chamber (or several, arranged radially) suggests a similarity to conventional jet design where rearward exhaust turbines feed forward coolant (rather than oxidiser) blades.

References

Croll G.J. (2006), The Natural Philosophy of Kazuo Kondo. Proc ANPA, Cambridge, arXiv:0712.0641v1

Croll G.J. (2009), A New Type of Jet Engine: Some A Priori Engineering Design Principles. Proc ANPA, Cambridge

Griggs, J.L., (1987) http://www.rexresearch.com/griggs/griggs.htm 10:00 10/01/11

PROGRESS IN CONSTRAINTS THEORY (CT)

Anthony Michael Deakin

1a Castle Rd, Rowlands Castle, Hants. PO9 6AP
amdeak.deakin@btinternet.com

PREAMBLE

We find that the Schrödinger quantization of a rudimentary classical archetype produces operator equations that express most of the rules of classical mechanics. These rules include the Riemannian connectivity of physical space; but this seems to emerge only at the macrophysical scale. The field equations are fourth order. But they are satisfied by the conventional second order solutions both as regards gravity and EM; extra physics is therefore to be expected. The Maxwell equations, which are deducible from the Feynman-Dyson quantum mechanical argument, are hardly effected. Gravitation is governed by the fourth order Kilmister equation; in conjunction with the Einstein equation, this dictates the allowed distributions of matter-energy-momentum and, perhaps, offers an explanation of the Dark Energy effect. Preliminary machine calculations suggest that the Kilmister equation has acceptable cosmological solutions.

Note: Each of the following sections begins with an Abstract in Italics.

LECTURE 1- IDEAS AND SOME RESULTS

1.1 The Classical Archetype

Begin with an abstract space $\mathscr{P}$ which is real, infinite and continuous. This is supposed of finite dimension n_d and unknown connectivity. A finite number n_p of special points are called particles. The n_d-tuple of real numbers that constitute a special point are called its coordinates. Denote the aggregate of coordinates by the $n_p n_d \times 1$ array $\underline{q}$; and suppose that each element of $\underline{q}$ is a continuous differentiable function of a single real, continuous variable $t \in [-\infty, \infty]$ called time. Suppose also that associated with $\mathscr{P}$ is a continuous, differentiable function $\theta(\underline{q})$ or a number of such functions. This is our classical archetype; but we have yet to gather evidence to show that $\mathscr{P}$ can be used to model physical space.

$\mathscr{P}$ consists of all possible Points which are either General or Special. A General Point consists of a unique, ordered n_d-tuple of unrestricted real numbers

$$\{q^1, q^2, \ldots, q^{n_d}\}; \quad q^\beta \in [-\infty, \infty]; \quad \beta = 1, 2, \ldots, n_d$$

A Special Point consists of a unique, ordered n_d-tuple of unrestricted real numbers called Coordinates. There are n_p special points.

$$\underline{q} \equiv \{q_\alpha^\beta\}; \quad q_\alpha^\beta \in [-\infty, \infty]; \quad \alpha = 1, 2, \ldots, n_p; \quad \beta = 1, 2, \ldots, n_d$$

We will not use this last notation $\{q_\alpha^\beta\}$ again.

We call this a Classical Archetype because a prime feature of classical physics is that all measurables are assumed continuous; and, when some are functions of others, those functions are assumed differentiable. Another prime feature of classical physics is that mater is assumed to be an agglomeration of structureless point particles.

1.2 Hierarchy Of Identities

Write an hierarchy of identities for $\dot{\theta}, \ddot{\theta}, \ldots;\ \ \dot{\theta} \equiv d\theta / dt$.

$$\dot{\theta} = \dot{q}^j \theta_{,j};\quad \theta_{,j} \equiv \frac{\partial \theta}{\partial q^j};\quad \dot{\theta} \equiv \frac{d\theta}{dt}$$

$$\ddot{\theta} = \ddot{q}^j \theta_{,j} + \dot{q}^j \dot{q}^k \theta_{,jk};\quad \theta_{,jk} \equiv \frac{\partial^2 \theta}{\partial q^j \partial q^k}$$

$$\dddot{\theta} = \dddot{q}^j \theta_{,j} + 3\ddot{q}^j \dot{q}^k \theta_{,jk} + \dot{q}^j \dot{q}^k \dot{q}^l \theta_{,jkl};$$

$$\ddddot{\theta} = \ddddot{q}^i \theta_{,i} + (4\dddot{q}^i \dot{q}^j + 3\ddot{q}^i \ddot{q}^j)\theta_{,ij} + 6\ddot{q}^i \dot{q}^j \dot{q}^k \theta_{,ijk} + \dot{q}^i \dot{q}^j \dot{q}^k \dot{q}^l \theta_{,ijkl};$$

$$i, j, k, l = 1, 2, \ldots n_c \equiv n_d n_p$$

etc.

The Einstein summation convention is in force; and, unless otherwise stated, all indices lie in the range $[1, n_c]$.

1.3a The Quantizations-Notations

We quantize these identities, as if they are equations of motion with t *as time, using the rules of Schrödinger. The usual choice of a mathematical* *model, from say classical physics, is thereby avoided. The results are no longer identities; they are constraints. The constraints are operator equations in the coordinate operators* $\underline{Q}$, *the corresponding conjugate operators* $\underline{P}$, *the operator* $\Theta(\underline{Q})$ *which represents the function* $\theta(\underline{q})$ *and the QM Hamiltonian* $H(\underline{P},\underline{Q})$. *The rules of quantization are a generalisation of those used by Schrödinger in his famous hydrogen model. The constraints are taken not as inconsistencies but rather as laws of nature.*

We use the following notations:

$$a \to A; \quad \dot{a} \equiv \frac{da}{dt} \to \dot{A} \equiv \overset{1}{A}; \quad \frac{d^n a}{dt^n} \to \overset{n}{A}; \quad n = 1,2,....$$

where '$\to$' means 'is represented by the operator'.

$$[A,B] \equiv AB - BA; \quad \lfloor A,B \rfloor \equiv \frac{i}{\hbar}[A,B]$$
$$[A,B,C] \equiv [A,[B,C]]; \quad [A,B,C,D] \equiv [A,[B,C,D]];$$
$$\lfloor A,B,C \rfloor \equiv \lfloor A,\lfloor B,C \rfloor \rfloor; \quad \lfloor A,B,C,D \rfloor \equiv \lfloor A,\lfloor B,C,D \rfloor \rfloor$$

$$p_j \to P_j; \quad q^k \to Q^k; \quad A_{,j} \equiv \lfloor P_j, A \rfloor; \quad A^{:k} \equiv \lfloor A, Q^k \rfloor$$

$$\{A_1, A_2, ..., A_n\} \equiv \tfrac{1}{n!} \sum_{perm} A_1 A_2 ... A_n$$

The commas on the LHS of this last formula are inserted, if need be, only for clarity. The order of the arguments in $\{.\}$ is immaterial. Notice that if an element inside any of the brackets $[.], \lfloor . \rfloor, \{.\}$ is null then the bracket is null.

We combine these notations with the Einstein summation convention in a manner illustrated by the following examples:

$$\left[A^u, B_u^{vw}\right] \equiv \sum_{u=1}^{n_c}\left(A^u B_u^{vw} - B_u^{vw} A^u\right); \quad \left\lfloor A^u, B_u^{vw}\right\rfloor \equiv \frac{i}{\hbar}\left[A^u, B_u^{vw}\right]$$

$$\left\{A_x^{uvw}, A_{uv}^{yz}, A\right\} \equiv \frac{1}{3!}\sum_{u,v}^{n_c}\left(\sum_{perm} A_x^{uvw} A_{uv}^{yz} A\right)$$

where, in this last example and for simplicity, A happens to be an invariant.

1.3b The Quantizations- Rules (Substitution)

$a \to A; \quad b \to B$

where a,b are real, scalar observables and A,B are their representative Hermitian symmetric operators. Then for multiples, integer powers, sums (weighted by c-numbers α,β) and products

$$\alpha a^m \to \alpha A^m; \quad \alpha a + \beta b \to \alpha A + \beta B; \quad ab \to (AB + BA)/2; \quad m = 1,2,....$$

whether or not A and B commute. From these formulae Kauffman deduces

$$\sum_{j=1}^{n} \alpha_j a_j \to \sum_{j=1}^{n} \alpha_j A_j; \quad a_1 a_2 .... a_n \to \tfrac{1}{n!}\sum_{perm} A_1 A_2 .... A_n \equiv \left\{A_1, A_2, ..., A_n\right\} \text{ [Kauffman]}$$

whether or not the A_j mutually commute.

In the Q-diagonal Schrödinger representation (I is the unit operator)

$q^j \to Q^j \equiv q^j I; \quad a(\underline{q}) \to A(\underline{Q}) \equiv a(\underline{q})I; \quad -\infty < q^j < \infty;$ The $\underline{q}$ are flat

$$p_k \to P_k \equiv -i\hbar \frac{\partial}{\partial q^k}; \quad b(\underline{p}) \to B(\underline{P}) \equiv b(\underline{P})$$

In the P-diagonal Schrödinger representation

$$p^j \to P^j \equiv p^j I; \quad b(p) \to B(\underline{P}) \equiv b(\underline{p})I; \quad -\infty < p_j < \infty$$

$$q_k \to Q_k \equiv i\hbar \frac{\partial}{\partial p^k}; \quad a(\underline{q}) \to A(\underline{Q}) \equiv a(\underline{Q}); \; \mathscr{P} \text{ is flat}$$

$$P_j P_k = P_k P_j; \quad Q^j Q^k = Q^k Q^j; \quad Q^j P_k - P_k Q^j = i\hbar \delta_k^j I; \; j,k = 1,2,3; \text{ Identities}$$

We assume, in addition, that these relations hold for all the coordinates of all the particles in $\mathscr{P}$. That is $j,k = 1,2,....n_c$

1.3c The Quantizations- Rules (Time)

In the Schrödinger representations the operators are independent of t and the 'state' function depends on t. Rate is treated according to the rule

$$\dot{a} \to \dot{A} \equiv \lfloor H, A \rfloor$$

This rule derives from the invariant (with respect to change of basis) criterion

$$\langle \psi(\underline{q},t) | \dot{A} | \psi(\underline{q},t) \rangle = \frac{d \langle \psi(\underline{q},t) | A | \psi(\underline{q},t) \rangle}{dt}; \quad \psi(\underline{q},t) = \exp(-iHt/h)\psi(\underline{q},0)$$

where $\psi(\underline{q},t)$ is the 'state' function in the Hilbert space at time t after a measurement at $t = 0$. In the Heisenberg representations the operators depend on t and the state vectors are independent of t. The connection between a Schrödinger operator A and the corresponding Heisenberg operator $A(t)$ is

$$A(t) = U^{+}(t)AU(t); \quad U(t) \equiv \exp(-iHt/\hbar); \quad \text{'+' denotes Hermitian transpose}$$
$$\dot{A}(t) = \lfloor H, A(t) \rfloor = U^{+}(t)\lfloor H, A \rfloor U(t); \quad H(t) = H$$
$$U^{+}(t)U(t) = U(t)U^{+}(t) = I; \quad I(t) = I$$

1.3d The Quantizations- Rules (Derivatives And Derivations)

We define

$$\frac{\partial a}{\partial q^{j}} \equiv a_{,j} \rightarrow A_{,j} \equiv \lfloor P_{j}, A \rfloor; \quad \frac{\partial a}{\partial p_{k}} \equiv a^{;k} \rightarrow A^{;k} \equiv \lfloor A, Q^{k} \rfloor$$

These definitions are justified by the fact that the rules for calculating $A_{,j}$ and $A^{;k}$ can be summarized by the statements

$$A_{,j} = \frac{\partial A}{\partial Q^{j}}; \quad A^{;k} = \frac{\partial A}{\partial P_{k}}; \quad A \text{ a multinomial in the } \underline{P} \text{ and the } \underline{Q}$$

as if $A, A_{,j}, A^{;k}, Q^{j}, P_{k}$ are scalars providing that the order of non-commuting operators is preserved. If A is pure (either in the $\underline{P}$ or the $\underline{Q}$) then, it can be proved, that these results apply even when A is not a multinomial. It is necessary for the scalar partial derivatives $\partial a(\underline{p})/\partial p_{k}$ or $\partial a(\underline{q})/\partial q^{j}$, where $a \rightarrow A$, to exist. Kauffman calls commutators like the above derivations. Notice that

$$A_{,j,k} = A_{,k,j}; \quad A^{;j;k} = A^{;k;j}$$

are identities for any operator A. In particular, if A is pure in the $\underline{P}$ then,

$$A^{:j:k} = A^{:k:j} \Rightarrow a^{:jk} \equiv \frac{\partial^2 a}{\partial p_j \partial p_k} = \frac{\partial^2 a}{\partial p_k \partial p_j} \equiv a^{:kj}$$

using the P-diagonal Schrödinger representation. If A is pure in the $\underline{Q}$ then,

$$A_{,j,k} = A_{,k,j} \Rightarrow a_{,jk} \equiv \frac{\partial^2 a}{\partial q_j \partial q_k} = \frac{\partial^2 a}{\partial q_k \partial q_j} \equiv a_{,kj}$$

using the Q-diagonal Schrödinger representation.

1.4 The Quantizations- Results (The Constraints)

The constraints limit the forms permitted for Θ and H. They are:

$$Z_1 \equiv \{H^{:j}, \Theta_{,j}\} = \lfloor H, \Theta \rfloor; \quad Z_2 \equiv \{\lfloor H, H^{:j} \rfloor, \Theta_{,j}\} + \{H^{:j}, H^{:k}, \Theta_{,j,k}\} = \lfloor H, H, \Theta \rfloor$$

$$Z_3 \equiv \{\lfloor H, H, H^{:j} \rfloor, \Theta_{,j}\} + 3\{\lfloor H, H^{:j} \rfloor, H^{:k}, \Theta_{,j,k}\} + \{H^{:j}, H^{:k}, H^{:l}, \Theta_{,j,k,l}\}$$
$$= \lfloor H, H, H, \Theta \rfloor$$

$$Z_4 \equiv \{\lfloor H, H, H, H^{:i} \rfloor, \Theta_{,i}\} + 4\{\lfloor H, H, H^{:i} \rfloor, H^{:j}, \Theta_{,i,j}\} + 3\{\lfloor H, H^{:i} \rfloor, \lfloor H, H^{:j} \rfloor, \Theta_{,i,j}\}$$
$$+ 6\{\lfloor H, H^{:i} \rfloor, H^{:j}, H^{:k}, \Theta_{,i,j,k}\} + \{H^{:i}, H^{:j}, H^{:k}, H^{:l}, \Theta_{,i,j,k,l}\}$$
$$= \lfloor H, H, H, H, \Theta \rfloor$$

etc. In particular, the operators (expressed in the Q-diagonal Schrödinger representation)

$$\Theta(\underline{Q}) \equiv \theta(\underline{q})I; \quad \Theta_{,j} = \theta_{,j}I; \quad \Theta_{,j,k} = \theta_{,jk}I; \quad \Theta_{,j,k,l} = \theta_{,jkl}I \quad \text{etc.}$$

are pure in the $\underline{Q}$; the order of the suffices is immaterial. The Z_n are defined as the LHSs of the quantizations written in the above form; they are used below.

1.5 Hamilton's Equations

As a consequence of the Schrödinger definitions the $\underline{P}$, *the* $\underline{Q}$ *and* $H(\underline{P},\underline{Q})$ *satisfy an operator form of Hamilton's equations. If we reverse quantize these we get a scalar form of Hamilton's equations; but according to the definitions the* $\underline{q}$ *are special; they are flat. Consequently, if* $\mathscr{P}$ *has (say) Riemannian connectivity, then it must be flat.*

By inspection of the rules we see that

$$\dot{P}_j = -H_{,j}; \quad \dot{Q}^k = H^{;k}$$

These equations are the quantum analogue, in a flat space, of the classical Hamilton's equations

$$\dot{p}_j = -\frac{\partial H}{\partial q^j}; \quad \dot{q}^k = \frac{\partial H}{\partial p_k}; \text{ Hamilton's Equations}$$

Indeed these last can be regarded as the reverse quantization of the operator equations. But keep in mind that the axioms of quantization require that $\mathscr{P}$ is flat. So the operator Hamilton's equations are restricted to flat coordinates $\underline{Q}$ that are related to their conjugates $\underline{P}$. Their classical counterparts $(\underline{p},\underline{q})$ are allowed to be curvilinear and independent.

1.6 The Meaning Of The Constraints

We show that each constraint corresponds to a term in the power series expansions of $\theta(t)$, *the Heisenberg operator* $\Theta(t)$ *and its expectation* $\langle\Theta(t)\rangle$. *So it is desirable that the constraints are all satisfied up to as high a level as possible.*

The scalar function $\theta(\underline{q})$ can, in principle, be expanded as a function of t. The resulting series can be quantized in terms of operators in the Schrödinger representation

$$\theta(t) = \theta_0 + \dot{\theta}_0 t + \ddot{\theta}_0 t^2 / 2 + \ldots. \rightarrow \Theta(t) = \Theta + \dot{\Theta} t + \ddot{\Theta} t^2 / 2 + \ldots.;$$
$$\overset{s}{\theta}_0 \equiv \left.\frac{d^s\theta}{dt^s}\right|_{t=0}; \quad s = 0,1,2,\ldots.$$

The Schrödinger operators $\dot{\Theta}, \ddot{\Theta}, \ldots.$ are generated according to the rule

$$\left.\frac{d^s\theta}{dt^s}\right|_{t=0} \rightarrow \overset{s}{\Theta} \equiv \left[H, \overset{s-1}{\Theta} \right]; \quad \Theta \equiv \overset{0}{\Theta}; \quad s = 1,2,\ldots$$

and so

$$\theta(t) \rightarrow \Theta(t) = \exp(iHt/\hbar)\Theta\exp(-iHt/\hbar); \text{ Heisenberg Representation}$$

Write the constraints in the form

$$Z_s = \overset{s}{\Theta}; \quad s = 1,2,\ldots \Rightarrow \Theta(t) = \Theta + \sum_{j=1}^{\infty} \frac{Z_j t^j}{j!} \Rightarrow \langle\Theta(t)\rangle = \langle t|\Theta|t\rangle = \langle\Theta\rangle + \sum_{j=1}^{\infty} \frac{\langle Z_j\rangle t^j}{j!}$$

Failure of the n^{th} constraint means the failure of the n^{th} differential identity and a corresponding inability to predict $\theta(t), \Theta(t)$ and $\langle\Theta(t)\rangle$.

1.7 The First Constraint

If $\theta(\underline{q})$ is arbitrary the first constraint requires H to be quadratic in the $\underline{P}$ (Newtonian physics). With the additional assumption, that coordinate rates (with respect to t) are continuous, the coefficients in $H(\underline{P},\underline{Q})$ are restricted; they are either constants or functions of the coordinates.

The first constraint can be written in various forms:

$$\overset{1}{Z} = \overset{1}{\Theta} \equiv \dot{\Theta} \Rightarrow \left\{H^{;j}, \Theta_{,j}\right\} = \lfloor H, \Theta \rfloor$$

$$\Rightarrow -\frac{1}{2\hbar^2}\left((HQ^k - Q^k H)(P_j\Theta - \Theta P_j) + (P_j\Theta - \Theta P_j)(HQ^k - Q^k H)\right) = \frac{i}{\hbar}(H\Theta - \Theta H)$$

It can be shown that, if the function $\theta(\underline{q})$ is arbitrary then, the most general form permitted for H is quadratic in the $\underline{P}$

$$H \equiv \left\{C^{uv}, P_u, P_v\right\} + \left\{E^j, P_j\right\} + U; \quad C^{uv} = C^{vu}; \quad u, v, j = 1,2,\ldots n_c$$

The Hermitian operators C^{uv}, E^j, U do not depend on the $\underline{P}$ and commute with the $\underline{Q}$. The rate operator of the k^{th} coordinate is

$$\dot{Q}^k \equiv \lfloor H, Q^k \rfloor = 2\left\{C^{uk}, P_u\right\} + E^k$$

It follows that, if any of the C^{uk} or the E^k are non-scalar matrices then, the spectrum of $\dot{Q}^k$ is not continuous; thus if, we require the spectrum of $\dot{Q}^k$ to be continuous then, the C^{uv}, E^j must be scalars. A form, consistent with this assumption, is

$$H \equiv \mathrm{K}\{G^{uv}, P_u P_v\} + \{F^j, P_j\} + V; \quad G^{uv} = G^{vu}; \quad u, v, j = 1, 2, \ldots n_c$$

where K is a conventional scalar (with physical dimensions $mass^{-1}$) and, in the Q-diagonal Schrödinger representation, the coefficients G^{uv}, F^j, V are either scalar constants or scalar functions of the $\underline{q}$

$$G^{uv}(\underline{Q}) = g^{uv}(\underline{q}); \quad F^j(\underline{Q}) = f^j(\underline{q}); \quad V(\underline{Q}) = v(\underline{q}); \quad Q^j = q^j I$$

1.8 The First Two Constraints (Theta Equation)

The operator equation that is satisfied when the first two constraints hold is called the Operator Theta Equation. In the Schrödinger representation it becomes a fourth order PDE called the Scalar Theta Equation
$g^{vj}(g^{uk}\theta_{,jku})_{,v} = 0;\ j,k,u,v = 1,2,\ldots n_c \equiv n_d n_p$. *In this PDE* $\theta(\underline{q})$ *is the dependent variable; and the* $g^{vj}(\underline{q})$ *are the coefficients of the quadratic terms in* $H(\underline{P},\underline{Q})$.

If $\theta(\underline{q})$ is arbitrary and the first two constraints

$$\{H^{\cdot j}, \Theta_{,j}\} = \lfloor H, \Theta \rfloor; \quad \{\lfloor H, H^{\cdot j} \rfloor, \Theta_{,j}\} + \{H^{\cdot j}, H^{\cdot k}, \Theta_{,j,k}\} = \lfloor H, H, \Theta \rfloor$$

hold then it can be shown that

$$-\frac{\hbar^2 K^2}{3} G^{vj}(G^{uk}\Theta_{,j,k,u})_{,v} = O; \quad j,k,u,v = 1,2,...n_c; \text{ Operator Theta Equation}$$

The numerical factor $-\hbar^2 K^2/3$ can, of course, be cancelled; we retain this factor because the operator on the LHS is the imbalance across the second constraint given that the first constraint holds. In the Q-diagonal Schrödinger representation

$$G^{uv} \equiv g^{uv}(\underline{q})I; \quad g^{uv} = g^{vu}; \quad \Theta = \theta(\underline{q})I$$
$$g^{vj}(g^{uk}\theta_{,jku})_{,v} = 0; \quad j,k,u,v = 1,2,...n_c; \text{ Scalar Theta Equation}$$

Notice that the theta equations are independent of $f^j \to F^j, v \to V$.

1.9 A Conjecture

Because, in the Schrödinger representation the coefficients in quadratic $H(\underline{P},\underline{Q})$ are scalars, they are all candidates for θ; so the theta equations are, in fact, QM field equations for the coefficients in quadratic $H(\underline{P},\underline{Q})$. We conjecture that satisfaction of the first two constraints is all that is required for classical mechanics using these field equations.

In the Newtonian mechanics of a single particle of mass m it appears that

$$K = \frac{1}{2m}$$

The imbalance across the second constraint, given that the first constraint holds, is proportional to $\hbar^2 K^2 = (\hbar/m)^2/4$. We conjecture that the imbalance across the n^{th} constraint, given that all the constraints below ($< n$) hold, is proportional to $(\hbar / m)^n$. Given that m is a laboratory mass, $\hbar / m$ is a very small quantity with physical dimensions length × velocity. We have carried out a crude analysis which suggests that, given laboratory conditions, the imbalance in the constraints for $n > 2$ is negligible provided that the lower constraints hold.

1.10 The Flat Coordinate Space $\mathcal{C}$

We now define an infinite continuous, real space $\mathcal{C}$, of dimension $n_c \equiv n_p n_d$, in which a single representative point $\mathcal{Q}$ has coordinates $\underline{q}$. $\mathcal{C}$ is clearly related to $\mathcal{P}$, and, to avoid topological complications, it must have the same sort of connectivity.

The coordinates $\underline{q}$ are shared by $\mathcal{C}$ and $\mathcal{P}$. If $n_p = 1$ then we can choose $\mathcal{C} \equiv \mathcal{P}$.

1.11 $\mathcal{C}$ And $\mathcal{P}$ Are Flat-Riemannian

Reverse quantize quadratic $H(\underline{P}, \underline{Q})$ to give a scalar quadratic Hamiltonian $H(\underline{p}, \underline{q})$ and we find that the scalar Hamilton's equations require that the motion of $\mathcal{Q}$ satisfies the generalised geodesic equation; t is proportional to the geodesic distance in $\mathcal{C}$ and the coefficients $g^{ij}(\underline{q})$ are elements of the fundamental tensor of $\mathcal{C}$. So $\mathcal{C}$ and hence $\mathcal{P}$ are flat-Riemannian.

$$g^{uv} \to G^{uv}; \quad f^{j} \to F^{j}; \quad v \to V$$
$$H(scalar) \equiv \mathrm{K} g^{uv}(\underline{q}) p_u p_v + f^{j}(\underline{q}) p_j + v(\underline{q}) \to$$
$$H(operator) \equiv \mathrm{K}\left\{G^{uv}, P_u P_v\right\} + \left\{F^{j}, P_j\right\} + V; \quad G^{uv} = G^{vu}; \quad u, v, j = 1,2,..n_c$$

Now suppose, for simplicity, that

$$f^{j} = 0 \ \forall\ j; \quad v = 0$$

then

$$H(scalar) \equiv \mathrm{K} g^{uv}(\underline{q}) p_u p_v$$

and the scalar Hamilton's equations give, after eliminating the $\underline{p}$,

$$\ddot{q}^{j} + \Gamma^{j}_{kl}\dot{q}^{k}\dot{q}^{l} = 0; \quad \Gamma^{l}_{ij} \equiv \frac{1}{2} g^{lk}\left(g_{ik,j} + g_{jk,i} - g_{ij,k}\right); \text{ the Geodesic Equation}$$

The generalised geodesic equation holds if $\exists\, f^{j} \neq 0$ and/ or $v \neq 0$

$$\ddot{q}^{m} + \Gamma^{m}_{kl}\dot{q}^{k}\dot{q}^{l} =$$
$$-g^{um}_{,v}\dot{q}^{v} f_u - g^{um}\left(f^{j}_{,u}(\dot{q}_j - f_j) - \frac{1}{2} g^{jk}_{,u}(\dot{q}_j f_k + \dot{q}_k f_j - f_j f_k) + 2\mathrm{K} v_{,u}\right) + f^{m}_{,v}\dot{q}^{v}$$

1.12 The Evidence About $\mathscr{P}$

This is the evidence about $\mathscr{P}$ that we require. The abstract space $\mathscr{P}$ is flat-Riemannian; the coordinates of structureless point particles $\underline{q}$ and the invariant t are continuous, real and unrestricted; the geodesic distance is *proportional to t; so we can regard t as the proper time. We can regard θ as a physical characteristic of the system of particles; and t can be taken as the time of the consciousness and of the clock of a single observer contemplating that system. This is all that is required to allow Newtonian and SR mechanics.*

1.13 Properties Of $\mathcal{C}$ and $\mathcal{P}$ That Emerge Only At The Macrophysical Scale

Note that, because the equations of motion for $\mathcal{Q}$ are derived from the reverse quantization $H(\underline{p},\underline{q})$, the Riemannian connectivity of $\mathcal{C}$ and $\mathcal{P}$ and the interpretation of t, as proportional to geodesic distance in $\mathcal{C}$, are properties that seem to emerge only at the macrophysical scale.

1.14 The Space $\mathcal{C}'$ May Be Curved And Is Tangential To $\mathcal{C}$

Consider an, in general, curved Riemannian space $\mathcal{C}'$ which, when it is flat, is identical to $\mathcal{C}$. Let $\mathcal{C}'$ have coordinates $\underline{x}$ and a fundamental tensor $g'^{ij}(\underline{x})$. Suppose that $\mathcal{C}'$ is tangential to $\mathcal{C}$ at the points $\mathcal{X}$ in $\mathcal{C}'$ and $\mathcal{Q}$ in $\mathcal{C}$; also suppose that the coordinates $\underline{x}$ are geodesic at the pole $\mathcal{X}$. Then we can choose $d\underline{x} = d\underline{q}$; further the functions $g'^{ij}(\underline{x})$ may be chosen arbitrarily save that $g'^{ij}_{,k}(\underline{x}) = 0$ and $g'^{ij}(\underline{x}) = g^{ij}(\underline{q})$ at the points $\mathcal{X}$ and $\mathcal{Q}$.

1.15 Drop The Primes

We can now drop the primes and accept that $\mathcal{C}$ may be curved. When $\mathcal{C}$ is flat we denote the coordinates by $\underline{q}$; when $\mathcal{C}$ may be curved we denote the coordinates by $\underline{x}$.

LECTURE 2- A PRELIMINARY STUDY OF THE KILMISTER EQUATION

2.1 When $\mathcal{C}'$ Is Curved- The Kilmister Equation

We are prompted to ask: what tensor equation, defined in $\mathcal{C}$, reduces to the scalar theta equations for $\theta \equiv g^{ij}(\underline{x})$ at the pole of geodesic coordinates? The answer is the Kilmister Equation (derived by the late Clive Kilmister)

$$K_{ab} \equiv g^{ef}(R_{ab;ef} + \tfrac{2}{3}R_{ae}R_{fb}) = 0;\quad a,b,e,f = 1,2,..n_c$$

otherwise known as the K equation.

The scalar theta equations, defined in flat $\mathcal{C}$ for $\theta \equiv g^{lm}(\underline{q})$, are

$$g^{vj}(g^{uk}g^{lm}(\underline{q})_{,jku})_{,v} = 0;\quad j,k,u,v = 1,2,..n_c$$

These are the field equations for the $g^{lm}(\underline{q})$ in the flat space $\mathcal{C}$. By choosing the coordinates $\underline{x}$ in the curved space $\mathcal{C}'$ to be geodesic at pole $\mathcal{X}$ (the point at which $\mathcal{T}'$ is tangential to $\mathcal{T}$) we can set $d\underline{x} = d\underline{q}$ and $g'^{lm}(\underline{x}) = g^{lm}(\underline{q})$. Thus

$$g'^{vj}(g'^{uk} g'^{lm}(\underline{x})_{,jku})_{,v} = 0; \quad j,k,u,v = 1,2,...n_c$$

We are prompted to ask the above question for the space $\mathcal{T}'$.

2.2 The Motion Of Particles in $\mathcal{P}$ Given The K Equation

The solution process, by which, in principle, the K equation informs the motion of the particles in $\mathcal{P}$, is: $K_{ab} = 0 \rightarrow g_{uv}(\underline{x}) \rightarrow$ a geodesic $\underline{x}(t)$ in $\mathcal{T}' \equiv \mathcal{T} \rightarrow$ coordinates $\underline{x}(t)$ chosen to be geodesic at $\mathcal{X} \equiv \mathcal{Q}$ $\rightarrow d\underline{q}(t) = d\underline{x}(t) \rightarrow \underline{q}(t)$.

2.3 The K Equation Requires $n_c > 3$ If $\mathcal{T}$ Is Curved

Curvature of $\mathcal{T}$ is taken as a symptom of gravity; but we prove that the K equation cannot exhibit curvature unless $n_c > 3$. Other properties of the K equation may be summarised as follows: Contact is made with GR by setting $n_p = 1; \quad n_d = 4$; we then get the Relativistic K Equation. Because, in CT, it is axiomatic that each particle has only one time coordinate, the remaining three must be space-like. So, in the vicinity of $\mathcal{X}$, $\mathcal{T}$ is Minkowskian.

The Ricci tensor $R_{ab} \equiv R^{c}_{.abc}$ is the contraction of the $n_c^2(n_c^2 - 1)/12$ unique elements of the Riemann-Christoffel tensor $R^{a}_{.bcd}$. So the condition that $\mathcal{T}$ is flat

$$R^{a}_{.bcd} = 0$$

comprises $n_c^2(n_c^2 - 1)/12$ unique PDEs for the $n_c(n_c + 1)/2$ elements $g_{uv}(\underline{x})$. By contrast the K equation

$$K_{ab} = 0; \quad K_{ab} \text{ is symmetrical}$$

comprises $n_c(n_c + 1)/2$ unique PDEs for the elements $g_{uv}(\underline{x})$. It follows that if the K equation is satisfied, while curvature is allowed (flatness avoided) then,

$$n_c^2(n_c^2 - 1)/12 > n_c(n_c + 1)/2 \Rightarrow n_c > 3 \Rightarrow R^{c}_{.abc} \neq 0$$

2.4a If $\mathfrak{T}$ Is Truly Empty Then $\Lambda = 0$ And $R^a_b = 0$

Einstein's law of gravity for empty space $R^a_b = 0$ satisfies the K equation; and, if $\mathfrak{T}$ is truly empty, then the relativistic K equation requires that Einstein's Cosmological Constant Λ vanishes. But $\mathfrak{T}$ is not, necessarily, truly empty.

Einstein's equation is

$$G^a_b + \Lambda\delta^a_b + \chi T^a_b = 0; \quad G^u_v \equiv R^u_v - \tfrac{1}{2}R\delta^u_v; \quad n_p = 1, n_d = 4$$
$$\chi \equiv 8\pi G / c^4 = 2.076 \times 10^{-48}\, cm^{-1} gm^{-1} sec^2$$

where T_{ab} is the matter-energy-momentum-stress tensor. If $\mathfrak{T} \equiv$ space-time ($n_p = 1, n_d = 4$) is truly empty then

$$T_{ab} = 0 \Rightarrow G_b^a = -\Lambda\delta_b^a \Rightarrow R_b^a = \Lambda\delta_b^a$$

and the relativistic K equation requires that

$$K_b^a = 0 \Rightarrow \Lambda = 0 \Rightarrow R_b^a = 0$$

2.4b The K Equation And Singularities

The K equation cannot be valid at a true singularity. This has implications when interpreting machine calculations.

Machine calculations (Maple 12) show that the Schwarzschild metric satisfies the relativistic K equation. This is to be expected because it satisfies $R_b^a = 0$ identically. Unfortunately, the machine does not solve the simultaneous ODEs that are derived by substituting the general spherically symmetric (SS) metric

$$ds^2 = -e^{\lambda}r^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\psi^2 + e^{\nu}(dx^4)^2$$

directly into the relativistic K equation.

The Schwarzschild metric contains a singularity at $r = 0$. This is a symptom of two assumptions (which are also central to CT): a) the particle at the origin is a structureless point; b) $\mathcal{T}$ is continuous. Something similar is true of the Big Bang theory; this extrapolates the density of the universe back to a singularity at time zero. Before this point is reached classical theories, such as GR and the K equation, must break down. For this and other reasons the

Big Bang hypothesis is losing ground. Other hypotheses treat of finite densities at time zero and what happens before that instant; yet other hypotheses consider what happens when the universe is fully expanded. Thus we have only sort machine solutions of the K equation which represent well but finitely expanded models; see Section 2.8 .

2.5 The Relativistic K Equation Determines T_b^a

The K equation does not mention matter explicitly. Solutions of the relativistic K equation, given four sets of boundary conditions, determine the fundamental tensor of space-time; and, through Einstein's equations, the corresponding distribution of matter-energy-momentum-stress.

$$G_b^a + \Lambda\delta_b^a + \chi T_b^a = 0; \quad G_\nu^u \equiv R_\nu^u - \tfrac{1}{2}R\delta_\nu^u; \quad n_p = 1, n_d = 4$$
$$\chi \equiv 8\pi G / c^4 = 2.076\times 10^{-48}\, cm^{-1} gm^{-1} sec^2$$
$$K_b^a = 0 \rightarrow g_{uv}(\underline{x}) \rightarrow R_\nu^u \rightarrow T_b^a = -\frac{G_b^a + \Lambda\delta_b^a}{\chi}$$

2.6 Newtonian Gravity As A Perturbation Of Minkowski

Machine calculations), using the Minkowskian metric perturbed by Newtonian gravity, confirm analytical results.

We use quasi-Cartesian coordinates and time x, y, z, t with the metric

$$ds^2 = (1+2U)dt^2 + (-1+2U)ds_0^2; \quad ds_0^2 = dx^2 + dy^2 + dz^2; \quad |U| << 1; \quad c \equiv 1$$

This metric represents scalar gravity where the 'potential' is proportional to U. By setting

$$U = b\phi(x, y, z); \quad b << 1; \quad b \neq 0; \quad b \text{ is a pure number}$$

we ensure that U is 'small' and depends only on the spatial coordinates; so we can regard the gravity as Newtonian. We find

$$R_{aa} = -b\nabla^2\phi; \quad G_4^4 = -2b\nabla^2\phi; \quad K_{uu} = -b(\nabla^2)^2\phi$$

all other elements of the tensors R_{ab}, G_b^a and K_{uv} being identically zero. So the Newtonian approximation to the K equation is

$$(\nabla^2)^2 U = 0$$

In this case we find $G_1^1 = G_2^2 = G_3^3 = 0$. Thus we can define 'density' ρ and the 'stress' σ as

$$\rho \equiv T_4^4 = -G_4^4/\chi = 2\nabla^2 U/\chi; \quad \sigma \equiv T_i^i = -G_i^i/\chi = 0; \quad i = 1,2,3; \quad \text{not summed}$$

The first of these definitions is seen to be a rearrangement of Poisson's equation $\nabla^2 U = (\chi/2)\rho$. The SS solutions are particularly simple

$$(\nabla^2)^2 U = 0 \Rightarrow U = \frac{k_1}{r} + k_2 r^2 + k_3 r + k_4; \quad \rho = \frac{2}{\chi}\nabla^2 U = \frac{2}{\chi}(6k_2 + 2k_3/r)$$

2.7 Two Exact Solutions Using The Cosmological Metric

Machine calculations, given the Cosmological Metric (corresponding to an homogeneous, isotropic space-time), obtain two unique and exact solutions; but they represent empty model universes.

When studying the relevance of GR to cosmology we can adopt the Cosmological Principle. This asserts that space-time is everywhere homogeneous and isotropic. It turns out the symmetry so implied reduces, from ten down to one, the number of unknown functions (Schur's theorem). In quasi-polar coordinates the corresponding 4-metric is

$$ds^2 = dt^2 - \left[\frac{\mathrm{R}(t)ds_0}{1 + kr^2/4}\right]^2; \quad k = 1,0,-1; \quad c \equiv 1$$

where, in the E3 subspace, r is the radius to the origin, ds_0 is an element of length and $\mathrm{R}(t)$ is a pure number (which is an unknown function of time t). The physical dimensions are chosen so that the speed of light is unity and the length r is a pure number. With the above metric the tensor K_b^a is diagonal; so the K equation gives four ODEs for $\mathrm{R}(t)$; the solutions of these must be consistent. The two unique and exact solutions are

$$\mathrm{R}(t) = Ct + D; \quad C^2 + k = 0 \Rightarrow C = \pm 1; k = -1 \text{ and } C = 0; k = 0$$

The first solution $(k = -1)$ represents an open (hyperbolic) expanding/contracting model universe; and the second solution $(k = 0)$ represents a static model universe which is neither open nor closed. With both solutions the Riemann, Ricci, Einstein and Kilmister tensors are null. Thus the space is empty and Minkowskian.

2.8 Constant Density Perturbations Of The Exact Solutions

Perturbations of the exact solutions are possible. They can represent model universes that are well expanded, contain uniform, low density distributions of matter-energy-stress and have roughly constant Hubble parameters. All these model universes, once expanding, continue so to do.

Given the Cosmological Principle and certain coordinate systems it turns out that the Einstein tensor is diagonal and $G_i^i = G_j^j; G_j^i = 0; i \neq j; \quad i, j = 1,2,3$ (not summed). We can therefore define ρ and σ as above. The perturbation investigated is of the form

$$R(t) = Ct + D + af(t); \text{ either } C = \pm 1; k = -1 \text{ or } C = 0; k = 0; \quad |a| << 1; \quad a \neq 0$$

In the $k = -1$ case we get

$$G_4^4 = -\frac{6af'(t)}{(D+t)^2} = -\eta a; \text{ (say) where } \eta \text{ is a constant.}$$

$$\Rightarrow f(t) = \frac{\eta}{18}(D+t)^3 + C1; \quad C = \pm 1; k = -1; \quad C1 = 0 \text{ by convention}$$

$$G_1^1 = G_2^2 = G_3^3 = G_4^4 = -\eta a \Rightarrow \rho = \sigma = \eta a / \chi$$

with K_b^a either of second order in a or zero. In the $k = 0$ case

$$G_4^4 = -3\left[\frac{af'(t)}{D}\right]^2 = -a\eta \Rightarrow f(t) = \pm t\sqrt{\eta/(3a)} + C1; \quad C = 0; k = 0;$$

$|a| << 1; \quad \eta / a > 0; \quad C1 = 0$ by convention
$G_1^1 = G_2^2 = G_3^3 = -\eta a / 3$
$\Rightarrow \rho = a\eta / \chi$ and $\sigma = \rho / 3; \quad \eta / a > 0 \Rightarrow \rho = a^2(\eta / a) > 0$

with K_b^a either of second order in a or zero. Substituting the expressions for $R(t)$ into expressions for the Hubble parameter, given by McVittie, we get different but roughly constant results.

2.9 Modified Cosmological Metric ($k = -1$)

We have modified the cosmological metric so that it is possible for it to describe weak gravity as a local perturbation. One of these perturbation solutions expands at an almost constant rate, does not permit Newtonian gravity and may correspond to an early phase when the universe was filled only with radiation.

$$ds^2 = (1 + 2U)dt^2 + (-1 + 2U)\left[\frac{R(t)ds_0}{1 + kr^2/4}\right]^2; \quad k = 1,0,-1; \quad |U| << 1$$

where U is a scalar potential representing Newtonian gravity. We choose

$U = b\phi(r); \quad b << 1; \quad b \neq 0; \quad b$ is a pure number

imposing SS on U for simplicity. The perturbation solution for the $k = -1$ case is

$R(t) = Ct + D + af(t); \quad 1 >> |a|; \quad C = \pm 1$
$f(t) = C1 + C2/(D + t) + C3(D + t) + C4(D + t)^3; \quad \phi(r) = 1/2 - C3a/b$

where the $C1$ to $C4$ are constants of integration and, by convention, $C1 = 0$. Observe that $\phi(r)$ is constant; that is, Newtonian gravity is not allowed. The K equation is only satisfied if

$$a = \frac{b}{3C3} \Rightarrow \phi(r) = 1/6; \quad 1 >> |a|, |b|$$

where $C3$ is not small. Four distinct simultaneous ODEs are produced by $K_b^a = 0$. These could be paired six ways to calculate $f(t)$ and $\phi(r)$. In fact we pick $K_1^1 = 0$ and $K_4^4 = 0$ and we solve them with the functions $f(t)$ and $\phi(r)$ as unknown. We find

$$\rho = \frac{6a}{\chi}\left(\frac{C2}{D^4} - 3C4\right); \quad \sigma = \frac{2a}{\chi}\left(\frac{C2}{D^4} + 9C4\right); \quad t = 0$$

2.10 Modified Cosmological Metric ($k = 0$)

The other perturbation solution expands at a much lower rate, permits Newtonian gravity and has a matter-energy density ρ *which is a sum of two terms* ρ_0 *and* ρ_{00}. *The term* ρ_0 *is the square of a real and so is positive. The term* ρ_{00} *can have either sign and arises because the K equation is of fourth order.*

The perturbation solution for the $k = 0$ case is

$$R(t) = Ct + D + af(t); \quad 1 >> |a|; \quad C = 0$$
$$f(t) = \frac{1}{6}C5t^3 + \frac{1}{2}C6t^2 + C7t + C8; \quad b\phi(r) = C1 + C2r + C3/r + C4r^2;$$

where the $C1$ to $C8$ are constants of integration and, by convention, $C1 = 0; \quad C8 = 0$. Observe that $\phi(r)$ is of the same form as the SS solution for the Newtonian potential U; in other words, Newtonian gravity is allowed. Observe also that, when $t = 0$, $\mathrm{R}'(t) = aC7; \quad 1 >> |a|$; in other words the rate of expansion/ contraction is small. The Kilmister tensor vanishes, to a sufficient order, only if the constants $C2$ and $C4$ are of the same order as b and certain other constants of integration vanish

$$C2 \equiv bc2; C4 \equiv bc4; \quad C3 = 0; \quad C6 = 0; \quad C5 = 0; \quad c2, c4 \text{ finite}$$

$$\rho = \rho_0 + \rho_{00}; \quad \rho_0 \equiv \frac{3(aC7)^2}{D^2\chi}; \quad \rho_{00} \equiv 12\frac{bC4}{D^2\chi} = 12\frac{b^2c4}{D^2\chi}; \quad \rho_0 > 0$$

The final calculation sets _ $C2 = 0 \Rightarrow c2 = 0$ to remove dependence of the density on r; i.e., it is assumed that there is a continuum of matter-energy but there are no point particles.

2.11 Vacuum Energy

The term ρ_{00} is interpreted as a vacuum energy; if $\rho_{00} < 0$ this extra term could explain the Dark Energy effect; that is an expansion over and above that which is predicted by GR given the Big Bang. Note that the condensation of matter from a very high temperature radiation field is a quantum phenomenon and thus outwith the competence of the K equation.

We have

$$\rho = \rho_0 + \rho_{00}; \quad \rho_0 > 0$$

The term ρ_0 is positive. The term ρ_{00} can have either sign and arises because the K equation is of fourth order. The term ρ_0 is interpreted as ordinary mass density possibly including Dark Matter. It is postulated that

$$\rho = 0 \Rightarrow \rho_{00} = -\rho_0 \Rightarrow \rho_{00} < 0$$

Because estimates of ρ_0 are known this postulate gives a sign and a magnitude to ρ_{00}; and, because ρ_{00} is then negative, test particles, in the model universe, experience an acceleration of recession along the line joining them; that is the universe expands. Assign a typical estimate of

$$\rho_0 = 8 \times 10^{-30} gm\ cm^{-3};\ \text{roughly the closure value}$$

and, on a crude analysis, the acceleration of recession per unit of separation has an Hubble parameter of $14 km\ s^{-1} per\ 10^6 ly$. This value is at the lower bound of modern estimates of the Hubble constant (15 to 30 $km\ s^{-1} per\ 10^6 ly$). So the Dark Energy effect predicted is significant.

The postulate imagines that the universe began with equal amounts of positive and negative energy; so the total was zero. But the positive energy tends to clump whereas the negative energy tends to spread out.

2.12 Flat $\mathcal{C}$

We have studied, analytically, flat and almost flat $\mathcal{C}$ in order to investigate the links of CT with Newtonian mechanics and curvature as the central idea of GR.

It is well known that in a Riemannian space we can always transform coordinates so that $g^{jk} = \pm\delta_{jk}$ at a point. If, in addition, the space is flat then this condition holds throughout the space and the flat coordinates may be used likewise. The scalar theta equation is

$$g^{vj}(g^{uk}\theta_{,jku})_{,v} = 0; \quad j,k,u,v = 1,2,...n_c; \quad \theta = g^{lm}(\underline{q}), f^l(\underline{q}), v(\underline{q})$$

We may put the constant metric tensor in, e.g. in the Euclidean case,

$$g^{uv} = g_{uv} = \delta_{uv}; \quad u,v = 1,2,3 \Rightarrow g^{vj}(g^{uk}\theta_{,jku})_{,v} = (\nabla^2)^2\theta = 0$$
$$\Rightarrow (\nabla^2)^2 f^l = 0 \;\&\; (\nabla^2)^2 v = 0$$

or, in the Minkowskian case,

$$g_{11} = g_{22} = g_{33} = -1; \quad g_{44} = 1; \quad g_{jk} = 0,\ j \neq k$$
$$\Rightarrow \Box^4\theta = 0; \; \Box^2 \equiv -\nabla^2 + \frac{\partial^2}{\partial(q^4)^2}; \; x^4/c \text{ is coordinate time}$$

Alternatively, in the Newtonian mechanics, we may write the Hamiltonian and compare it with the general case. For a single particle

$$H \equiv \frac{1}{2m}\sum_{j=1}^{3} P_j^2 + V(\underline{Q}); \quad \underline{Q} \equiv \underline{q}I; \quad V(\underline{Q}) \equiv v(\underline{q})I \Rightarrow \mathrm{K} = \frac{1}{2m}; \quad n_p = 1; \quad n_c = 3$$
$$\Rightarrow g^{uv} = g_{uv} = \delta_{uv}; \quad u,v = 1,2,3 \Rightarrow g^{vj}(g^{uk}v_{,jku})_{,v} = (\nabla^2)^2 v = 0$$

For two particles

$$H \equiv \frac{1}{2m_1}\sum_{j=1}^{3} P_j^2 + \frac{1}{2m_2}\sum_{j=4}^{6} P_j^2 + V(\underline{Q}); \quad \underline{Q} \equiv \underline{q}I; \quad V(\underline{Q}) \equiv v(\underline{q})I; \quad n_p = 2; \quad n_c = 6$$

$$\Rightarrow g^{uu} = \frac{1}{2Km_1}; \quad g^{(u+3)(u+3)} = \frac{1}{2Km_2}; \quad g^{uv} = 0, u \neq v; \quad u,v = 1,2,3$$

$$\Rightarrow g^{vj}(g^{uk}v_{,jku})_{,v} = \left(\frac{\nabla_1^2}{m_1} + \frac{\nabla_2^2}{m_2}\right)^2 v = 0; \quad m_1, m_2 \neq 0$$

where $\nabla_1^2 v$ and $\nabla_2^2 v$ are evaluated, respectively, in the neighbourhoods of the two particles.

2.13 Flat $\mathcal{T}$ – Two Particles (v Invariant)

According to Einstein the expression of physical law must be independent of the choice of coordinates. In the case of two particles $v(\underline{q})$ must be an invariant independent of rotation and translation of the coordinate axes in $\mathscr{P}$. The only such invariants, associated with the classical model, are the distance in $\mathscr{P}$ between the two particles

$$l \equiv +[(q^1 - q^4)^2 + (q^2 - q^5)^2 + (q^3 - q^6)^2]^{1/2}$$

and functions of same. So v is a function of l. Therefore

$$\nabla_1^2 v = \nabla_2^2 v = \left(\frac{d^2}{dl^2} + \frac{2}{l}\frac{d}{dl}\right)v$$

and, after division by $(1/m_1 + 1/m_2)^2$,

$$\left(\frac{\nabla_1^2}{m_1}+\frac{\nabla_2^2}{m_2}\right)^2 v=0 \Rightarrow \left(\frac{d^2}{dl^2}+\frac{2}{l}\frac{d}{dl}\right)^2 v=0$$

The general solution of this equation is ($c_1,...,c_4$ are constants of integration)

$$v=\frac{c_1}{l}+c_2l^2+c_3l+c_4$$

If l is small enough, we may neglect the terms $c_2l^2+c_3l$. Then v provides an inverse square force (of attraction if $c_1<0$ and of repulsion if $c_1>0$) along the line joining the particles. In this sense the new field equation is successful. The neglected terms are supposed only significant when the distances are cosmological. Strictly speaking, because $\mathcal{C}$ is flat, we could say that the result does not apply to gravity; rather, we might say, it must apply, if it applies to any force in nature, to electrostatics.

2.14 Nearly Flat $\mathcal{C}$

We have already encountered, in the machine calculations (Art. (2.6)), an example of a slightly curved $\mathcal{C}$ that induces gravity on a single particle.

$$g_{11}=g_{22}=g_{33}=-1+2U;\quad g_{44}=1+2U;\quad g_{jk}=0,\, j\neq k;\quad n_p=1;\quad n_c=4$$
$$\Rightarrow g^{vj}(g^{uk}g^{lm}{}_{,jku})_{,v}=\sum_{j,k,l} g^{jj}(g^{kk}g^{ll}{}_{,jkk})_{,j}\approx 2\,(\nabla^2)^2U=0;\quad |U|<<1$$

since, by definition, U does not depend on the coordinate time q^4.

A more sophisticated calculation assumes that there are two particles in $\mathcal{P}$ and eight coordinates in a slightly curved $\mathcal{C}$; the fundamental tensor approximates that of a double Minkowski space-time. The calculation equates the coordinate accelerations, generated by a weak scalar potential ω, with the same quantities along a geodesic in $\mathcal{T}$. This relates the Christoffel symbols to the coordinate derivatives of ω. The Ricci R_{ab} tensor is thereby approximated. Substitution into the K equation yields the approximation

$$\left(\frac{\nabla_1^2}{m_1}+\frac{\nabla_2^2}{m_2}\right)^2 \omega = 0$$

To prove that, with this specification, $\mathcal{T}$ is can be curved we evaluate, approximately, the Riemann-Christoffel tensor. The non-zero components are

$$R^{\mu}_{.4\nu4} \approx \Gamma^{\mu}_{44,\nu} - \Gamma^{\mu}_{4\nu,4} = \frac{\omega_{,\mu\nu}}{2c^2 m_1}; \quad R^{\mu}_{.44\nu} \approx \Gamma^{\mu}_{4\nu,4} - \Gamma^{\mu}_{44,\nu} = -\frac{\omega_{,sp}}{2c^2 m_1};$$

$$R^{\mu}_{.8\nu8} \approx \Gamma^{\mu}_{88,\nu} - \Gamma^{\mu}_{8\nu,8} = \frac{\omega_{,\mu\nu}}{2c^2 m_2}; \quad R^{\mu}_{.88\nu} \approx \Gamma^{\mu}_{8\nu,8} - \Gamma^{\mu}_{88,\nu} = -\frac{\omega_{,sp}}{2c^2 m_2}$$

2.15a Motion Of A Charged Particle In Flat $\mathcal{T} \equiv \mathcal{P}$

We have also considered the motion of a charged particle in a field governed by Maxwell's equations. Given suitable definitions CT seems consistent both with the motion and the field.

The scalar Hamiltonian, with Hamilton's equations, gives

$$\ddot{q}^m + \Gamma^m_{kl}\dot{q}^k\dot{q}^l =$$
$$-g^{um}_{,v}\dot{q}^v f_u - g^{um}\left(f^j_{,u}(\dot{q}_j - f_j) - \frac{1}{2}g^{jk}_{,u}(\dot{q}_j f_k + \dot{q}_k f_j - f_j f_k) + 2\mathrm{K}\nu_{,u}\right) + f^m_{,v}\dot{q}^v$$

where, of course, $\mathcal{T}$ may be curved. Now consider a single particle in Minkowskian $\mathcal{P}$ and flat $\mathcal{T}$. With Galilean coordinates the metric of $\mathcal{T} \equiv \mathcal{P}$ is

$$g_{jk} = 0,\ j \neq k,\ g_{jj} = g^{jj} = -1, -1, -1, +1$$

Substituting the metric the acceleration becomes

$$\ddot{q}^\lambda = \dot{q}^j(f_{j,\lambda} - f_{\lambda,j}) - f^j_{,\lambda} f_j + 2\mathrm{K}\nu_{,\lambda};\quad \dot{q}^\lambda = -\dot{q}_\lambda;\quad \dot{q}^4 = \dot{q}_4;\quad f^\lambda = -f_\lambda;\quad f^4 = f_4$$
$$\ddot{q}^4 = -\dot{q}^j(f_{j,4} - f_{4,j}) + f^j_{,4} f_j - 2\mathrm{K}\nu_{,4}$$

where the Greek indices run from 1 to 3 and the Roman indices run from 1 to 4. In the next section we compare these equations of motion with those of a single charged particle in an EM field

$$c\frac{d(m\underline{\dot{q}})}{dq^4} = \frac{\varepsilon}{c}(\underline{u} \times \underline{b} + c\underline{e});\quad c\frac{d(m\dot{q}^4)}{dq^4} = \varepsilon\underline{e}.\underline{u};\quad \underline{b} \equiv \nabla \times \underline{a};\quad \underline{e} \equiv -\nabla\phi - \frac{\partial \underline{a}}{\partial q^4}$$

where: m is the rest mass, ε is the charge, q^4/c is the coordinate time, $\underline{\dot{q}}$ is the proper 3-velocity, $\underline{u}$ is the 3-velocity, $\underline{a}$ is the 3-vector potential, ϕ is the scalar potential, $\underline{b}$ is the 3-magnetic induction, $\underline{e}$ is the 3-electric intensity. These SR equations have received extensive experimental verification.

2.15b Motion Of A Charged Particle In Flat $\mathcal{C} \equiv \mathcal{P}$

Written in coordinate form the above vector equations of motion are

$$\ddot{q}^{\lambda} = \frac{\varepsilon}{mc}\left(\dot{q}^{\mu}(a_{\mu,\lambda} - a_{\lambda,\mu}) - \dot{q}^{4}(\phi_{,\lambda} + a_{\lambda,4})\right); \ddot{q}^{4} = -\frac{\varepsilon}{mc}\left(\dot{q}^{\mu}(a_{\mu,4} + \phi_{,\mu})\right); u^{j} \equiv c\frac{dq^{j}}{dq^{4}} \Rightarrow u^{4} = c$$

These equations can be compared, directly, with the constraints equations (see (Art. 2.15a)) by regarding the $\underline{\dot{q}}^{j}$ as arbitrary. The comparison yields

$$f_{\mu,\lambda} - f_{\lambda,\mu} = \frac{\varepsilon}{mc}(a_{\mu,\lambda} - a_{\lambda,\mu}); \quad f_{4,\lambda} - f_{\lambda,4} = -\frac{\varepsilon}{mc}(\phi_{,\lambda} + a_{\lambda,4});$$
$$f^{j}_{,\lambda} f_{j} = 2\mathrm{K}v_{,\lambda}; \quad f^{j}_{,4} f_{j} = 2\mathrm{K}v_{,4}$$

The first three of these equations relate the 4-curl of the f_j to curl $\underline{a}$ and grad ϕ. But the last two equations relate the gradient of v to $f^{j}_{,k} f_{j}$. If we define

$$4\mathrm{K}v \equiv f^{j} f_{j} \Rightarrow 4\mathrm{K}v_{,k} \equiv f^{j}_{,k} f_{j} + f^{j} f_{j,k} = 2 f^{j}_{,k} f_{j} \Rightarrow f^{j}_{,k} f_{j} = 2\mathrm{K}v_{,k}$$

then the original scalar Hamiltonian becomes

$$H \equiv \mathrm{K}\left(p^{u} + \frac{f^{u}}{2\mathrm{K}}\right)\left(p_{u} + \frac{f_{u}}{2\mathrm{K}}\right); \text{ scalar Hamiltonian in Minkowski space}$$

With appropriate definitions, this is,

$$H = \frac{1}{2m}\left(-\sum_{\mu=1}^{3}\left(p_\mu + \frac{\varepsilon}{c}a_\mu\right)^2 + \left(p_4 - \frac{\varepsilon}{c}\phi\right)^2\right); \quad \mathrm{K} \equiv \frac{1}{2m}; \quad f_\mu \equiv \frac{\varepsilon}{mc}a_\mu; \quad f_4 \equiv -\frac{\varepsilon}{mc}\phi$$

which can be shown to give the equations of motion of a single charged particle directly. Notice that the definitions (of the f_j in terms of the a_μ and ϕ) are sufficient but they may not be necessary. This identification of the f_j, v is classical because it is based the scalar Hamiltonian.

2.15c The Impact On Maxwell's Equations

When there is but a single particle in $\mathscr{P}$, and $\mathcal{T} \equiv \mathscr{P}$ is Minkowskian with Galilean coordinates, the theta equation is

$$\Box^2(\Box^2\theta) = 0; \quad \Box^2 \equiv -\nabla^2 + \frac{\partial^2}{\partial(q^4)^2}; \quad q^4/c \text{ is coordinate time; see (27a)}$$

Notice that solutions of the standard wave equation $\Box^2\theta = 0$ are also solutions of the theta equation. Because the f^j are candidates for θ these results require that

$$\Box^2(\Box^2\underline{a}) = \underline{0}; \quad \Box^2(\Box^2\phi) = 0; \text{ see Art. 2.15b; new equations}$$

With 3-vector notation Maxwell's equations, in the Lorentz gauge, can be written as:

$$\nabla.\underline{e} = 4\pi\rho;^* \quad \nabla \times \underline{e} = -\frac{\partial \underline{b}}{\partial q_4}; \quad \nabla.\underline{b} = 0; \quad \nabla \times \underline{b} = \frac{4\pi \underline{j}}{c} + \frac{\partial \underline{e}}{\partial q_4}^*$$

$$\nabla.\underline{a} + \frac{\partial \phi}{\partial q^4} = 0;^* \quad \nabla.\underline{j} + c\frac{\partial \rho}{\partial q^4} = 0; \quad \Box^2 \underline{a} = \frac{4\pi}{c}\underline{j}; \quad \Box^2 \phi = 4\pi\rho$$

$$\underline{b} \equiv \nabla \times \underline{a}; \quad \underline{e} \equiv -\nabla\phi - \frac{\partial \underline{a}}{\partial q^4}$$

Given three* of these equations as definitions, five of the first eight equations can be deduced as identities. To do this one needs the vector identities

$$\nabla \times (\nabla \phi) = \underline{0}; \quad \nabla.(\nabla \times \underline{a}) = 0; \quad \nabla \times (\nabla \times \underline{a}) = \nabla(\nabla.\underline{a}) - \nabla^2 \underline{a}$$

We can also show, using vector the identities, that the components of the force fields satisfy the wave equation $\Box^2 \underline{e} = \underline{0};\ \Box^2 \underline{b} = \underline{0}$. The new equations require that the current and charge densities satisfy the standard wave equation $\Box^2 \underline{j} = \underline{0};\ \Box^2 \rho = 0$. It seems that the assumptions $\underline{j} = \underline{0}$ and $\rho = 0$ are appropriate. The Maxwell equations with $\underline{j} \neq \underline{0}$ and $\rho \neq 0$ reduce the behaviour of an host of charged particles to that of a continuous medium characterised by $\underline{j}$ and ρ.

2.16 An Hamiltonian That Satisfies All The Constraints

We have experimented with various simple Hamiltonian forms, from conventional classical mechanics, to see how far up the hierarchy of constraints they can be made to go.

The one-dimensional form

$\frac{P^2}{2m}$; $n_c = 1$; m a scalar constant

satisfies *all* the constraints provided that θ satisfies the scalar theta equation; and, if $\theta \equiv 1/(2m) =$ a constant then, that is automatic.

2.17 A Post Lecture Note

Since this lecture was given Louis Kauffman and I have examined the derivation of the Kilmister equation thoroughly. Despite initial misgivings we find it to be sound; see Section 2.1 .

The SS solution of the scalar theta equation, for the Newtonian gravitational potential U in an almost flat quasi-Euclidean space, is

$$(\nabla^2)^2 U = 0 \Rightarrow U = \frac{k_1}{r} + k_2 r^2 + k_3 r + k_4; \quad |U| << 1; \quad k_4 = 0$$

where the radial acceleration of a test particle is proportional to

$$-dU/dr$$

; see Section 2.6. This, it is supposed, is the gravitational potential due to a stationary point mass at the origin. The k_1/r term represents the familiar inverse square force postulated by Newton. The remaining terms must be either very small or zero to have escaped experimental detection. Analysis shows that the term $k_2 r^2$ causes the distance between test particles to accelerate/ decelerate at a rate proportional to r. This has been confirmed, by algebraic calculation with the K equation, for a uniform low density

model universe; see Sections 2.10/ 2.11. An extra postulate (e.g. that the Universe began with zero energy) is necessary to establish the sign and magnitude of k_2. *As has been remarked, this may explain the Dark Energy effect. The* $k_3 r$ *term produces a constant radial acceleration/ deceleration. Geoffrey Constable has drawn our attention to a 2004 report that the Voyager probe, then at 86 AU from the sun, was not quite in its right place; it seemed to be experiencing a small* ($10^{-9} ms^{-2}$) *constant acceleration towards the sun. If the effect is real, and CT provides an explanation then, the effect is somehow due to non-uniformity in the distribution of local matter; it could not appear, therefore, in the results of calculations so far performed with the K equation. Note that the extra terms* $k_2 r^2 + k_3 r$ *must make an apparent contribution to the Dark Matter effect.*

'God be in my Heart and in my Thinking': Towards Self-awareness in Maths and Science

© Nicola Graves Gregory Feb 2011
Nicky7@ecobiz.co.uk

'Nel mezzo del cammin di nostra vita ...' In the middle of the road of our life' - this is how Dante begins the 'Divine Comedy'. When we come to consciousness, we never start at the beginning. We always start in the middle, from where we are. We find ourselves in the middle of life, in the middle of our world - our hugely heterogeneous, ambiguous world. We only ever see beginnings from the outside.

- We are always in the middle.
- If we take this seriously, i.e. choose to live coherently, congruently with this understanding, then there are implications for our maths and science.
- We need to base ourselves in the awareness of this 'middle' position, the one in us who is aware of being aware, aware of being conscious, in Vedic or Buddhist terminology, the witness.
- This is a turn of the spiral on Descartes' self-questioning. Descartes wanted certainty, absolute certainty. He wanted to build an edifice of certain knowledge. So he took a few basic premises, clear and distinct ideas, and built. We have been living in that building in the centuries since.
- It is of the essence of the acceptance of the 'middle' position that absolute certainty of knowledge (in the sense of sapere, savoir, wissen) is not possible, but a lot of very useful, temporary (maybe centuries, millennia long) knowledge, science, is.
- Knowing (cognoscere, connaitre, kennen) has been sadly neglected. It is time to make amends.
- The witness is aware that all language is limited, all spoken truths are not the truth (Tao).
- The Cartesian position depended on splitting body and mind. We have been living with the results of that – splitting upon splitting.
- The witness position makes a more holistic knowledge (sapere and cognoscere) possible.

- The witness position is one of self-awareness. Our witness notices what we notice and observes us noticing. Our noticing always has a slant, a spin which comes from our education and motivations. (To the optimist, the glass is half full. To the pessimist, the glass is half-empty. To the engineer, the glass is twice as big as it needs to be.)
- The aim is to develop knowledge from the witness position. This was Pythagoras' teaching. His holistic mathematics is '*ta mathemata*', 'those things which have been learned'. Later on part of maths was abandoned, philosophy, ontology, wholeness. The quantitative aspect *logistike* took over.
- When we look at the maths and science which have been created on the basis of the body-mind split we see that we are technologically proficient, and ethically irresponsible.
- How do we start again now, beginning again in the middle? Starting from the 'middle' position, we recognise that the pursuit of knowledge is a human endeavour. We are concerned with what we as humans might know about this universe and our place in it, not what absolutely is, nor simply what we can make the world do for our venal desires.
- We need to review our existing maths and science as a product of human endeavours.
- To develop self-awareness in the act of doing science or maths, is a psychological development, not just a matter of technical expertise.
- We need to start from the awareness of the profound gift of human life on this beautiful, amazing planet.

****

It is a chimaera to think that we can ever know the ultimate [sic] beginnings of this world, or the absolute foundations, in the sense of possessing them, being able to write them down. We can attempt to move towards such goals asymptotically, and learn along the way.

The linguist, George Lakoff, in 'Women, Fire and Dangerous Things', speaks of a conceptual middle, namely, first level categorisation which corresponds to the level of action and interaction. Roger Brown, in 'Social Psychology', lists various properties of this level:

it is the level of distinctive actions
it is the level which is learned easiest and at which things are first named
it is the level at which names are shortest and used most frequently
it is a natural level of categorisation, as opposed to a level created by 'achievements of the imagination'.

This level is intermediate between abstract generality and high specificity, e.g. 'dog' lies between 'animal' and 'great dane'.

Euclid coined the term '*koina*' for this natural level of categorisation in mathematics, the initial terms which we all have a basic idea of, which we can roughly agree on, but can not define precisely. It has been and is, a major mathematical endeavour to examine these terms, deepen our understanding and refine our terminology in these foundational (or quasi- foundational) areas.

Euclid himself did not go down this path. He carried out the encyclopaedic undertaking of recording all the extant Greek mathematical work with demonstrations of the theorems according to axiomatic, deductive logic. It was this method, the demonstration form, that was revolutionary. It created the model for all subsequent mathematical proofs and became the cement for the construction of the vast mathematical edifice which followed.

Descartes took this method as the inspiration for his rationalist, foundationalist philosophy, wanting to create a system of certain knowledge, based on 'clear and distinct ideas'. But it does not actually generate new mathematics, it only checks the validity of mathematical theorems once they have come into human consciousness. Imre Lakatos argues in *Proofs and Refutations,* that the heuristic value of the proof is more important than any certainty factor; it is in the process of attempting to prove and disprove theorems that we discover where our ideas are unclear, and new ideas can come into being.

We might picture this by imagining the *koina,* our normal level of discourse, as the surface of a sphere. We can choose to excavate towards the centre, or build outwards into space. Interestingly, the analogy could work with excavation being either to more specific, individual instances or towards more generally applicable, deeper concepts.

As Sylvester said (1876), "a *mathematical idea should not be petrified in a formalised axiomatic setting, but should instead be considered as flowing as a river. One should always be ready to change the axioms, preserving the informal idea.*" Here we return again to Dante's middle: we are here now between past and future.

So we are always in the middle in (at least) 2 ways, conceptually and temporally. If we take this seriously, i.e. choose to live coherently, congruently with this understanding, then there are implications for our maths and science.

When we recognise that we are always in a conceptual middle, we see that to deepen our understanding of life, of reality, we need to investigate our *koina*, the undefined terms, the intersubjectively agreed basic terms of discourse, to uncover assumptions of which we are otherwise unaware.

When we recognise that we are always in a temporal middle, now, between the past and the future, then to the extent that we wish to be more responsible as people, we look to the past to see what we have learned, what mistakes we have made, in order to make informed decisions about the future. We do this as individuals. Businesses, institutions do this. But academic disciplines, such as maths and the sciences, are more nebulous and this reviewing role within the discipline itself does not exist.

There are academic disciplines of history of maths and science, and of philosophy of maths and science, but they are epiphenomenal. They retrospectively describe what the actors, the participants, are doing. They are an audience. They do not take an active part in the unfolding drama. They have no impact on the decisions that are made on a daily basis around the world. This is not a responsible, grown up way to go on!

As regards the conceptual middle, it is important to recognise that the sphere of discourse concerns not only the external world but also ourselves. We do not understand the external world and we do not understand ourselves. We are attempting to understand both better. And these are not separate tasks.

We see the external world through the filter of our concepts. 'Beginner's mind', 'a child's view', 'thinking out of the box' – these are what we need to reapproach this amazing world afresh. We need to question our questions. Our verbal questions are already framed in the concepts that we want to liberate ourselves from.

Our very perceptions are culturally determined. We learn to use our eyes to focus on physical objects very early on when we emerge from the womb. Clearly this ability has important survival value, but we stop ourselves seeing in other ways; we make assumptions in the act of seeing physical objects,

As Lawrence Wright says in *Perspective in Perspective,*

"*We prefer an ordered world, regular patterns, familiar forms, and when flaws or distortions occur, provided they are not too gross, our mind's eye tidies them up. We see what we want or expect to see ...*"
"*... We never see the whole of any solid object at any given moment; even to see the whole surface of so simple an object as a sphere, six successive viewpoints are needed ... But we accept at a single glance that it is a sphere: we complete it on the assumption that it is a consistent shape, and the simplest possible one. In fact it might be any one of countless forms that are only partly spherical. There is a fountain court at Granada where visitors are told the story of a Moorish Caliph about to appoint a Grand Vizier. He invited the candidates to identify an object lying or floating in shallow water. All but one promptly said it to be an orange. One picked it up, and identified it as half an orange; he got the job.*"

So what can we do now? Where can we stand? To act in accordance with the 'middle' position, the awareness of the limitations of our normal, everyday concepts, we need to base ourselves in our awareness of being aware: in Vedic or Buddhist terminology, the witness stance. These traditions have disciplines to help people develop this faculty. Given the spiritual, esoteric nature of the Pythagorean school, this was probably part of their training also, and hence at the roots of our mathematics.

The literal meaning of the Greek word, '*esoterikos*' is 'inner'; it came to have its present meaning from the nature of the inner membership of the Pythagorean brotherhood. Postulants had to wait around on the outside for some time (being one of the '*exoteriki*') to show they were serious, before being allowed to enter and become one of the '*akousmatiki*' (listeners). After some further proving period they were allowed to participate, becoming one of the '*esoteriki*'.

It was Pythagoras who originally coined the term 'mathematics'. The Greek word is '*ta mathemata*' which means literally, 'those things which have been learned'. Given the mystery nature of the Pythagorean school it is unlikely that '*ta mathemata*' was simply a matter of learning external formulae (times tables or the like). It is much more likely that it was a holistic learning, of which the formulae were part and possibly represented more than might at first be seen.

To begin to rediscover, or invent self-awareness in maths and science now will involve a number of areas of work. One area is to look at maths and science in historical, social, political contexts, the temporal threads, and ask intellectual history and social psychoanalytic questions. We could see maths and science as a form of cultural dreaming and ask what might the various mathematical and scientific ideas tell us about our consciousness, the nature of the consciousness that created them (as distinct from what they say about physical reality). We could look at mathematical and scientific ideas in relation to the social and political contexts pertaining at the times of their arising and ask what might the ideological content be?

Another aspect of self-awareness is awareness of our motivations. What is the motivation for scientific inquiry? Roger Penrose called his recent book on the history of science, '*The Road to Reality*'. Probably most scientists would assent to the notion that science attempts to discover what is fundamentally real in the world around us. But there are many assumptions in this notion of reality which we start to see when we observe that another way of stating this now would be, 'how things work'.

This languaging is indicative of the attitude to nature now. There was a major change in attitude from the scholastic period to the Enlightenment, from the desire to understand nature to the desire to control nature. Nature had successfully been established as inanimate by the Inquisition which chose brute force over verbal argument and either killed those who thought otherwise or scared them into silence. The Enlightenment motivation was epitomised in Francis Bacon's famous dictum, '*Knowledge is power*', published in 1597 in his '*Meditationes Sacrae*'.

Our current science, which came from this motivation, has been very successful in developing technological power. But this has begun to backfire. The lack of self-awareness has resulted in a lack of responsibility and a lack of the precautionary principle, and we are seeing the dangerous results, in planetary pollution, climate change and escalating weaponry, of the work of a sorcerer's apprentice, of knowledge without wisdom.

The present notion of physical reality as separate from us, comes from Descartes' creation of a rigid duality between *res cogitans* and *res extensa,* the infamous mind-body split. We can not know reality in an abstract, separate-from-reality way. We experience reality. This is a process, and we are in it. As Schopenhauer said, "*...materialism is the philosophy of the subject who forgets to take account of himself.*"

We are in the world, in Life. We are part of a greater wholeness, whether we call it the Universe, Bohm's implicate order, or God. This is the other side of self-awareness. I can be aware of myself as a conscious entity with respect to some thing or being or process that I am investigating. I need also to be aware that I am <u>in</u> some thing, being, process, greater than myself. We are always in the middle, in between. We are wholenesses within a greater whole. We are not split mind-bodies. We are spiritual (which can be taken as 'awareness based'), corporeal beings.

To act in accordance with the awareness of our 'middle' position implies attempting to develop self-awareness in the very act of working mathematically and/or scientifically. So what kind of maths and science might correspond to the awareness that we are in and of this world, not separate, objective observers? Our awareness of the world is not just head awareness, it is a body, sense awareness and a heart awareness. We feel the world with our hearts.

God be in my head, and in my understanding;
God be in mine eyes, and in my looking;
God be in my mouth, and in my speaking;
God be in my heart, and in my thinking;
God be at mine end, and at my departing.

The 16th century Sarum Primer prayer articulates beautifully the scholastic stance to knowledge. 'God' here (for the agnostics, or (God forbid) the atheists) can be taken to mean Bohmian wholeness, or a state of oneness with the whole. '*God be in my head and in my understanding*': 'standing under' is a receptive stance; whereas in '*God be in my heart and in my thinking*', thinking is active. The prayer (very interestingly) states the desire for our active thinking to be based in our hearts, in oneness, and for our heads to understand, to be receptive, in oneness.

Stephen Harrod Buhner, in 'The Secret Teachings of Plants' describes in detail the physiological events involved in heart-brain entrainment. For example, "*Sympathetic and parasympathetic nerve pathways and the baroreceptor system directly link the heart and brain, allowing information and communications to flow freely. Messages from the heart to the brain during this shift to coherence, significantly alter the brain's functioning, especially in the cortex, which profoundly affects perception and learning.*"

This is an empathic form of thinking. When we start down this path, we allow ourselves to open to the deeply different wonders of so many worlds within this world. Barbara Mclintock who won the Nobel Prize for her work on transposons and corn genetics, said that it was the plants who told her what to do. Temple Grandin is a Professor of

Animal Science. She considers that it is her autism, which allows her to empathise with animals.

This is very different from the materialist stance from Cartesianism, strengthened by the power of Leibniz' and Newton's subsequent differential calculus to describe mechanical change. Ironically being part of a greater whole was a reality for Newton. The greater whole was God. Newton, like Paracelsus and Tycho Brahe, was an alchemist. He needed to keep this hidden for fear of religious oppression. Alchemy was a scientific practice albeit a very different form from present mainstream. Goethe developed his own different form of science which recognised the importance of the witness position, but this was largely overlooked in the mainstream thrust of materialist science from the 18th century on.

By the beginning of the 20th century, mechanical, materialist science was reaching some limits. The emergence of quantum physics could be seen as Nature suggesting that it was time for humanity to develop some self-awareness in its investigative approach, politely pointing out that at some levels, observation is an intervention, a sharp prod even.

We need to take into account limits of different ways of thinking. Aristotle was clear when he first enunciated deductive logic, that it was only true within limits. We've forgotten this. We've not asked what are the areas of validity of our different mathematical ideas. Instead we've been attempting to scientise all areas of human inquiry by importing quantitative maths, as if this will somehow make them more valid.

It may be time to attempt to develop various theories of different wholenesses, rather than <u>one</u> theory of everything. Natural philosophy as a philosophy of nature would include a philosophy of life, of living beings. Wholeness is one of the key attributes of living beings, going down to the level of cells. Materialist scientists have ignored the importance of wholeness, believing they could discover the nature of matter by breaking it up. Many people have pointed out the weakness of this approach: David Bohm simply states that *"fragments are not the same as parts"*.

I've characterised science as asking what is the reality of the world around us, Is there a corresponding question which maths is asking? Nowadays there are many different views of what maths is. Removed from its esoteric origins, the external part became such an increasingly powerful activity that under the Baconian paradigm its ontology was scarcely questioned. Nowadays maths is generally defined by extension rather than by essence. But in such definitions usually include, a reference not only to number, space, pattern, but also to truth. In fact in the absence of a belief in God, mathematical truth became a kind of godhead, a refuge, the locus of permanent certainty.

In the late 19th century, doubts arose about maths as the home and guarantor of truth and certainty for science, as a result of strange things that had begun to emerge in maths, non-Euclidean geometries and the like. This generated a concern to establish mathematical foundations.

Bishop Berkeley and others had, much earlier, raised concerns about the lack of philosophy and ontology in the development of maths, but Girard voiced the prevalent, pragmatic mood of mathematicians and scientists, when he stated that he used irrational numbers even though they were meaningless, because they were useful. Centuries later, the Copenhagen interpretation of quantum physics was similar: in layman's terms, 'We don't know what it means, but these are the equations that work'.

A lot of work went into mathematical foundations and a lot came out of it - Russell and Whitehead's Principia, and Wittgenstein's, set theory et al. But eventually Gödel proved that Hilbert's dream of establishing the completeness and internal consistency of maths within maths was impossible. Foundations are a building metaphor. We do not build houses on themselves. We always need some agreed *koina*. What was important in this foundational work was what came out of it – not an increase in certainty but a deepening and broadening of ideas.

Between the 13th century and the 18th century in Europe, much larger scale human institutions came into existence. In the political sphere nation states, organised around capitalist economies, replaced feudal systems. Hierarchical social organisation remained the norm; similarly in the realm of ideas. It is only in the late 20th century that the idea of the possibility of non-hierarchical organisation has arisen.

In the hierarchy of present day materialist maths and science, maths and physics have been the rulers. But this mathematics is only the quantitative part of the original Greek discipline, '*logistike*', the discipline for dealing with number as quantity or '*hule*'. '*Arithmetike*' was the branch concerned with the '*eidos*' or form, or qualitative aspect, of numbers: part number theory, part numerology and aspects bridging these. (Interestingly Vanessa Hill's recent Pythagorean 'Nature's Code' is rooted in biology)

It is not a simple matter to return to the Pythagorean oeuvres to rediscover their wholeness of mathematics. We do not have any extant Pythagorean texts because it was a mystery school. But also our cultural context is so different that it is extremely difficult to comprehend what a fuller import of that mathematics might have been.

Vedic maths has become known in the West recently as a very powerful method for mental calculation. This is because it works more in the way our minds work with numbers, i.e. heterogeneously, not just following one method for all numbers. It uses the fact that we have a number system base 10. Vedic maths is effective not only in the path of knowledge of the external world, but also psychologically, spiritually. Unlike the Pythagorean case, Vedic texts are still extant, which gives us the possibility of investigating these mathematica for their wider and deeper meanings.

The Pythagoreans insisted on the difference between one and oneness. Oneness is a wholeness and the distinction between one and oneness is at the core of holistic mathematics. Furthermore oneness and no-thingness are one and the same.

In incarnate reality we can only apprehend a nuance of oneness, sometimes in approaching sleep or through meditation Experienced practitioners may reach nirvana, but that's not oneness. To get a sense of oneness we may begin by sitting in a relaxed, upright posture and closing our eyes (when our eyes are open, visual information is generally responsible for approximately 70% of brain activity). We are advised to allow thoughts to flow through our minds, not to fight them, but simply parenthesise them with the thought– "That's a thought". We then focus on sound, and attempt to listen without naming (or judging). This is an extremely interesting, challenging exercise, designed to bring us more into now. The approach to now is through infinitesimals, moments, discrete, non-zero infinitesimals.

When we listen to music or language our minds perform time packaging. Each phoneme, for example, 'p', is a sound shape, a shape in time. Words are shapes of shapes. When we learn to speak however, we understand words, the meaningful wholes, before we grasp the notion of individual phonemes. We approach these much later when we learn to write and read. Similarly with music, we first hear melodies (whole musical forms) and only later can we separate out the individual musical (pure) tones. We humanoids appear to be hard-wired for wholeness, meaningful wholes.

Bohm's implicate order is one take on oneness, the great wholeness. In oneness, there are no distinctions, so no things. As soon as there is a one, then there is the other, the complement, which makes 2 things, and the distinction between them, which is a 3rd thing: they are heterogeneous, of very different natures, but one can not exist alone. Interestingly Peano's set theoretical derivation of the integers, which was intended as a simpler foundation, involves summing heterogeneities.

No-thingness, the full emptiness, is a fundamental concept in Eastern thought, Vedanta, Buddhism, Taoism, but not in western monotheist thought. The confluence of Hindu and Arabic thought whereby the Hindu symbol for zero, together with the Arabic numerals, gave rise to our present number place system, was key to the 16th and 17th century developments in maths and hence the technoscientific (aka industrial) revolution.

In Europe before that, they used the hugely cumbersome Roman numeral system and it was a considerable mathematical endeavour to calculate when Easter should fall (being a lunar festival, unlike the solar, fixed Christmas). The symbol for zero was a technical import which galvanised mathematical progress. But the Eastern philosophical notions did not accompany it. So the huge, mathematical leaps forward were seen as alien, which appears in the names, 'negative' and 'imaginary' numbers, as well as the fuzzy thinking around infinitesimals in differential calculus.

There has also been a long-term confusion about one and oneness, which has been ignored, because wholeness has not been of interest in the philosophy of maths since ancient Greek times. Hubristic positivism has been the main stance towards ancient Greek views, i.e. 'now we understand what they didn't understand'. This is to overlook the fact that when we learn something new, we mostly forget what it was like not to know it. We do not remember how it was not to speak, not to see physical objects etc.

It is much more difficult to try to get inside earlier or different mindspaces. How did they understand? And given how they understood, what did they see? What was that reality like – qualitatively? What might we learn from them?

In the meantime when we just accept ones (but consciously), accept multiplicity, which is where we mostly live and love and have our being. Then when we start adding all these homogeneous ones, 1+1+1+ ... we get the natural numbers, 1,2,3,4,5,6,7, ..., and we discover all this extraordinarily beautiful patterning – odds, evens, perfects, triangles, squares, Pythagoras, Fermat and hey presto!, the primes even appear to be connected with quantum reality via the Riemann hypothesis.

Is mathematics Rilke's unicorn in the Sonnets to Orpheus?

This is the creature there has never been.
They never knew it, and yet, none the less,
They loved the way it moved, its suppleness.
Its neck, its very gaze, mild and serene.

Not there, because they loved it, it behaved
As though it were. They always left some space.

And in that clear unpeopled space they saved,
It lightly reared its head, with scarce a trace

Of not being there. They fed it, not with corn,
But only with the possibility
Of being. And that was able to confer

Such strength, its brow put forth a horn. One horn.
Whitely it stole up to a maid, to be
Within the silver mirror and in her. (transl. J.B.Leishman)

Rilke contributes here to the debate about the nature of mathematical existence. The unicorn enters the mirror of self-reflection and the anima, the soul; this offers the possibility of self-awareness. 'Poem' and 'fact' have similar etymological roots: 'from the Greek '*poiein*', 'to make'; 'fact' from the Latin '*facere*', 'to make' and 'to do'.

Praise be for this beautiful, mathematical world, this unicorn. An interesting question arises: to what extent can we be aware of the unicorn and aware in our sensory, personal being in the world? For example, if I count potatoes to make sure I am cooking the right amount for the guests at a meal, how aware am I of the individual natures of each potato? This is an important question for the development of human wholeness.

Abstract thought is the product of head thinking. It is immensely valuable and powerful, and potentially dangerous. Logic is amoral.

We in the Western world have grown up in a society permeated by materialist ideology. Even if we had a spiritual and/or religious upbringing, materialism and rectilinear *logistike* lives in our architecture, our technology, the entire social fabric of our lives. Our lives are ruled by quantitative or meaningless numbers, National Health numbers, National Insurance numbers, passport numbers, credit ratings. From birth we are taught to see and name physical objects, so we forget how to see other things. We are trained to orient our senses in specific ways outside; we are discouraged from developing inner awareness. The subject-predicate form of Western language structures our head thinking. So how can we begin to develop other ways of thinking? Can heart thinking and body awareness help?

Stephen Harrod Buhner, having developed heart thinking in his herbalist studies, is sceptical about the possibility of heart thinking in mathematics. I disagree: I think that many, if not most significant mathematical advances have probably come as a result of mathematical heart thinking, from opening the heart, from love.

When we dream we allow our hearts to cocreate with our heads and pure mathematicians have been allowed to dream more freely than scientists. Scientists dream as well, Kerkule's discovery of benzene rings is the most famous example. Still there are more constrictions on science in daylight thinking. Materialist forms have dominated science for 3 centuries now. Someone said recently that, after Descartes scientists thought of the workings of the world as complicated clockwork, and that the teachings of quantum physics suggest we need to think of the world as mind. The implications of this are enormous, but this change in attitude has not become mainstream.

When we open our hearts, we allow space for the realities of other beings; we enter conversations, rather than interrogate. It's part of interview technique to ask open-ended questions. What?, how?, why?, encourage people to talk about themselves, whereas when?, where?, how much?, etc. are closing questions: you get one word answers and you don't get very deep. At first it looks as if science has managed to get pretty deep with its persistent, closing approach. It's certainly managed to achieve a lot, from space travel to nanotechnology. But human beings who live in and with and under this technoscience continue to make war on each other and other Gaian species, with increasingly effective technoscientific weapons. The arms industry, the herbicide and pesticide industries are all killers.

As David Bohm says, "*Nature will respond in accordance with the theory with which it is approached*". Climate change is a result of 3 centuries of approaching her with a mechanist, materialist theory. The time has come to approach Nature, a greater whole of which we are part, with respect, to be interested in her being, not what we can make her do for us.

In craniosacral therapy, we adopt a witness awareness, a listening stance. Our bodies are universes of billions of intelligent cells, organised in intelligent communities, tissues, organs. The essence of craniosacral therapy is to do as little as possible, not to try to put things right, rather just to be present as simply and deeply as possible, to witness, to listen to the body's stories. I've sometimes called it psychotherapy for the body. I attempt to allow my attention to be directed by the intelligence of the client's body; in this way it can use the energy of my attention (homoeopathic input) to do what it wants to do. Startlingly deep healing can occur. If there is a scientific explanation of this phenomenon, it would surely be at a quantum physical level where an observation is an intervention. Attention is our finest energy.

In complementary approaches to healing we are concerned with the specifics of cases more than general theories, with qualitative assessments more than quantitative measurements. Western medicine like other forms of Western science can do amazing things, but the cost is huge. The pharmaceutical industry whose stated aim is to cure human ills, is the 4th biggest killer of people in the US and UK and the side effects of these drugs on Gaia's flora and fauna, through effluents, is also lethal.

The overarching paradigm of complementary medicine is wholeness: this is the etymological root of 'healing'. The greater wholeness contains both polarities, specificity: generality and quality: quantity. Here we are coming to the advaitist core of the Bohmian stance: advaita is at the heart of Vedic philosophy; literally it means non-duality.

Our culture has elevated the general over the specific and quantity over quality. This has been at a great cost to our individual lives. Alienation, the Western 20th century malaise, is not feeling at home in ourselves, in our inner lives, in our bodies. This is not surprising when we've been brought up to believe that only what is outside and measurable, is real.

We see the cultural dominance of the external over the internal, in that, when we speak of 'the 5 senses', we mean only the 5 externally oriented senses and we ignore our internal senses, including proprioception, our most fundamental sense, which gives us our primary experience of wholeness. Body awareness includes the external and internal senses.

This dominance of the external and general goes hand in hand with a dominance of the visual sense over the others. Vision sets the bar for measure of the external world; it allows us to measure continua. Hearing gives a different, discrete kind of measure. It relates to counting, beginning with the rhythm of now, now, now, ...

The ancient Greeks separated geometry, dealing with continuous lengths, from *arithmetike*, dealing with whole numbers. Geometry corresponded to visual space; *arithmetike*, to music, 1:2, 2:3, 3:4, being fundamental tonal relationships. Centuries later when *logistike*, quantitative mathematics, came to rule supreme, the two were conflated and there was no philosophical questioning of the meaning implications.

For Pythagoras and Plato, wholeness was paramount: numbers were whole numbers; lines were lengths which could be measured practically as fractions, but were considered by mathematicians as ratios. There was a beautiful wholeness to the system including, of course, the harmonic musical ratios. The square root of 2 did not fit into this system, and the wholeness was shattered. It was over 2,000 years before the ideal of wholeness again came into mathematical purview in Hilbert's programme, only to be broken by Gödel's work. (I've written about this in more detail in '*Who Carved up the Integers? They Never Died*'.)

Our current maths and science is based on the visual sense, in that it is primarily quantitative, and on the Cartesian, real number line, which emerged from *logistike*. This maths and science is powerful in the external world, and beautiful. Are there ways that we can reclaim the human, sensory connection? Michael Atiyah has suggested that algebra relates more to hearing, as it relates to time and rhythm, more than space and lines. Topological ideas have connections with touch. How might we begin to develop maths relating to other senses? What might it be like?

We are corporeal, sensory, spiritual processes in continual interaction with each other and our worlds, with nature. We are fluid, changing interfaces. Present day science still views humanoids as intelligent, biological machines. To develop maths and science in keeping with the spiritual, organic beings that we are, we might consult Vedic wisdom. There the senses appear as dyads, organs of reception correspond to organs of expression or action. Some seem straightforward to us, for example, hearing and speech, others less so, for example, seeing and moving.

We can also return to proprioception as our key starting place. We find ourselves, in the oneness of our individual bodies, in the middle of a multiplicity of heterogeneous sense information. What would this mean to begin our numbering here? We could also note our human form which has 6 orientations rather than 3 dimensions: left – right, up – down, front – back, are not simple opposites; they have qualitative differences.

If we begin to reclaim mathematics as a <u>human</u> <u>activity</u> to help us orient ourselves in the world, many more things are possible and many more things begin to make sense. Our 5 external senses are described in the Vedas as organs of reception. We might even begin by investigating the organ of expression which corresponds to seeing, namely movement. Traditional Indian dance is highly mathematical.

If we also want to follow the path of self-awareness in maths, we could think of movement in its meaningful aspect, as gesture. In this way we might begin to translate the beautiful abstractions of the vast architecture of existing mathematics into human life experience. Rudolf Laban in the 20th century developed a human movement analysis related to the cube. This is an extraordinarily complete explication of human movement and effort – gesture.

Counting is a gesture; it is a rhythm, a beat – now, now, now, ... The rhythm of our hearts begins when we are embryos, in the circulation of blood cells. The heart comes into being from the flow of blood: the flow creates the structure. Measuring length is a gesture, a different gesture, one of extension, a continuity stretching from here to there, from here to here.

Jaap van der Wal, has pioneered the investigation of gesture in our embryonic phase. This is one of immense becoming – huge flux, structures forming, being destroyed, new ones coming into being. All such changes take place in the dark, in a contained space. We humanoids begin this adventure in the holy whole, the hole of the womb. Early embryonic developments all occur by first creating holes. To imagine (or remember) these changes is awesome, sensuous topology.

In recent years, there have been a number of experiments with dancing mathematics, Marcus du Sautoy, Francoise Chaitin-Chatelin as well as Indian mathematical dance. The point is not merely for dance to be a mathematical educational tool. The movement is part of the mathematics. The Fibonacci series <u>lives</u> in plants. Similarly music is a form of applied mathematics.

Maths is clearly our most abstract language, the purest metaphorical language. Unfortunately, in the absence of trust in the great wholeness of which we are part, we've been desperate for some kind of certainty and have got very hung up on mathematical proof. It's important, but as Imre Lakatos pointed out, the real motivation of attempting to prove theorems, is not to establish ultimate certainty, but rather to reveal ideas hidden in our initial *koina* and thus grow mathematics.

To do mathematics, create mathematics <u>is</u> to discover and abstract pattern, order, structure, processes from phenomena and discover the relationships between these abstractions. This is a form of meaning making. Meaning is a process. In mathematics, we define terms using the *koina* and return to excavate for deeper understandings within those *koina*. This is not just mathematical development but an inherent part of human growth. As Viktor Frankl writes, making meaning is an essential human characteristic.

We probably need more praxis in these initial ways before we can really begin to imagine how a maths that recognises equality between beings, or a non subject-predicate based maths might show up. We need to find ways to unlearn as well as to learn (thank you, Lewis Carroll). (The membrane of the sphere could serve as a metaphor here as well.) For example, in order to understand and manipulate numbers and other mathematical expressions (and alphabet letters), we have had to numb ourselves to their reality as symbols. Marie-Louise von Franz devotes nearly a whole book, 'Number and Time', to the symbolic meanings of the numbers 1 to 5. Again the question arises as it did with counting potatoes, of the possibility of dual awareness. The polarities are abstract and general complementing sensory/ intuitive and specific.

The possible directions I am suggesting for maths and science involve radical change, but they are really only a continuation of the Pythagorean mathematical telos and the Goethean approach to science. The changes have already begun. From early in the 20th century quantum physics and Gödel's incompleteness theorems demanded shifts in the philosophy of maths and science away from materialist positivism. As C.S.Lewis said, *"We all want progress, but if you're on the wrong road, progress means doing an about-turn and walking back to the right road; in that case, the man who turns back soonest is the most progressive."*

When we allow ourselves to recognise the validity of the many different wholenesses in our world, to recognise that there are different forms of relationship within them and between them, then we open to a greater plurality of mathematics. We can begin to look at the different forms they might take – different theories for different wholenesses - wholenesses of cells, of organs, of tissues, of organisations, of minds, of processes, of buildings, of stories, of plants, of animals. We can look at the maths of the in-between spaces, even between specific and general.

We live in interesting times (as the Chinese might say). There is huge flux. A few years ago the number of people on the planet reached the same number as cells in a human brain. It's time to review our deeply held cultural ideas, and be open to vision, intuit, what ideas need to come in, to grow, if we humanoids are to develop or even survive. Our mathematical ideas and ideas about maths are important ones to reconsider.

Greek mathematics was a maths of stasis. Plato's reality was a timeless space outside of consensus reality. Their mathematical discoveries (pre-irrationals) corroborated this thesis. They developed a calculus, but an integral calculus, one of summation, not of change.

The Renaissance, then the Enlightenment, brought a very different world view. The differential calculus is the pinnacle of the huge mathematical advances of that time: the introduction of a sign for zero, the Hindu-Arabic decimal number system, the idea of the number line, the Cartesian grid, and equations with variables. This calculus allowed mathematical description of change. This was world changing. The following three centuries were dominated by an unfolding of, actualisation of, those ideas.

V.I.Arnold has characterised this maths as "*divided into three parts: cryptography (paid for by CIA, KGB and the like), hydrodynamics (supported by manufacturers of atomic submarines) and celestial mechanics (funded by military and other institutions dealing with missiles, such as NASA)*". This was not wholly tongue in cheek.

The change that was mathematised in the differential calculus, was quantitative change between measurable variables. It modelled mechanical change, not qualitative change or emergence. It was only in the 20^{th} century that the latter idea appeared in systems (aka complexity) theory. The possibility of non-hierarchical organisation was a necessary precursor.

Mae Wan Ho discovered a living example of this when she looked through an electron microscope at a worm and dancing rainbows. They subsequently discovered that these were caused by the harmonious quantum activity between all the living tissue of the worm. These are individually free, yet harmoniously related activities at multiple levels. She calls it a quantum jazz ballet.

It would be a worthy goal for present day maths and science now to attempt to envision such a possibility at our macroscopic level: individual freedom combined with harmonious relationships with other beings, beings around us, within us and the greater wholenesses within which we have our being.

We are now in a position to begin to contemplate other kinds of mathematics, mathematics of becoming and relating, mathematics whose telos is to seek the meaning of the whole wherever there is duality, and to discover and create beauty and harmony.

A project for maths and science now: to research the nature of being, becoming human in this richly beautiful, conscious universe.

The Process Category of Reality

Michael HEATHER & Nick ROSSITER
Northumbria University, Newcastle NE2 1XE, UK,
michael.heather@trinity.cantab.net; nick.rossiter1@btinternet.com
http://www.computing.unn.ac.uk/staff/CGNR1/

Extended Abstract

The Alternative Natural Philosophy Association (ANPA) has no preferred foundations nor methods. Rather a prime aim as its name indicates is to consider scientific alternatives to the mainstream. Nevertheless a number of its members in quite diverse applications seem to subscribe to some version of PROCESS. This seems more to arise by default rather than by any deliberate policy or even concerted effort to concentrate on the same paradigm. Incremental advances in interpolated knowledge promote a conservative mainstream with less scope to explore alternative methods. Studying the frontiers of knowledge on the other hand is more concerned with extrapolation from the known to the unknown and encourages the exploration of alternatives. For similar reasons also many of the substantive topics of interest to ANPA members tend to be fundamentals at the frontiers rather than incremental advances within existing knowledge.

Process is concerned on the one hand with the global/local distinction and on the other hand by the juxtapositioning between the stationary and the non-stationary. The universal and the particular, the static and the dynamic are

integrable at the level of metaphysical reality and are formally representable and accessible in natural categories that follow physics. Process relates not just to the non-stationary but subsumes both the static and the dynamic. One is contained in the other but which way round? Such problems like Zeno's paradox of the arrow's dynamic flight consisting of only static positions are avoided in the 17th century French logic school of the Port Royal, following from Aristotle's first and second intentions, as distinguishing the intension from the extension. Information systems are an important example of the need to represent the world by computer. The limitations of modelling reality are evident in database methods like ACID and in Codd's pure relational model.

Natural categories as a metaphysics provide mixed levels for intension/extensions. Intension and extension alternate in a preorder, that is with arbitrary beginning. This is the natural role of the categoreal arrow with an identity arrow as intension and a distinguishable valued arrow for extension. The simplest identity arrow is treated as an object, the next higher identity arrow (the functor) composed of extensional arrows between objects makes a category with ordinary functors as extensional arrows between categories. The next higher level is a category of these categories with objects as categories and functors between them. The highest level arrow is again an identity natural transformation which composes the previous level of categories as objects with natural transformations between them. It is this final identity natural transformation that constitutes a topos. The whole is just a recursive system with closure at four levels consisting of three open interfaces. The identity natural transformation is scale invariant and any higher level would only be a reformulation of this same level. This is process and the Universe is an instantiation of process but the World is even greater than the physical Universe consisting of all the relations between physical entities and all the relations between those relations. The interaction of subjective human behaviour in the global world of physics, biology and economics is very topical.

Relationships in nature are explicable in natural categories with the single concept of adjointness that consists only of a pair of contravariant arrows inducing a monad. Mathematical categories other than the Cartesian closed are possible but natural categories being derived from physics only recognise the existence of Cartesian closed categories where every object is the domain of a covariant arrow and the co-domain of a contravariant arrow. This recursive structure applies at any of three possible levels and in general occurs between a pair of categories where adjointness of the pair of arrows contravariant to one another between the categories induces a monad (or 'triple') and a co-monad (or 'co-triple'). Each arrow has a dual role. One is the contingent arrow of intension (in the monad) and the determinant arrow of extension (in the co-monad) while the other arrow is the contingent arrow of extension (in the co-monad) and the determinant arrow of intension (in the monad). Finitary mathematics seeks to model various features of the adjointness between natural categories in a range of topics, including currying, lambda calculus, the Yoneda lemma and its embedding. From the ANPA's long-term perspective, however the four level Combinatorial Hierarchy based on the Frederick Construction is a binary model of the four-level metaphysical preorder of Process presented here.

1 Process

Although its roots go back a further twenty years, the Alternative Natural Philosophy Association (ANPA) has from its first meeting in 1979 given preference to rigorous formal argument. 'Disciplined thought' is essential with alternative methods. Without pre-existing agreed defined terms, ANPA would otherwise have no *modus operandi*. For there has to be some common ground of reasoning. Process seems to arise naturally as both a consequence and a catalyst in the ANPA context. A continuing example is the process basis for the fine structure constant [5].

A fundamental structural significance in the world is the way the local connects into the global such as in the McLuhan Global Village where everything is connected [20]. The temporal analysis is the distinction between stationary and the non-stationary. Philosophically this global/local distinction is not at all new. It is at the root of Zeno's paradox of the arrow's dynamic flight consisting only of static positions.

The noun 'process' or the participle 'processing' commonly describe an act of transforming an existing object by some procedure to another form as in a manufacturing or business administration procedure. Wikipedia deals with its entry for PROCESS in up to 40 different fields of knowledge, including philosophy, science, engineering, computing, chemistry, biology, law, business and even the 'process haircut' [32]. There are variations in the meaning of the word depending on context. For instance in business, process describes activities or tasks that produce a specific service or product for customers. Interestingly Wikipedia does not include physics in its lists of fields for process.

The whole subject of cybernetics can be viewed as a process operating in nature as in Wiener's definition [30] involving comparison of communication in the animal and the machine [24]. Process describes the way that both animals (biological systems) and machines (non-biological or "artificial" systems) can operate according to cybernetic principles. This was an explicit recognition that both living and non-living systems can have purpose. Wiener considered that systems theory seeks to deal with the local/global divide [24], treating systems as equivalent to process but the latter is the higher form. The early specification of the working of the brain in cybernetics by Ashby [2] amounts to the concept of process but it was von Bertalanffy of the early founders of cybernetics that explicitly related the latter to process [6, 7].

Most writers trace process to the 'all is flux' of Heraclites in contrast to Parmenides, who is more usually associated with a static view. However, process is more than flux and also subsumes permanence. It is rather the Heraclites' *logos* which was taken up by the Greeks of Alexandria and the Judeo-Christian tradition to identify *logos* with God and the second person in the Trinity. The whole theory of evolution is process too but one where the origin of species does not unfold in a linear fashion. Evolution appears a foundational natural process encompassing both emergence and change. Ordering is adjointness and includes both static and dynamic aspects. It is a paradox that process includes invariance[1] which describes no change under a transformation. Indeed scale invariance turns out to be an important phenomenon of process and a relevant aspect to ANPA because of the interest in dimensionless universal constants such as the scale invariant fine struc-

[1]The subject of invariance was mainly developed in the 19th century by Arthur Cayley. Saunders Mac Lane [21] traces the early origins of category theory to Cayley.

ture constant. Fractal patterns arising from scale invariant physics are studied piecemeal with use of special sets like the Mandelbrot, Julia etc. However general methods are restricted because a set cannot be a member of itself in the way that a reflective subcategory can have itself as an object. Information systems like the web also exhibit properties of scale invariance but we do not have space here to pursue this aspect of process which arises in exponential categories.

There is always the problem of where to begin. That statement may be formally expressed as *a pre-order of categories* or just as well as *a category of preorders* for both lack beginning and ending. However within process we can but focus on the *category of reality* in the sense of the category where objects and relationships between objects exist to make up the physical world. This is metaphysical process and the Universe is an instantiation of process but the World is even greater than the physical Universe consisting of all the relations between physical entities and all the relations between those relations. Physical relations connect directly from higher-order relations. This is treated bottom-up but because of the holistic nature of process it is driven top-down. A topical example is the recent realisation of how subjective human behaviour affects the objective syntax of world economies. Current practical examples of applied recursion across levels is deduplication in structured data storage [15] or functional DNA nanostructures that can be integrated into larger structures as miniature circuit boards in bioengineering [26].

In this sense the *World* is greater than the physical Universe of cosmology. There is a unique arrow from the source of the World to every object in it and a unique resultant arrow between any pair of objects as in Figure 1.

For we are concerned with the higher order of relations between

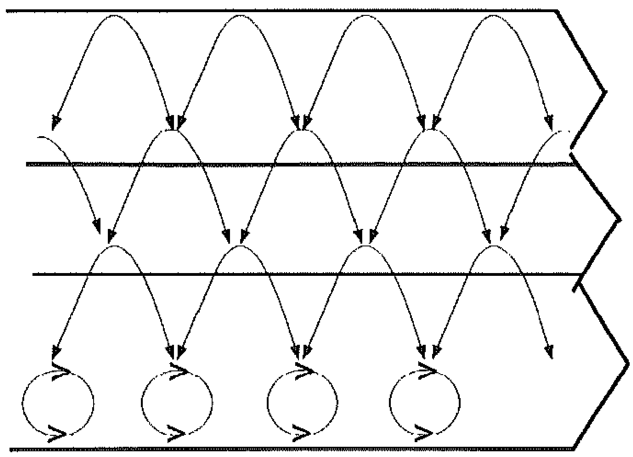


Figure 1: A Schematic World-Universe Relationship. The lowest horizontal arrow is a category consisting of a row of notional elementary objects connected by a row of vertical arrows which are themselves connected by a higher row which are in turn connected by a yet higher row.

physical objects and the relations between those relations which together with the physical objects of the Universe make up the World. This then embraces the whole of human affairs and activity including the arbitrary disciplines of philosophy and theology. Existence in categories is identifiable with the object which as we shall see is the condition known categorically as Cartesian. Ordering is adjointness and includes both static and dynamic aspects.

This empirical knowledge that every entity that exists is related to every other object that exists is no more than a definition of the Universe to include everything naturally accessible. This provides a unique direct arrow between any pair of objects that is the composition of all possible arrows between them. This is the structure given the label preorder. Figure 3 presents a two-dimensional representation for the context of the objects C and A of a preorder. There is but one unique arrow between any pair of objects in a preorder and that arrow as the figure shows is the limit of all other possible arrows whether directly between the pair or indirectly via any other pair in the preorder. We cannot assume any orientation

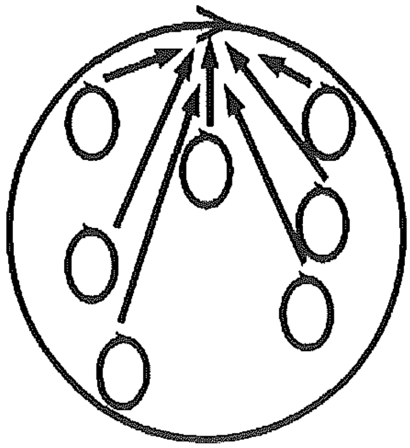


Figure 2: Terminal Object. A category with a terminal object has an arrow from every object to it. In preorders this arrow is unique as in Figure 3

for it or even presume the concept of a dimension. It is possibly easier to imagine than to draw, although our common perception wants us to imagine it in three dimensions or possibly in higher order algebraic or geometric dimensions but lies easily in higher-order geometric dimensions. A process preorder does not exist in any space whether algebraic or geometric. Rather it should be space free. This is the quantum world. However the effect between entities is mutual and the arrow is therefore two-way but not symmetrical because the opposing directions give rise to a natural parity in their mutuality. This is the ultimate reality of the quantum world. Whether it is the quantum or the physical world that is true reality seems just a matter of personal preference.

Newtonian physics treated the universe as some container either rectilinear or spherical but embedded in time. Such a structure is representable, for example by Yoneda or Curry techniques[2], to first-order as a number. This is the classical model which can be verified by measurement in first-order predicate logic because as Gödel has shown first-order predicate logic is complete for a

[2]See [3] at pages 118 and 190 respectively

closed world. However Gödel has also shown that such a logic is not complete for an open world and any model based on number and relying on axioms is not complete whether open or closed [10]. This effectively sets a limit to Wigner's 'unreasonable effectiveness of mathematics in the natural sciences' [31].

2 Metaphysics

If we want to identify a category with reality, existence requires designation of one object as the terminal object, as shown in Figure 2. This is the condition known as 'Cartesian'. It is also possible to designate another as the source of the process as initial object. This is the condition known as 'co-Cartesian' but is not a necessary and sufficient condition and may therefore result in over-specification and a too constrained system. There is a free functor mapping from the preorder on to any of its partial orders. It is natural to identify the terminal object with the covariant identity functor. If the initial object exists it would exist as the contravariant identity functor of the category. Nevertheless although these are arbitrary terms the use of the labels 'terminal' and 'initial' imports an interpretation and requires the existence of some axiom of choice, which is an axiom/assumption of set theory. The ANPA *Statement of Purpose*[3] Clause 1 states that 'The primary purpose of the Association is to consider coherent models based on a minimal number of assumptions. Here we are raising the stakes from models to metaphysics but nevertheless attempting to keep to a minimum of assumptions[4]. The *Statement* might be better expressed as 'a minimum of assumption'. Here we try to make no assumption at all beyond that the World exists. We try to keep open issues

[3]as regularly published in its Proceedings including in these proceedings for ANPA 31.

[4]There is some philosophical difficulty here with ANPA's 'minimal number of assumptions' when dealing with supposition because the number is not necessarily a measure of quantity or quality.

about terms such as 'terminal' and 'initial' because they may be related to what cosmologists currently tell us about the fabric of the physical Universe consisting mainly of dark matter and dark energy with only 4% in familiar forms.

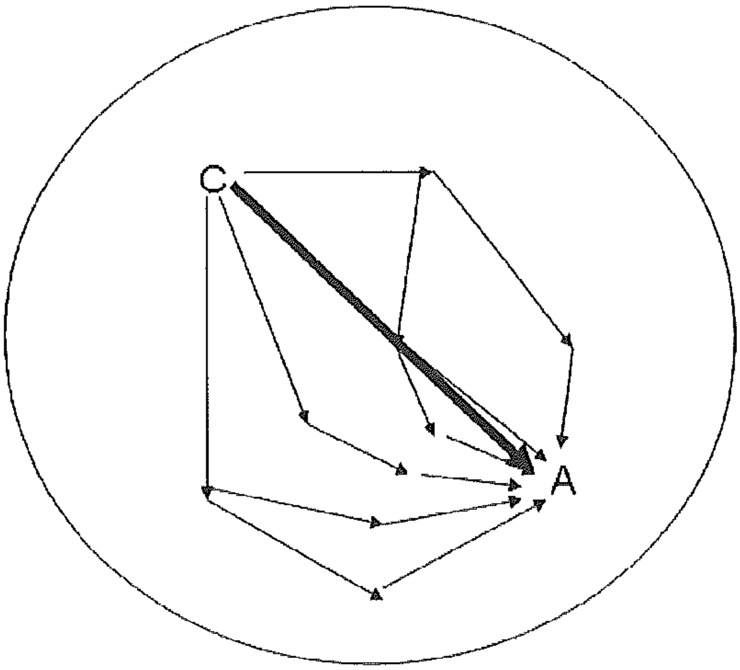


Figure 3: The unique arrow from C to A as a limit of other arrows in a unidirectional Preorder. The co-limit sums over all others

ANPA is mainly concerned with fundamentals at the frontiers rather than incremental advances within existing knowledge. But how do the general and the particular relate within the structure of the world? Any formal description needs to be able to combine both the global and the local. This is possible with natural categories by substituting metaphysical process in the interpretation of Whitehead's later *Process and Reality* [28] for that in his earlier *Principia* [27], which was the starting point for the traditional finitary category theory of Eilenberg and Mac Lane [21]. It is the difference between a metaphysics and modelling which are separated by two levels as in Figure 4 (diagram 16 in [25]). Metaphysics is one level up from reality in human perception while models are one level down. The limitations of modelling reality can be seen in information systems where there is a need to represent the world on computers. Problems are evident in database methods like ACID [11, 23] and in Codd's pure relational model [8]. In database design, data normalisation is used to attempt to match the logical data structures to the physical world. This method of design has a number of unsatisfactory features. Firstly it is di ffi cult to enforce the laws of the physical world in the operational database and secondly the theoretical underpinning, based on set theory, is not natural because of the problem of representing arrows across multiple levels as functions.

3 Finitary categories model natural (metaphysical) categories

Whitehead developed his theory *Process and Reality* in what he terms speculative metaphysical categories These are in great contrast to the formal principles he enunciated with Bertrand Russell in *Principia Mathematica* and Whitehead devotes Chapter 1 of his later work ([28] pp.4-26; [29] pp.3-17) to explaining in a general philosophical context why they had to be speculative. For the second half of the last century has seen substantial advances in the development within finitary mathematics of formal categories based on the concept of the arrow and initiated by Eilenberg and Mac Lane [21]. The phrase 'finitary mathematics' is a term first coined by the mathematician David Hilbert[5] and effectively describes the whole mainstream of twentieth century mathematics built up on a system of proofs in set theory and number from incompletely specified axioms. The adjective finitary is itself a little misleading as finitary mathematics includes topics like infinity and transfinite numbers as these are modelled on the finite concept of

[5] according to Feferman [10] Hilbert never defined finitary mathematics and it collapsed at its foundations under the weight of Gödel for the reasons mentioned above.

number.

Nevertheless it is possible to ascend the staircase in Figure 4 from categories as mathematical models to metaphysical categories and extend that ladder to categories that are no longer speculative but which can now be made formal thanks to the work of Eilenberg, Mac Lane and a large number of pure mathematicians world-wide who have refined and extended their original interpretation of the humble arrow based only on the four properties:

1. a morphism from domain to co-domain
2. identity from an indistinguishable domain and co-domain
3. associativity
4. composition,

There are two distinctions important for process that we need to draw. One is between metaphysical categories and finitary categories in respect of the use of number in physics; the other is between sets and either types of categories in respect of the representation of intension and extension. We will first consider the natural numbers then look at intension and extension as an intrinsic property of parity to be found in adjointness.

4 The finitary category of the natural number

Because it relies heavily on experiment, physics as a discipline has become identified with measurement and number as its prime conceptual tool. Consequently it has become very bound up with sets which equate to number. However it is an assumption that qualities and quantities are representable as number. The physics and the mathematics have become merged so that space is a complex

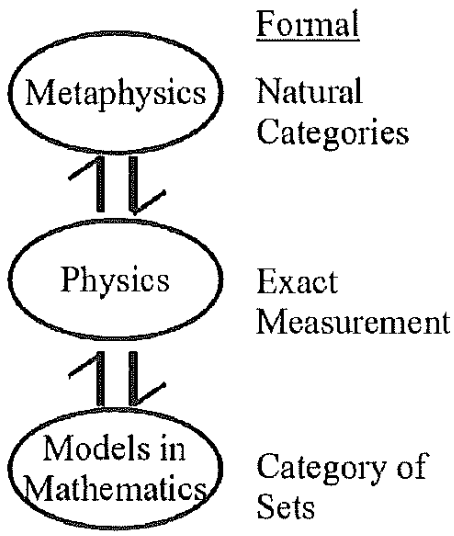


Figure 4: The Staircase of the World from Metaphysics to Models

number whether it is Newton's Universe as a container or the infinite Hilbert space of quantum mechanics. These it should not be forgotten are just numbers. This is fine to the extent of first order models for which Gödel (as mentioned above) has shown to be consistent but it is not sufficient for open or other higher order systems for Gödel has shown these to be neither consistent nor decidable when relying on axioms of sets or number. This applies as much to the use of statistical methods as the interpretation of measurement. It may be possible to reduce any problem to first order but any conclusions will then be subject to the assumptions in the reduction. This is particularly insidious in treating open systems as closed. However openness is not just bound up with the concept of order for it contains a deeper logical strand of constructivism as associated with the intuitionism of Brouwer. Boolean logic suffices for a closed system but an open system requires the logic of Heyting (See Figure 9 below).

Metaphysical categories have therefore no natural concept of number. Finitary categories as a model relying on the concept of sets has consequently to introduce the concept of number[6]. This is achieved by postulating a Natural Number Object with a recursive definition on arrows comparable to recursive functions generating the set of natural numbers. This requires importing some undefined successor function. While this may be natural in mathematics it is not natural in physics where systems are open either externally or internally. An obvious example is radioactivity where atoms decay according to some preorder and it is not therefore possible to identify a successor before the event of decay. Of course it was to explain such events that the notion of randomness was invented but this is normally dealt with by some theory of statistical probability which leads back to the concept of number and does not provide an exact solution. This lack of a predefined successor is a feature of all open systems and a chief cause of problems of interoperability in global systems.

Open physics lacks a concept of number and this questions the use of finitary models in physics. The existence of multiverses must surely be the largest incarnation of the number concept. The Panel 1 lists nine current possible theories recently identified by Greene [12]. These can also be compared with Barrow's views of multiverses [4].

It is instructive to review Greene's list from the process perspective. The list does not claim to be exhaustive and is an example of undecidability demonstrating how the use of number leads to degeneracy with many possible forms. This degeneracy is well borne out in the thorough examination of n-categories carried out by Leinster [19]. It may well be a comparable defect in string theory that allows variations in physical laws. In process categories physical laws arise from properties of adjointness whose *bonum esse* is

[6]First carried out by Lawvere [17] and now to be found in standard category theory texts, such as [3] at p 177.

Panel 1 : MULTIVERSES – Current Possible Multiverses recently identified by Greene [12]

1. Infinite space may contain a number (possibly an infinity) of universes that may lie beyond our sight.
2. Uncountable other universes with different characteristics may have been created with ours during a fleeting period of superfast accelerating expansion.
3. String theory suggests our universe is one of many 4-dimensional *'brane worlds* floating in a higher-dimensional space-time.
4. A simple cycle of universes with variations in physical laws as possible in string theory.
5. More complex versions of cyclic universes.
6. Quantum mechanics allows/requires many worlds to exist in parallel formed by a branching of the wave function.
7. the universe is a holographic projection.
8. We are just one of a set of artificial universes created in simulation on a super-advanced computer.
9. The philosophical necessity that every possible universe must be realised somewhere.

uniqueness. Furthermore about half the items in the list depend on some idea of infinity. But infinity belongs to mathematics, not to physics. It was David Hilbert the proponent of finitary mathematics who with the paradox of his Hotel Infinity recognised that infinity is always beyond reach and therefore cannot plausibly exist in physical reality. Infinity in finitary mathematics seems no more than a model of repleteness[7] under the free functor in process categories. The last item that postulates every possible universe is also derived from probability theory applied to infinity. That too fails at the Gödel hurdle of 'number'.

As anthropocentric variants on our universe with complicated

[7] Johnstone ([16] at p.3) defines the condition of repleteness as "that any object of the ambient category isomorphic to one in the subcategory is itself in the subcategory".

theories reminiscent of epicycles, multiverses bear an almost Ptolemaic resemblance. The super-advanced computer is a science-fiction vision of current commercial computers. They have not been thought through with respect to quantum computation nor any general attention paid to boundary conditions nor to the relativistic nature of time which Whitehead would carefully respect [28].

5 Logical structure of World representation as adjointness

In terms of natural categories, process is adjointness. This is the formal metaphysics of real existence such that every physical entity in the Universe affects every other. There is at the most a single pair of arrows in opposite directions between any pair of objects. These are limits of all the possible paths around the Universe between any given pair. This limit reduces to a single function as an abstraction in lambda calculus or as a resultant in vector analysis (for first order models lose the resolution of the contravariant pair). There are four levels involving three interfaces. The uppermost level is the intension and the lowest is the extension corresponding respectively to the global and the local. The intermediate interface connects intension and extension, that is snaps the local into the global for all time and space. Any set-theoretic approach finds this latter mechanism, which is essential to all studies of globalisation and interoperability, very difficult if not impossible as recognised by Russell's paradox.

Nevertheless in finitary categories the mathematics of adjointness has been developed in this concept termed a Cartesian closed category, derived as an abstraction of the Cartesian product but this description from historic origins may by its simplicity mislead as to its great power and content. The finitary approach is to distinguish the two properties of Cartesian closed and locally Cartesian closed but in process categories it is that natural distinction between intension and extension. This paper provides an introduction to that formal description of the mathematical structure of the World as found in nature.

To the global/local distinction must be added the stationary against the non-stationary. Both the static and the dynamic are formally representable and accessible in the logic of natural categories. Process relates not just to the non-stationary but subsumes both the static and the dynamic. One is contained in the other but which way round? Such problems, like Zeno's paradox of the arrow's dynamic flight consisting of only static positions are avoided in the 17th century French logic school of the Port Royal [1] (harking back to Aristotle's first and second intentions) by distinguishing the intension from the extension. Aristotle referred to them as first and second intentions. Because of their extended meaning these terms were recognised in the subject of logic by retaining the older spelling with an "s" rather than a "t". When the old subject of logic was superseded around 1900 by symbolic logic based on set theory, the intension/extension relationship became rather lost until the development of computer programming revived it with the need for rigorous typing.

The intension-extension relationship is recursive; thus in the diagram of Figure 4 metaphysics is the intension for reality as its extension and reality itself becomes the intension for models as possible extensions. In the natural categories of metaphysics process is adjointness. This is no more than the formal metaphysics of real existence that every physical entity in the Universe affects every other. There is at the most a single pair of arrows in opposite directions between any pair of objects. These are limits of all the possible paths around the Universe between any given pair. This limit is that of the preorder in Figure 3. Mathematical categories other than the Cartesian closed are possible but process categories being derived from physics only recognise the existence of Cartesian closed categories which has the property of adjointness. Every object is the domain of a covariant arrow and the co-domain of a contravariant arrow. This recursive structure of intension/extension applies at any level but is best studied between a pair of categories (identity functors $\mathbf{1}_F$ and $\mathbf{1}_G$) where adjointness of the pair of arrows (F and G, contravariant to one another) induce a monad consisting of a triple $< T, \eta\ , \mu >$and a co-monad consisting of the co-triple $< S, \epsilon\ , \delta >$Figure 5 shows the adjointness between the categories, intension and extension.

Each arrow has a dual role. F is the contingent arrow of intension and the determinant arrow of extension while G is the contingent arrow of extension and the determinant arrow of intension. T is just the composition GF and S the composition FG. Each of these compositions may be compared in Figure 6 at the next level up with the contribution they make to their respective identity functors by means of the creative unit of adjunction $\eta : \mathbf{1}_F \longrightarrow GF$; and the qualitative co-unit of adjunction $\epsilon : FG \longrightarrow \mathbf{1}_G$. Comparision at the even higher level of order is provided by the unit of potentiality $\mu : T^2 \longrightarrow T$; and its co-unit $\delta : S \longrightarrow S^2$. There are special cases of the latter two which may be interpreted [25] as in the 'dimension of time' with the unit of anticipation where potentiality is by hindsight and the co-unit of anticipation by foresight. Although there are never more than two basic adjoint functors $F \dashv G$, the combined composition of their two compositions T and S may be resolved into the three basic functors of Figure 7 to be found in standard category theory texts, where$\Sigma$is the existential qualifier,$\Pi$the universal quantifier and

Δ the stability diagonal pullback functor. The interplay of left and right adjointness with left and right exactness is a little subtle [13] and can be better understood in the exploded diagram of Figure 8 which is repeated in Figure 9 to show an exploded view of the natural intuisionistic logical structure of the Cartesian closed category.

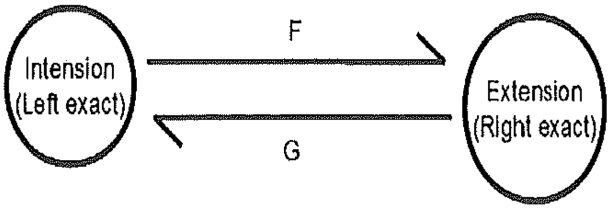


Figure 5: Adjointness $F \dashv G$

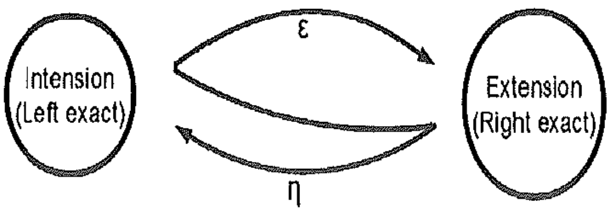


Figure 6: Adjointness expressed with natural transformations η and ϵ

6 The natural World structure as a Cartesian closed category

Relationships in nature are therefore all explicable in process categories with this single concept of adjointness [18] that consists only of a pair of contravariant arrows inducing a monad. In finitary categories the mathematics of adjointness has been developed in what

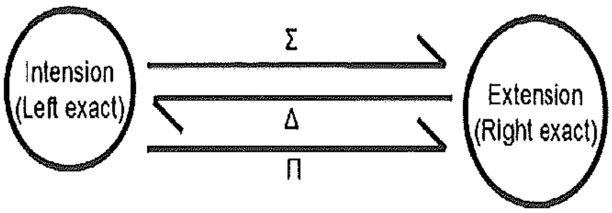


Figure 7: Adjointness$\Sigma \dashv \Delta \dashv \Pi$

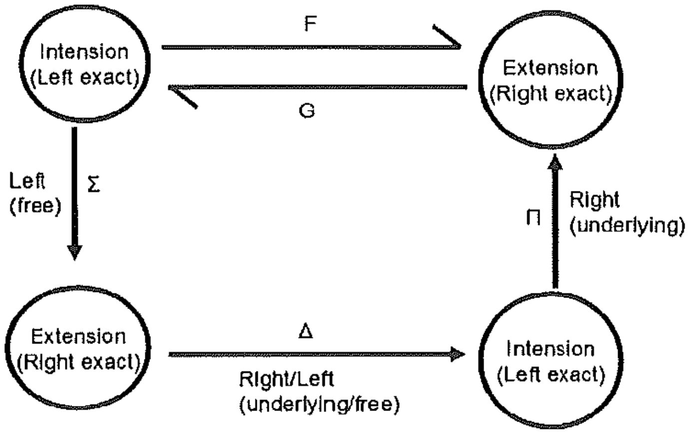


Figure 8: Explosion$\Sigma \dashv \Delta \dashv \Pi$of the Arrow Functors of Adjointness $F \dashv G$

is termed a Cartesian Closed Category, derived as an abstraction of the Cartesian product but this description from historic origins may by its simplicity mislead as to its great power and content. The finitary approach is to distinguish the two properties of Cartesian closed and locally Cartesian closed but in process categories it is that natural distinction between intension and extension that provides a formal description of the mathematical structure of the World as found in nature. It is the simple principle that everything in the world is related to everything else in the world that provides the formal structure of the relationship relevant to any scientific study or technological application requiring an understanding of these relationships.

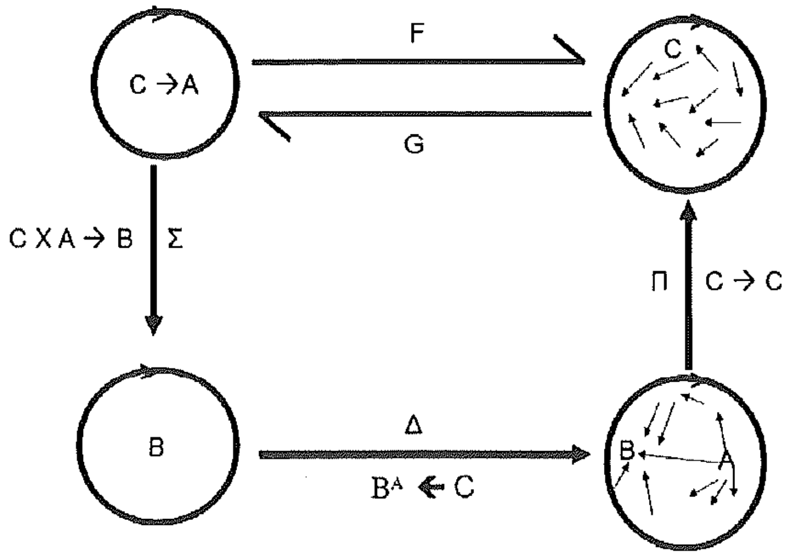


Figure 9: Intuisionistic structure of the Cartesian closed category: Exploded view of Heyting logic.

An early example is the representation of information in computers, that needed some implementable model of real-world relationships. Some variation of the hierarchical was possibly the most common structure attempted in different knowledge systems. But the most successful measured by the volume of commercial transactions was by far the simple relational model based on lists or tables manipulated as sets embodying an intension/extension relationship.

The Cartesian closed category (CCC) is a fundamental category of category theory. Its features and their definitions are to be found in its standard textbooks but most if not all come from the stationary viewpoint of set theory, not from process. That set theory itself does not rest on unequivocal foundations may raise few problems in pure mathematics where axioms may be defined at will and may well be adequate too in applied mathematics to a first order. However, many problems requiring mathematical solutions today arise in more complex situations. Transactions in information systems [22] are a case in point as of the nature of process. Thus a common approach in databases [9] is to adopt the principles under the acronym ACID stating the requirements for Atomicity, Consistency, Isolation and Durability. The aim is to ensure that a transaction involving a series of operations is indivisible, enforces all rules, provides results only on termination and guarantees to hold the results under any circumstances. The transaction concept has been implemented efficiently on many database systems but in information systems as a whole the idea lacks the abstraction needed for successful business modelling. The alternative approach in natural philosophy is that of process as explored in the 20th century [23].

While in the formal language of category theory the world may be described as 'Cartesian-closed', this term may give a false impression that it has a Cartesian coordinate system which is unfortunate but the phrase has arisen historically in that context because it embodies the fundamental concept of the Cartesian product. In fact it is much more than a simple product and these terms need to be examined further. For while natural categories and metaphysics provide us with a process structure for the world, we can only begin to investigate it here. Intension and extension alternate in a preorder, that is with an arbitrary beginning of an intension with an extension which itself becomes an intension of the next extension and so on as in Figure 10 [14].

7 The Topos: Archetype of Natural World

The archetype of the natural world is the topos, in its early days formally defined as a Cartesian closed category with subobject classifiers and informally as a generalised set. Johnstone in his preface to [16] lists thirteen alternative descriptions that have been ap-

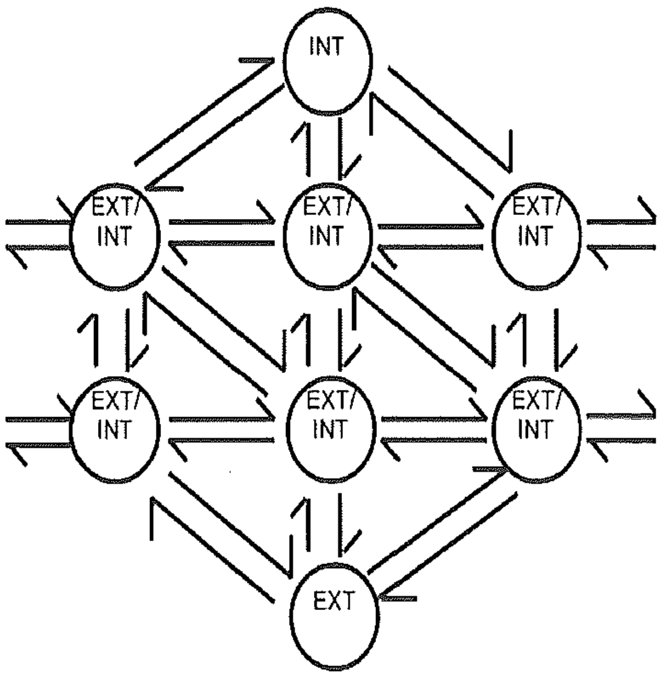


Figure 10: Alternate Intension/Extension Pairs in Nature

plied to the topos (pp.viii &sq). Many of them like for instance "A topos is a generalised space" still carry hangovers from sets. We would recommend as an informal definition: "The category of categories of catgories". To some this may only confirm categories as "abstract nonsense" but it is accurate and and makes explicit the recursion. The topos sums up all that we have said in this paper. It is the ultimate intension existing as an identity natural transformation in any extension given by the internal categories, subject to the locally Cartesian closed condition with the preorder structure and an intuitionistic logic that is the Heyting and which is more general than the Boolean. There is a unique arrow from the source of the World to every object in it and a unique limiting arrow between any pair of objects.

To satisfy its holistic nature the World must emerge top-down. That is to say no more than that if the Big Bang happened it potentially contained everything that ever existed[8]. However it is easier to explain bottom-up by treating the role of the arrow as a natural expression of process with an identity arrow as intension and a distinguishable valued arrow for extension. However while in natural category theory the simplest identity arrow may be treated as an object, it is convenient to begin with a category of three composing objects as a generalisation of any possible category. This is shown in Figure 11 with the next higher identity arrow (the functor) composing extensional arrows between objects. The next higher identity arrow is the locally Cartesian closed natural transformation composing categories with ordinary functors as extensional arrows between categories as shown in Figure 12. The highest level arrow is also a natural transformation which composes structures of categories and functors. It is this identity natural transformation that constitutes the full Cartesian closed category of a topos as in Figure 13. However, the natural arrow is double-headed as a composition of the adjoint functors but with a parity as previously discussed above. Although as just explained it may be easier to understand these diagrams bottom-up in the way that models are usually built-up, nevertheless process can only exist as a whole and the full diagram represents a natural occasion or "actual event" as first introduced by Whitehead [28]. From the long-term perspective of ANPA however the four-level Combinatorial Hierarchy based on the Frederick Construction is a binary model of the four-level metaphysical preorder of Process presented here.

The whole is just a recursive system with closure at four levels consisting of three open interfaces. Figure 13 shows the three interfaces for composing arrows (ordinary, functor, natural transformation) with the four levels (identity arrow, identity functor/category,

[8]formally (μ,δ) in the description above for the monad/comonad.

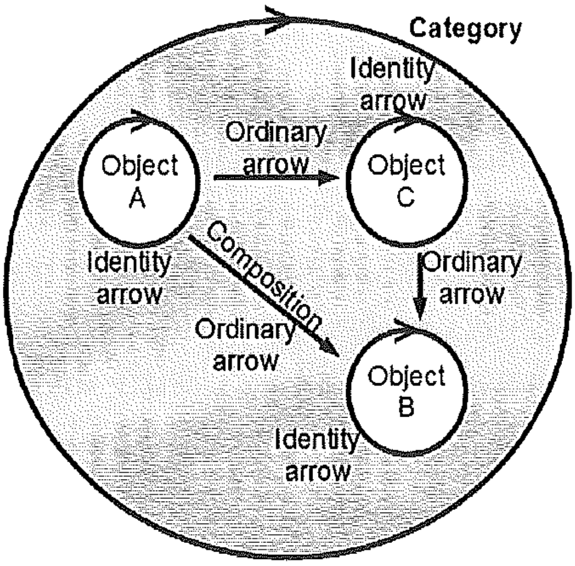


Figure 11: A category consists of ordinary arrows composing between identity arrows as objects

higher-order identity functor/category, identity natural transformation/topos). The diagram shows well the natural recursive nature of the structure. It also demonstrates connectivity from any object to any other object. It is possible therefore, as shown in Figure 14, to get from any object A to any object B directly: $B = \theta A$, or indirectly with possible local variations through any other internal path: $\theta'' \circ \theta'$A= B. This is a natural structure because it is obtained from simple induction applied to the notion of process without any assumptions. As a final comment it is interesting to compare briefly the World as a topos with the long-term study by ANPA of the four-level Combinatorial Hierarchy based on the Frederick Construction as a binary model of the four level metaphysical preorder of process presented here. The subobject classifiers of any intension are the Boolean truth values (0,1) as the initial and final objects of a topos that is both Cartesian and co-Cartesian. The intension generates by process the possible extensions but is limited by scale invariance to four levels of three

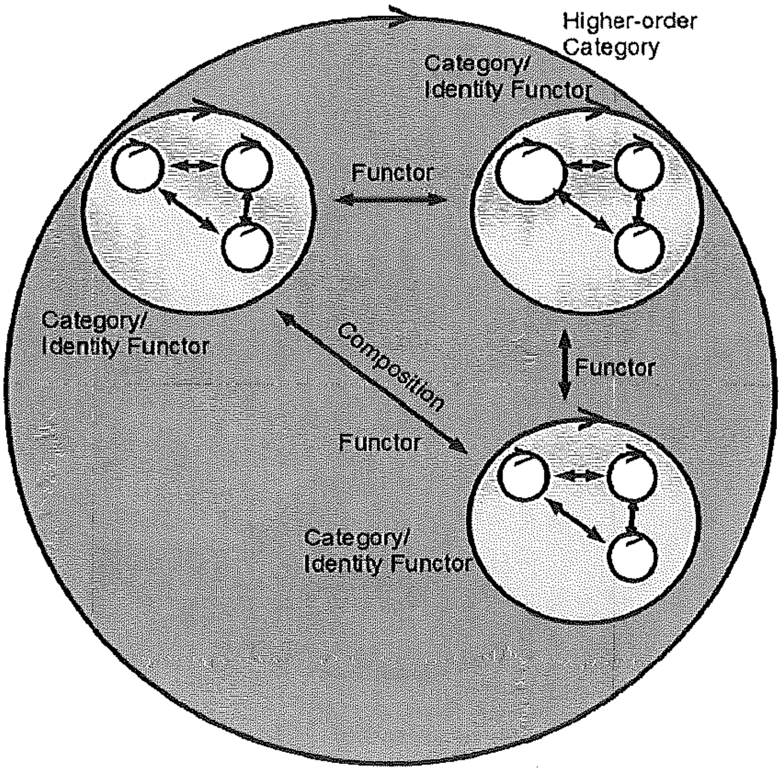


Figure 12: Functors between Categories compose to form higher-order categories. A category is just an identity functor

interfaces which the Frederick construction as a model predicts and correlates to a great precision with the fine structure constant.

Appendix I: Finitary Approaches to Cartesian closed categories

The relationship between natural categories and finitary category theory is symbiotic as part of the general mutual independence on one another of pure and applied mathematics. Natural categories being metaphysical are at the highest possible level and therefore lack a higher vantage point from which to view them. Finitary category theory on the other hand is a model relying mostly on the catgegory of sets. Being finitary the subject can be advanced by a number of categorial proofs. Understanding cat-

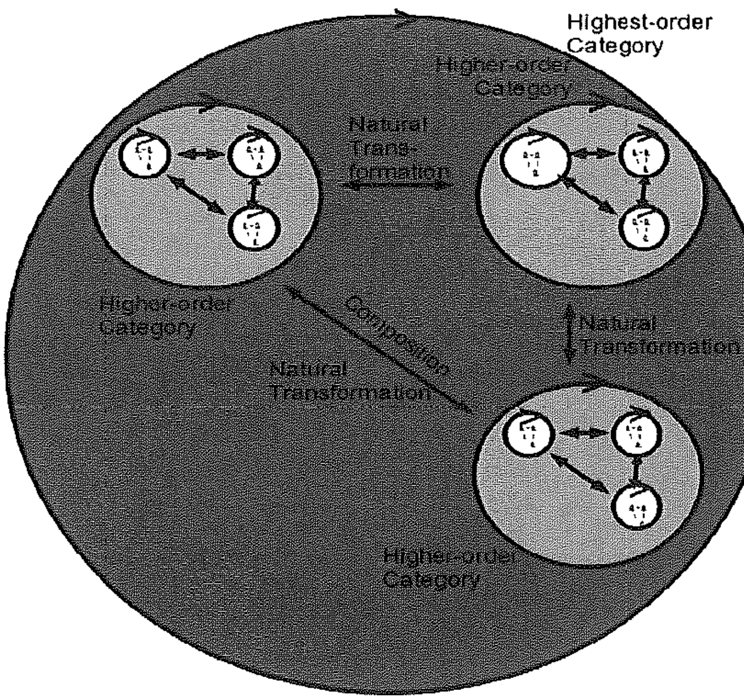


Figure 13: Natural Transformations of Composing Functors themselves compose in the highest possible category, a Topos

egories on the other hand has only pure induction to guide by empirical reality through the natural metaphysics. This is an important example of the three-tier general scheme of metaphysics, physics and models of Figure 4. Because of the symbiosis between the pure and applied approach to formalism it is instructive to compare the traditional treatment of Cartesian closed categories in finitary category theory. Seminal texts are that of Barr & Wells [3] for applications in computer science and Mac Lane's work in pure mathematics [21]. It is to be noted that their treatment is syntactical rather than semantic and the deep applied significance may not be too obvious in these syntactical descriptions.

Appendix I(a): Treatment by Barr and Wells

The classical approach as followed by Barr & Wells ([3] pp.142-

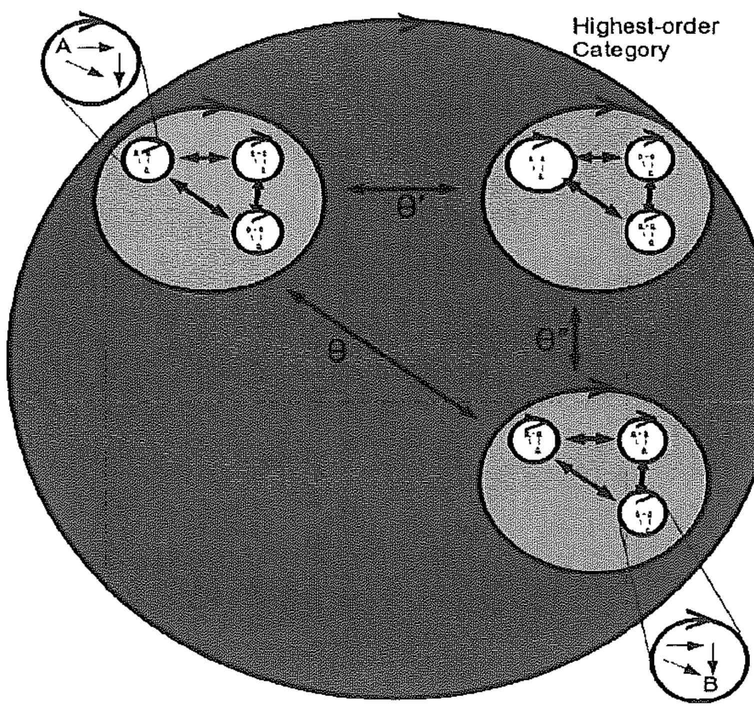


Figure 14: A topos showing natural path from any object A to any object B

160) defines a category **C** as Cartesian Closed if it satisfies the three conditions reproduced from their description in Panel 2.

Traditionally the family of arrows (historically known as a Hom functor) from A to B is written as $[A \longrightarrow B]$ or denoted as B^A and then called the exponential object with A as the exponent. It is possible to add some semantic detail to the statements CCC-1 to CCC-3 in the panel and draw formal diagrams to indicate further aspects. In basic terms the definition above requires a terminal object T as an upper limit closing the category from above. This has to be independently defined for the category of sets because there is no syntactical connection between the extension and the intension of a set. It lies in the semantics unexpressed and the connection has to be made in the mind of the user. A natural category on the other hand exists as an intension identity arrow typing, by means of a contravariant arrow, every object in its possible extensions. A pair of objects has a product with projections

Panel 2 : Three Conditions for a Cartesian Closed Category ([3] p.143)

CCC-1	There is a terminal object 1
CCC-2	Each pair of objects A and B of **C** has a product $A \times B$ with projections $p_1 : A \times B \longrightarrow A$ and $p_2 : A \times B \longrightarrow B$
CCC-3	For every pair of objects A and B, there is an object $[A \longrightarrow B]$ and an arrow eval : $[A \longrightarrow B] \times A \longrightarrow B$ with the property that for any arrow $f : C \times A \longrightarrow B$, there is a unique arrow $\lambda f : C \longrightarrow [A \longrightarrow B]$ such that the composite $C \times A \xrightarrow{\lambda f \times A} [A \longrightarrow B] \times A \xrightarrow{\text{eval}} B$ is f

where there is only one path between the product and the related object. More precisely:

CCC-1 For any object A in the category, there is exactly one arrow $A \longrightarrow T$, where T is the terminal object and the category is closed on top T. This is quite straight forward in finitary categories where the elements of a set are defined as independent of one another and can only be related by functions. In natural categories there is no such independence because of the nature of process every object in the world is related to every other. The semantics of CCC1 would then express the wholeness of the category.

CCC-2 expresses the property that any pair of objects may combine and any such combination may be resolved into one or other of its components. This appears fairly obvious at the syntactical level but provides the basis of relationships at the semantic level. Any combination is dependent on context which qualifies any relationship.

The first limb of **CCC-3** provides for currying to change a function on two variables to a function on one variable. For function $f : C \times A \longrightarrow B$, let $[A \longrightarrow B]$ be the set of functions from A to B. Then there is a function: $\lambda f : C \longrightarrow [A \longrightarrow B]$ where

$\lambda f(c)$ is the function whose value at an element $a \epsilon A$ is $f(c, a)$. This is equivalent to the typed lambda calculus. Typical examples of currying with integers often given are:

$$f : \text{multiply}(\_, 2) \longrightarrow R \text{ curries to } \lambda f : \text{double}(\_) \longrightarrow R$$

$$f : \text{exponentiate}(\_, 2) \longrightarrow R \text{ curries to } \lambda f : \text{square}(\_) \longrightarrow R$$

The use of 'double' and 'square' are examples of semantic expressions used to bridge conceptually the gap between intension and extension in set theory. This is the finitary syntactical version of the property in the universe that there is a single direct connection between any pair of two entities, that is the resultant of all possible connections between them as illustrated in the diagram of Figure 3. The language used by Barr & Wells in these definitions is not purely categoreal but as not uncommon in finitary category theory it is often necessary to resort to hybrid descriptions involving set theoretic concepts as with the use here of lambda calculus, invented by Church to express for the purposes of set theory the concept of typing as a limit. Lambda calculus was known, from early on and for similar reasons, to be logically inconsistent. It is subject to the Kleene-Rosser paradox, which is another incarnation of Russell's paradox.

In the second limb of **CCC-3**, for every pair of objects A and B, there is an object $[A \longrightarrow B]$ and an arrow eval : $[A \longrightarrow B] \times A \longrightarrow B$ with the property that for any arrow $f : C \times A \longrightarrow B$ there is a unique arrow $\lambda f : C \longrightarrow [A \longrightarrow B]$ such that the diagram in Figure 15 commutes.

In Figure 15 C is the product object and eval is a function mapping all A objects and their associated B objects onto B. The

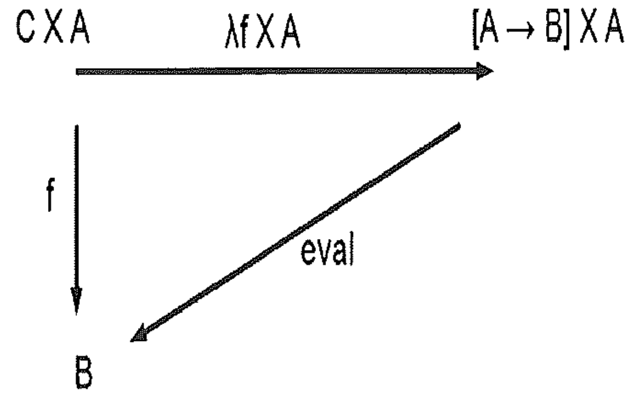


Figure 15: Commuting Diagram for Rule CCC-3 (second limb) for a Cartesian closed category

semantics is very profound in that it leads to the Heyting logic mentioned previously which is only possible in finitary category theory by arbitrary enhancement but is naturally inherent in process categories where it is essentially the metaphysics of causation.

The problems which arise from the lack of formal integrity between the intension of a set and the extension of its elements carry over into the concept of 'locally Cartesian closed'. Natural categories have the property of being both Cartesian closed and locally Cartesian closed. As arbitrary models finitary categories may have the former property without the latter. Categories with both properties are treated as strong and those that are not also locally Cartesian closed as weak. In the former products are extended to pullbacks and Barr & Wells rely on this to distinction to define locally Cartesian closed ([3] at p.353). Categories are locally Cartesian closed when the category **C** has pullbacks and either the pullback functor has a right adjoint or for every object A in **C**, the slice category $\mathbf{C/A}$ is Cartesian closed. Pullbacks express relationships over objects in a particular context so locally Cartesian closed categories provide more expressiveness for finitary categories in representing the real world. Figure 16 compares the product and pullback.

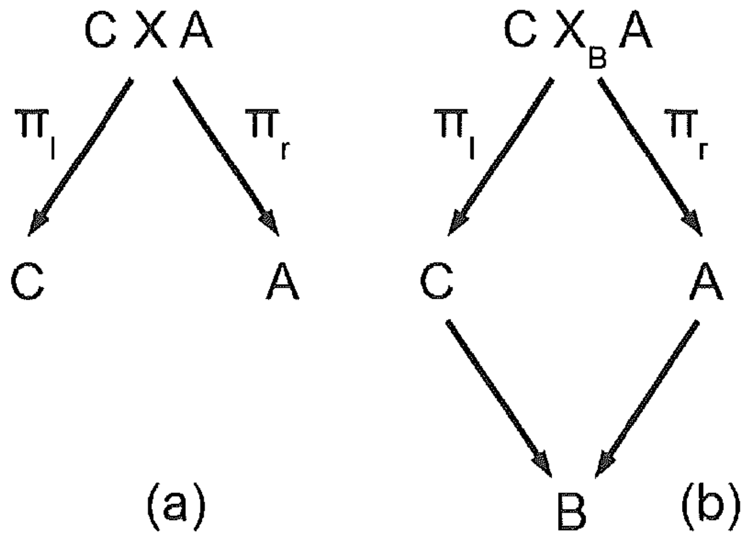


Figure 16: Comparison of Constructions (a) Product $C \times A$ and (b) Pullback $C \times A$ in context of B

Some greater insight on their application to the real world comes from the first chapter in volume I of Peter Johnstone's *Sketches of an Elephant* [16]. A category is Cartesian closed if it has a terminal object, products of pairs of objects and equalizers of pairs of morphisms. A category is locally Cartesian closed if it has a terminal object and pullbacks of pairs of morphisms ([16] A1.2 p.11). A Cartesian closed category is locally Cartesian closed if it has pullbacks. The property of Cartesian-ness is stable under slicing ([16] A1.2.6). That is the stability functor$\Delta$is in adjointness with the existential functor $\exists \dashv \Delta$ and with the universal functor $\Delta \dashv \forall$ for a pullback category. The approach by Barr & Wells to Cartesian closed categories can be adjusted to a more abstract view using adjointness. In the potentially adjoint relationship $F \dashv G$, the free functor F creates binary products and the underlying functor G checks for exponentials, that is one path. The free functor $\_ \times A$ takes an object C to its product with A, that is $C \times A$. The underlying functor G takes a product object $C \times A$ to an object B. Figure 17 shows the diagrams that must both commute for adjointness to hold, diagram (a) for the left adjoint and (b) for the right adjoint. A comparison of Figures 15 and 17(b) shows that in the former the arrows $\_ \times A(f)$ and ϵ correspond respectively to $\lambda f \times A$ and eval in the former. The counit of the adjointness is therefore the evaluation map.

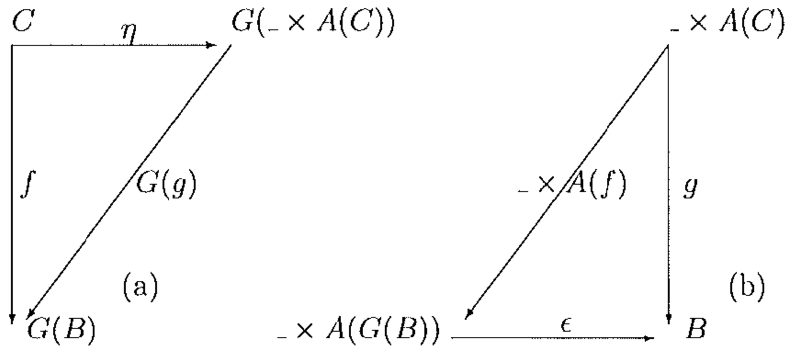


Figure 17: Roles in Adjointness of a) η, the unit and b) ϵ, the counit of adjointness respectively. Free functor is $\_ \times A$.

Finitary Cartesian closed categories can be readily extended from binary products to finite products and this is demonstrated by Barr & Wells ([3] pp.191-196). For any objects $A_1, A_2, \ldots,$ and A of a Cartesian closed Cartesian closed category, there is an object $[A_1 \longrightarrow A]$ and an arrow:

$$eval : [A_1 \longrightarrow A] \times A_1 \longrightarrow A$$

such that for any $f : \pi_{A_2} \longrightarrow A$, there is a unique arrow:

$$\lambda_1 f : \pi_{A_2} \longrightarrow [A_1 \longrightarrow A]$$

Finite products give construction of n-tuples which Barr & Wells

[3] show can represent strings through constructions such as the Kleene closure (p.340) and the Kleisli category (pp.366-367). These seem attractive but being derived from sets they have to be treated with caution for use in information systems. For they are still models and prone to the same difficulties we have already discussed. For instance the issue arises with the significance of order. This is not new. Historically there has been some debate about whether $A \times B$ is 'the same' as $B \times A$. Barr & Wells for instance are compelled to acknowledge the difficulty with finite products as in Panel 3.

Panel 3 : Problems with Equivalence of Products in CCC-3 ([3] p.144)

> Condition **CCC-3** appears to treat the two factors of $C \times A$ asymmetrically, which is misleading since of course $C \times A \equiv A \times C$. Even that last isomorphism is misleading since $C \times A$ and $A \times C$ could be taken to be the same object. Products are of indexed sets of objects, not necessarily indexed by an ordered set, even though our notation appears to suggest otherwise. It gets even worse with n-ary products ...

In applications such as relational databases a product is regarded as an associative operation so that $A \times (B \times C)$ is regarded as equivalent to $(A \times B) \times C$, at least at the data level. But this is the problem: extensionally the product operation is associative. However, intensionally a different answer is obtained depending on the order of the operations. So the product operation is not associative in Cartesian closed systems.

Appendix I(b): Treatment by Mac Lane

Mac Lane ([21], pp.87-88) defines Cartesian closed in tabular form using the diagonal functor$\Delta$for product and the terminal object in category C in **Set** as reproduced here in Panel 4.

Mac Lane asserts the existence of a Cartesian Closed Category as equivalence with adjointness, as in Panel 5.

Panel 4 : Left and Right Adjoints in Cartesian Closed Category **C** in **Set** after ([21] pp.87-88)

Functor	*Adjoint*	*Unit*	*Counit*
$\Delta : C \longrightarrow C \times C$	Left: Coproduct $\amalg : C \times C \longrightarrow C$ $< a, b > \longmapsto a \amalg b$ Right: Product $\Pi : C \times C \longrightarrow C$ $< a, b > \longmapsto a \times b$	(pair of) injections $i : a \longrightarrow a \amalg b$ $j : b \longrightarrow a \amalg b$ Diagonal arrow $\delta_c : c \longrightarrow c \times c$ $x \longmapsto < x, x >$	'folding map' $c \amalg c \longrightarrow c$ $ix \longmapsto x, jx \longmapsto x$ (pair of) projections $p : a \times b \longmapsto a$ $q : a \times b \longmapsto b$
$C \longrightarrow 1$	Left: Initial object s Right: Terminal object t	 $c \longrightarrow t$	$s \longrightarrow c$

Panel 5 : Assertion of Cartesian Closed Category as Equivalence with Adjointness ([21] p.97)

> To assert that a category C has all finite products and coproducts is to assert that products, terminal, initial and coproducts exist, thus the functors $C \longrightarrow 1$ and $\Delta : C \longrightarrow C \times C$ have both left and right adjoints. Indeed the left adjoints give initial object and coproduct, respectively, while the right adjoints give terminal object and product, respectively.

Mac Lane ([21] at pp.97-98) thus by using just adjoints at both the category level and the object level is able to define "Cartesian Closed Category'. He puts it this way: a category **C** with all finite products specifically given is called *Cartesian closed* when each of the following functors in Panel 6 has a *specified* right adjoint (with a specified adjunction) in Panel 7.

Panel 6 : Functors and Maps involved in Adjointess ([21] p.98)

$C \longrightarrow 1$,	$C \longrightarrow C \times C$,	$C \xrightarrow{-\times b} C$,
$c \longmapsto 0$,	$c \longmapsto < c, c >$,	$a \longmapsto a \times b$,

The first adjoint in Panel 7 specifies the terminal object and the second the product and its projections. The third specifies the evaluation map as shown in Panel 8.

Mac Lane's treatment, in common with that of Barr & Well's,

Panel 7 : Right Adjoints for Cartesian Closed Category ([21] p.98)

$t \longleftarrow\!\!\!| \; 0$, $\quad a \times b \longleftarrow\!\!\!| < a, b >$, $\quad c^b \longleftarrow\!\!\!| \; c$,

Panel 8 : Evaluation map as condition for adjointness in Cartesian Closed Category ([21] p.98)

The third required adjoint specifies for each functor $- \times b : C \longrightarrow C$ a right adjoint, with the corresponding bijection
$\hom(a \times b, c) \equiv \hom(a, c^b)$
natural in a and in c. By the parameter theorem (to be proved in the next section), $< b, c > \longmapsto c^b$ is then (the object function of) a bifunctor $C^{op} \times C \longrightarrow C$. Specifying the adjunction amounts to specifying for each c and b an arrow e
$e : c^b \times b \longrightarrow c$
which is natural in c and universal from $- \times b$ to c. We call this $e = e_{b,c}$ the *evaluation* map.

is restricted to the category of **Set**. From the point of real-world systems such as information and database systems, this is unsatisfactory as in the Boolean world there is a reliance for negation on the closed world assumption. What is required is an open system, through the free functor F, with Heyting intuitionistic logic to give negation in an approach which does not violate Gödel's principles.

References

[1] Arnauld, Antoine, & Nicole, Pierre, *Logic or the Art of Thinking*, Translated Buroker, Jill Vance, Cambridge (1996).

[2] Ashby, W R, *Design for a Brain*, Chapman & Hall (1952).

[3] Barr, M, & Wells, C, *Category Theory for Computing Science*, 1st edition, Prentice Hall 432pp (1990).

[4] Barrow, John D, *The Book of Universes* Bodley Head (2011).

[5] Bastin, T, & Kilmister, C W, *The Origin of Discrete Particles* Series on Knots and Everything **42** 196pp (2009).

[6] Bertalanffy, L von, *The Theory of Open Systems in Physics and Biology* Science **111** 23-29 (1950).

[7] Bertalanffy, L von, *General Systems Theory* Penguin, London (1971).

[8] Codd, E F, *The Relational Model for Database Management* Addison-Wesley (1990).

[9] Connolly, T, & Begg, C, *Database Systems* Pearson (2004).

[10] Gödel, Kurt, Die Vollständigkeit der Axiome des logischen Funktionenkalküls *Monatshefte für Mathematik und Physik* **37** 349-360 (1930). The completeness of the Axioms of the Functional Calulus of Logic. Feferman and co: Collected Works 102-122 German even, 103-123 English odd. Translated by Stefan Bauer-Mengelberg, read and approved by Gödel after some accommodation. Subsequent minor additions by Jean van Heijenoort (p.59 collected works **1** p.59, Feferman) (1986).

[11] Gray, Jim, Thousands of Debit/Credit Transactions-Per-Second: Easy and Inexpensive, *Microsoft Technical Report Series* MSR-TR-2005-39 (2005).

[12] Greene, Brian, *The Hidden Reality: Parallel Universes and the Deep Laws of the Cosmos* Allen Lane Hardback (2011).

[13] Heather, M.A. & Rossiter, B N, Don't Confuse Left and Right Exactness! A Note correcting the Distinction in the Proceedings of ANPA 27, *ANPA 29* Exactness, Cambridge, 157-159 (2008).

[14] Heather, Michael, Livingstone, David, & Rossiter, Nick, Logical Foundations for the Infrastructure of the Information Market, I-ESA, Berlin, 25-28 March, *Enterprise Interoperability III, New Challenges and Industrial Approaches*, Mertins,

Kai, Ruggaber, Rainer, Popplewell, Keith and Xu, Xiaofei, (edd) 624- 637 (2008).

[15] IBM, Data deduplication in Tivoli Storage Manager, http://www.ibm.com/developerworks/wikis/display/tivolistoragemanager/Data+deduplication+in+Tivoli+Storage+Manager+V6.2+and+V6.1 (2010).

[16] Johnstone, P T, Sketches of an Elephant, A Topos Theory Compendium, *Oxford Logic Guides* 43, Clarendon (2002).

[17] Lawvere, F W, An elementary theory of the category of sets, *Proc Nat Acad Sci* **52** 1506-1511 (1964).

[18] Lawvere, F W, Adjointness in Foundations, *Dialectica* **23** 281-296 (1969).

[19] Leinster, Tom, *Higher Operads, Higher Categories*, London Mathematical Society LNS 298, Cambridge (2004)

[20] McLuhan, Marshall, *The Gutenberg Galaxy: The Making of Typographic Man* Toronto 263pp (1962).

[21] Mac Lane, S, *Categories for the Working Mathematician*, 2nd ed, Springer-Verlag, New York (1998).

[22] Rossiter, Nick, Heather, Michael, & Nelson, David, *A Natural Basis for Interoperability*, in: Enterprise Interoperability: New Challenges and Approaches, Doumeingts, G, Muller, J, Morel, G, & Vallespir, B, (edd) 417-426 (2007).

[23] Rossiter, B N, Heather, M A, & Sisiaridis, D, Process as a World Transaction, *Proceedings ANPA* 27 Conceptions, 122-157 (2006).

[24] Rossiter, Nick, & Heather, Michael, *Free and Open Systems Theory*, EMCSR-2006, Cybernetics and Systems, 18th Eu-

ropean Meeting on Cybernetics and Systems Research, University of Vienna, 18-21 April 2006, Trappl, R, (ed) 1 27-32 (2006).

[25] Rossiter, Nick, & Heather, Michael, The Natural Metaphysics of Computing Anticipatory Systems, *IJCAS* in press, 18pp (2011).

[26] Rothemund, P W K, Folding DNA to create nanoscale shapes and patterns, *Nature* **440** 297-302 (2006).

[27] Whitehead, Alfred North, & Russell, Bertrand, *Principia Mathematica* **1** Cambridge University Press (1910).

[28] Whitehead, Alfred North, *Process and Reality: An Essay in Cosmology*, Macmillan, New York (1929).

[29] Whitehead, Alfred North, *Process and Reality: An Essay in Cosmology*, corrected edition, edd D R Gri ffi n, D W. Sherburne, New York, The Free Press (1978).

[30] Wiener, Norbert, *Cybernetics, or Control and Communication in the animal and the machine* John Wiley (1948).

[31] Wigner, E P, The Unreasonable Effectiveness of Mathematics in the Natural Sciences, *Comm Pure and Applied Math* **13** 114 (1960).

[32] Wikipedia, Process, http://en.wikipedia.org/wiki/Process.

Higher Categories and Self-Reference

Louis H. Kauffman
Department of Mathematics, Statistics
and Computer Science (m/c 249)
851 South Morgan Street
University of Illinois at Chicago
Chicago, Illinois 60607-7045
<kauffman@uic.edu>
Tel: 1-312-996-3066
Fax:1-312-996-1491.

Abstract

The paper introduces the notion of a categorical pair, *a pair of categories* (C, C') *such that every morphism in* C *is an object in* C'. *Categorical pairs are precursors to 2-categories. Arrows in* C' *can express relationships between the morphisms of* C. *In particular we show that by using a model of the linguistic process of naming, we can ensure that every morphism* F *in* C *has an indirect self-reference of the form* $a \longrightarrow Fa$ *where this arrow occurs in the category* C'. *This result is shown to generalize and clarify known fixed point theorems in logic and categories, and is applied to Goedel's Incompleteness Theorem, the Cantor Diagonal Process and the Lawvere Fixed Point Theorem.*

Key Words. *category; categorical pair; 2-category; indicative shift; self-reference; indirect self reference.*

1 Introduction

The purpose of this paper is to introduce a categorical pattern that generalizes and complements the Lawvere Fixed Point Theorem. We produce a construction for indirect self-reference that applies directly both to situations in ordinary language and to Goedel's Theorem on the incompleteness of formal systems. Our construction can be summarized very succinctly and so we begin the paper with a self-contained account of the construction, and then devote the rest of the paper to discussion about how this *indicative shift* [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12] can be seen in a number of different contexts. The indicative shift is defined in Section 2. The shift formalizes an operation on names that can also be regarded as an *expansion* of a name in the sense that if "A" is the name of A then the expansion E"A" refers to A"A", the result of appending the contents of the name to the name. Thus if we regard the name as pointing to its contents as in

$$\text{"}A\text{"} \longrightarrow A$$

then

$$E\text{"}A\text{"} \longrightarrow A\text{"}A\text{"}.$$

Self-reference results when one expands the name of the expansion operator.

$$E\text{"}E\text{"} \longrightarrow E\text{"}E\text{"}.$$

The paper is organized as follows. In Section 2 we construct a categorical context for the indicative shift by considering a pair of categories (C, C') where every morphism in the first category is an object in the second category. Arrows in the second category can be interpreted as references between arrows in the first category. In this sense the second category of the categorical pair defined in this section takes the place of a meta-language in a logical context. We prove three basic results about self-reference in this section that we call the First, Second and Third Self-Reference Theorems. The First Self-Reference Theorem gives conditions under which an indirect self-reference can occur. In such a situation there is a formula F that can internally talk about an entity n. Formally we write Fn as a composition of morphisms in the first category. In the second category there may be an arrow between these morphisms of the form

$$n \longrightarrow Fn.$$

This arrow constitutes a reference from n to Fn. Informally, n names Fn and Fn talks about n, so Fn refers to its own name, making an indirect self-reference. We show how to prove that such references always exist under the conditions of the indicative shift.

The formality of the indicative shift is very simple. Suppose that

$$a \longrightarrow F$$

is an arrow in C'. Then it is either given or constructed (First and Second Self-Reference Theorems) that there is an operator $\sharp$ (independent of a)such that

$$\sharp a \longrightarrow Fa.$$

This is the indicative shift. It follows that if

$$g \longrightarrow F\sharp$$

then

$$\sharp g \longrightarrow F\sharp g,$$

producing the desired indirect self-reference.

The Second Self-Reference Theorem assumes that the pair of categories (C, C') is a 2-category, and shows that the indicative shift with $\sharp a = aa$ follows naturally from the properties of composition in a 2-category. The Third Self-Reference Theorem is a generalization of the formalism of the Lawvere Fixed Point Theorem. This is explained in Section 4. We end Section 2 with an application of its ideas to an example of Raymond Smullyan. Smullyan's example is a miniature version of Goedel's incompleteness theorem. In Section 3 we show how the Self-Reference Theorems apply to Goedel's Theorem when that theorem is seen as the production of a statement that asserts its own unprovability in a given formal system. In Section 4 we discuss the relationship of these ideas with the Lawvere Fixed Point Theorem, and we discuss how the Lawvere Theorem relates to Goedel's Theorem via its analog with the Cantor diagonal process. The difference between our categorical approach and that of the Lawvere Theorem is that we formalize indirect self-reference and this occurs most naturally by using higher categories.

In Section 5 we discuss how these category ideas and the indicative shift apply to ordinary language. More work needs to be done in relating these formalisms to ordinary language, for it is in ordinary language that the line between the categories (between level and meta-level) is easily erased. Names of names are still names in ordinary language, and in the language of categories, objects and morphisms can become interchangeable. From the mathematical side one can approximate the situation of language by using higher categories or even reflexive categories (where ideally there is a 1-1 correspondence between objects and morphisms) rather than the 2-categories of Section 2. We give an example of a reflexive category (in that every object is a morphism) by taking the generating arrows and the objects to be the arcs of an oriented knot diagram. Section 6 is an Epilogue that reviews and discusses the ideas and results of the paper.

We also have not addressed questions of infinity and cardinality in this paper except in drawing the relationship between these results and the Cantor diagonal argument that is directly in back of formalisms for higher cardinalities. The Lawvere Fixed Point Theorem allows contexts with non-standard subobject classifiers and hence alternate views of the infinite. We expect that the present categorical constructions related to the indicative shift will also shed light on the epistemology of infinity. This will be pursued in a subsequent paper.

2 The Indicative Shift

The indicative shift defined in this section formalizes an operation on names that can also be regarded as an *expansion* of a name in the sense that if "A" is the name of A then the expansion E"A" refers to A"A", the result of appending the contents of the name to the name. Thus if we regard the name as pointing to its contents as in

$$\text{"}A\text{"} \longrightarrow A$$

then

$$E``A" \longrightarrow A``A".$$

Self-reference results when one expands the name of the expansion operator.

$$E``E" \longrightarrow E``E".$$

In those contexts where one thinks of expanding a name to its contents it is convenient to use the symbol E for the shift operator. In this section we shall adopt the symbol $\sharp$ for the shift. When we use the symbol $\sharp$ we are thinking of the shift at the point where a name is given to a contents. At that point, there is an initial pointing of the name to the contents before the name is directly associated with the contents. Then a shift occurs where the name is associated with the contents and the abstract name is associated with the reference to these contents. These points of language are further discussed in Section 5.

The reader should recall that a *category* consists in a collection of *objects* and a collection of *morphisms*. To each morphism f there is associated an ordered pair of objects (A, B). We write $f : A \longrightarrow B$ to denote the morphism and call A the *domain* of f and B the *codomain* of f. Given morphisms $f : A \longrightarrow B$ and $g : B \longrightarrow C$, there is a morphism $g \circ f : A \longrightarrow C$, called the *composition* of f and g. Composition of morphisms is associative. Every object A comes equipped with an identity morphism 1_A whose composition (with A in the role of domain or codomain) with another morphism does not affect that morphism. This is the complete definition of a category.

We take as given that in a category one can say whether two objects are equal and whether two morphisms are equal. We wish to model situations where equality is replaced by reference. We speak this way for motivation and use the word reference as it is used in ordinary language where one may say that the name of a person refers to that person, or that the title of a paper refers to the text or to the contents of the paper. We wish to model situations where one distinguishes between the morphisms of a given category and certain patterns of reference that are seem among these morphisms at a second level. A good example, that we shall later consider is the reference of a Goedel number to its corresponding decoded text. One might regard the Goedel number as a stand-in for its corresponding text, and it may be important to simply regard the Goedel number as a name of that text.

Let there be given a category C and suppose that the set of morphisms of C are seen as the objects in another category C'. We shall call the morphisms in C' *reference arrows* for the mophisms of C, and we shall call the pair of categories (C, C') a *categorical pair*. We make no further restrictions on a categorical pair other than that there are two categories C, C' with the morphims in the first category forming the objects in the second category.

In the literature, there is a notion of *2-category* (and of higher categories). A 2-category is a categorical pair with extra structure. In the notation of our reference arrows, the extra structure is as follows. One may have

$$\alpha : a \longrightarrow b$$

and

$$\beta : d \longrightarrow e,$$

arrows in C' where ad and be are both legal compositions in the base category C of the categorical pair. Then it is natural that there should be a referential arrow

$$\alpha \circ_0 \beta : ad \longrightarrow be,$$

usually called horizontal composition of these arrows in C'.

Along with horizontal composition we have *vertical composition* which is simply the given composition of arrows in C'. We can denote vertical composition by $\circ_1$. Thus if

$$\alpha : a \longrightarrow b, \gamma : b \longrightarrow c$$

then

$$\gamma \circ_1 \alpha : a \longrightarrow c.$$

Now suppose that we have two possible vertical compositions

$$\alpha : a \longrightarrow b, \gamma : b \longrightarrow c$$

$$\beta : d \longrightarrow e, \delta : e \longrightarrow f$$

where ad, be and cf are each legal compositions in the base category C of the categorical pair. Then it is natural to demand the compatibility

$$(\alpha \circ_0 \beta) \circ_1 (\gamma \circ_0 \delta) = (\gamma \circ_1 \alpha) \circ_0 (\delta \circ_1 \beta).$$

A categorical pair (C, C') that satisfies this compatibility (called the *interchange law*) is called a 2-*category* [15].

Consider a categorical pair (C, C'). Let a and b be morphisms in C and let $a \longrightarrow b$ be a morphism in C' with domain a and codomain b. Remember that while a and b are morphisms in the initial category C, they are objects in the referential category C'. We call this arrow a *reference* from a to b. Now for simplicity, assume that C has only one object so that any two morphisms in C can be composed. Let there be given a special morphism $\natural$ in C with the following property:

The Indicative Shift. *If $a \longrightarrow b$ is a reference arrow in C', then there is a uniquely associated reference arrow $\natural a \longrightarrow ba$. Here $\natural a$ and ba denote the compositions of these morphisms in the initial category C.*

A categorical pair with these properties is called a *referential pair.*

First Self-Reference Theorem (SRT1). Let (C, C') be a referential pair. Let F be any morphism in the category C. Then there exists a morphism a in C and a reference arrow in C' such that $a \longrightarrow Fa$.

Proof. Let

$$g \longrightarrow F\sharp$$

be any arrow from some morphism in C to the composition $F\sharp$. (We could take it to be the identity arrow for $F\sharp$ if neccessary. Recall that there is an identity arrow for $F\sharp$ as an object in the category C'.) Then apply the indicative shift and obtain:

$$\sharp g \longrightarrow F\sharp g.$$

Thus with $a = \sharp g$, we have $a \longrightarrow Fa$. This completes the proof. //

Remark. We interpret an arrow of the form $a \longrightarrow Fa$ as a model of an expression Fa that is talking about (in the internal language of compositions in C) its own "name" (which is the morphism a from the point of view of the category C').

Remark. Note that if, in the above proof, we had used the identity morphism in C' on $F\sharp$,

$$F\sharp \longrightarrow F\sharp,$$

then the indicative shift produces

$$\sharp F\sharp \longrightarrow F\sharp F\sharp$$

and we have $a = \sharp F\sharp$ with $a \longrightarrow Fa$. Since every object in C' has an identity reference arrow, self-reference is given at this level. The First Self-Reference Theorem guarantees indirect self-reference in the sense of the existence of arrows of the form $a \longrightarrow Fa$. Such arrows arise via the indicative shift as the proof shows.

Remark. We can regard the referential arrows of the category C' as generalizations (categorifications) of the *equality* of morphisms in the base category C. If the referential arrows are themselves taken to be equalities then the indicative shift would state that if $a = b$ as morphisms in C, then $\sharp a = ab$. In other words, in this degenerate form, we would have $\sharp a = aa$ for all morphisms a in C.

The First Self-Reference Theorem would then correspond to the following calculation. If

$$g = F\sharp$$

then

$$\sharp g = F\sharp g.$$

Hence

$$gg = Fgg.$$

The reader will recognise that this is exactly the form of the proof of the Church-Curry Fixed Point Theorem for Lambda Calculus [1]. See the Epilogue (Section 6 of this paper) for more discussion of this point. Assuming the Indicative Shift is a strong assumption about the categorical pair, generalizing the Church-Curry Fixed Point Theorem to a context that encompasses indirect self-reference.

With this motivation, we can formulate a second self-reference theorem that is close to the flavor of the lambda calculus. Now we will assume that the pair (C, C') is a 2-category and that C has one object so that any two morphisms in C can be composed. For a morphism a in C we define

$$\sharp a = aa,$$

the composition of a with itself. We call the 2-category a *lambda pair* if $\sharp$ is itself a morphism in C.

Second Self-Reference Theorem (SRT2). Let (C, C') be a 2-category that is a lambda pair as defined above with $\sharp a = aa$. Then, given a morphism $a \longrightarrow F$ in C', there is a corresponding morphism $\sharp a \longrightarrow Fa$. With respect to this indicative shift we obtain indirect self-reference from any morphism $a \longrightarrow F\sharp$ by taking the corresponding shift to $\sharp a \longrightarrow F\sharp a$. Note that this morphism is the same as $aa \longrightarrow Faa$. This final conclusion is a direct generalization of the Church-Curry Fixed Point Theorem.

Proof. Suppose we have a morphism

$$a \longrightarrow F$$

in C'. Let $a \longrightarrow a$ be the identity morphism for a in C'. Then we have the horizontal composition of these two morphisms:

$$aa \longrightarrow Fa.$$

Hence we have, as desired, the shift morphism

$$\sharp a \longrightarrow Fa.$$

The rest of the Theorem follows in the same pattern as the proof of $SRT1$. //

Third Self-Reference Theorem (SRT3). Let (C, C') be a 2-category that is a lambda pair. Let F, a and α be morphisms in C. Suppose that $Fa \longrightarrow \alpha F\sharp$ where $\sharp x = xx$ for any morphism x in C. Then $Faa \longrightarrow \alpha Faa$. Thus α receives indirect self-reference from Faa.

Proof. This result is the immediate consequence of horizontal composition in the 2-category of the arrow $Fa \longrightarrow \alpha F\sharp$ and the identity arrow $a \longrightarrow a$, coupled with the fact that $\sharp a = aa$.//

Remark. This Third Self-Reference Theorem is a 2-categorical analog of the Lawvere Fixed Point Theorem [13, 14]. We will explain this connection in more detail in Section 4. The reader should note that direct application of the indicative shift to the arrow $Fa \longrightarrow \alpha F\sharp$ does not produce this self-reference.

The Smullyan Categorical Pair. An exercise related to Goedel's Theorem due to Raymond Smullyan [16] can be naturally formulated in terms of categorical pairs. The first category C consists in (as morphisms) all words in the alphabet $\{\sim, P, R\}$. where a word is any ordered string of these symbols. The category C has a single object.

Composition in C consists in concatenation of strings. The objects in the second category C' consist in strings X in the alphabet $\{\sim, P, R\}$ and strings $[X]$ enclosed in brackets. Thus every morphism in C is an object in C'. Some objects in C' are not morphisms in C. For example, $[\sim PRR]$ is an object in C'. Only the following types of arrow in C' are allowed, where X is an arbitrary string in that alphabet.

1. $PX \longrightarrow X$
2. $\sim PX \longrightarrow [X]$
3. $RX \longrightarrow XX$
4. $\sim RX \longrightarrow [XX]$

The reader will note that by substituting R for X in item 3. we obtain the self-reference

$$RR \longrightarrow RR,$$

and by substituting $\sim R$ for X in item 4. we obtain an expression that brackets its referent $\sim R \sim R \longrightarrow [\sim R \sim R]$. This category contains both self-reference and indirect self-reference.

Smullyan has an amusing interpretation of this formalism. He tells the story of a machine that prints strings from the category C (he does not use categorical terminology, but we will describe it that way). The arrows in C' are interpreted as restrictions and descriptions of the machine's actions. For codomains of arrows in C', the absence of a bracket means *printablity* and the presence of a bracket (as in $[X]$) means *unprintablity*. Meaning is mediated by an arrow in C'. The category C' contains the semantics for the categorical pair.

1. If the machine can print the string PX then it can print the string X. In other words PX means that X is printable.

$$PX \longrightarrow X.$$

2. If the machine can print the string $\sim PX$ then the string X is not printable (as an isolated string) by the machine. The string $\sim PX$ means that X (alone) is not printable.

$$\sim PX \longrightarrow [X].$$

3. The printing of the string RX means that XX is printable.

$$RX \longrightarrow XX.$$

4. The printing of the string $\sim RX$ means that XX is not printable.

$$\sim RX \longrightarrow [XX].$$

Thus we can interpret the Smullyan Machine in terms of the category C' by saying that each morphism in C' is interpreted as a delineation about the printing of its codomain (or the contents of the bracket for its codomain) in terms of the contents of its domain. Each of the special string types (lets us call them *interpretable* strings) $\{PX, \sim PX, RX, \sim RX\}$ might be printable by the machine, and if printed, they each tell what the machine can further print. It is given that *whenever the machine prints one of these special strings then it tells the truth*. We deduce that the machine cannot print the string

$$\sim R \sim R,$$

for this string asserts its own unprintability. Thus, while the Smullyan Machine always tells the truth when it prints an interpretable string, there are interpretable strings that are true but unprintable! This Smullyan categorical pair is an intriguing miniature version of Goedel's Incompleteness Theorem, with printablity replacing provability.

The Universal Building Machine. We can interpret the expansion operator E described at the beginning of this section as a universal building machine. Then "X" designates a blueprint for the construction of X. (Of course here we indulge in a heirarchy of names. Really X is the name of an actuality and "X" is the name of the blueprint for constructing this actuality.) Then we have

$$E\text{``}X\text{''} \longrightarrow X\text{``}X\text{''},$$

meaning that the universal builder E takes the blueprint "X" and produces the actuality X appended to a copy of its blueprint. The 2-categorical morphism is a morphism between the composition of the building machine and the blueprint and the composition of the actuality and its blueprint. The universal building machine will build itself when supplied with its own blueprint.

$$E\text{``}E\text{''} \longrightarrow E\text{``}E\text{''}.$$

3 Goedel's Theorem

In this section we point out how Goedel's Incompleteness Theorem can be regarded as an application of the First Self-Reference Theorem (SRT1). We let C be a category associated with a formal system S that is assumed to be consistent and rich enough to contain the standard theory of integers and first order logic. The morphisms of the category C will consist in all finite texts of S. Two morphisms can be composed if one can be substituted into a unique free variable in the other. Thus ab means that b is substituted for the unique free variable in a. Otherwise ab is not defined. We also assume that there is a given method of Goedel numbering that associates a unique natural number to each text in S.

We now define the morphism $\sharp$. For g a natural number that is a Goedel number of some text T with one free variable in S, $\sharp g$ *is a text in S that represents the Goedel* *number of* Tg. That is, $\sharp g$ represents the Goedel number of the text that results from substituting g for the free variable in T. This means that $\sharp$ itself stands for a text in the formal system S that accomplishes the program of decoding a number g to its text T and then calculating the Goedel number of Tg. Thus $\sharp g$ can be regarded, in the system S, as a representative for the Goedel number of Tg. This completes the description of the category C.

Now we describe the category C'. An arrow in C' is always of the form

$$g \longrightarrow F$$

where g is the Goedel number of a text F. Note that g may also be a text in S that represents that Goedel number. Note also that if F is a text with one free variable, then it follows from $g \longrightarrow F$ that

$$\sharp g \longrightarrow Fg$$

since $\sharp g$ represents the Goedel number of Fg. Hence the indicative shift applies in this category. We can apply the First Self-Reference Theorem and obtain that if F has one free variable then so does $F\sharp$, and if

$$g \longrightarrow F\sharp$$

then

$$\sharp g \longrightarrow F\sharp g.$$

This result says explicitly that $F\sharp g$ is a text in S that refers to its own Goedel number. One then replaces F by B where Bg stands for a text in S that asserts the *unprovability* of the proposition with Goedel number g. With this in hand we have from the First Self-Reference Theorem that

$$\sharp g \longrightarrow B\sharp g.$$

so that $B\sharp g$ asserts its own unprovability in the system S. This completes our sketch of how **SRT1** applies to Goedel's Theorem

Remark. We can place Goedel's Theorem in the context of the Second Self-Reference Theorem **SRT2.** Regard Goedel numbers g as morphisms in a category by defining gh to be the the result of substituting h in the free variable of the decoding of g (if there is such a free variable). Then we see that $\sharp g = gg$ is a concise description of the $\sharp$ operator as we have defined it above. With this definition of composition of Goedel numbers we have a category C and can construct that category C' of arrows from Goedel numbers to texts in the formal system just as we did in the above paragraphs. Now, if $g \longrightarrow F$ where F is the decoding of g, then by definition $gg = \sharp g$ is the Goedel number of Fg where Fg denotes the result of substituting g into the free variable in F. Thus we have $gg \longrightarrow Fg$ as the horizontal composition of $g \longrightarrow F$, and the identity arrow $g \longrightarrow g$ and $gg \longrightarrow Bgg$ as the horizontal composition of $g \longrightarrow B\sharp$ and the identity arrow $g \longrightarrow g$.This gives exactly the 2-categorical structure of the Second Self-Reference Theorem. The reader should note that in the category C we have both Godel numbers and texts as morphisms. In the category C' we have identity morphisms that carry numbers to numbers and texts to texts, but otherwise arrows in C' carry numbers or representatives of numbers to texts.

In this way we see clearly that the categorification of the Church-Curry Fixed Point Theorem that is implicit in the Second Self-Reference Theorem applies to Goedel's Theorem, showing how the indirect self-reference central to the Goedel construction comes from categorically changing an equality to an arrow.

4 Lawvere's Fixed Point Theorem

Lawvere's Theorem [13, 14] is a direct generalization of Cantor's diagonal argument. Recall Cantor's argument. We work in the category of sets. Let $[A, B]$ denote the collection of set theoretic mappings from A to B. Let $Z = \{0, 1\}$ and note that a subset A of a set X can be regarded as a mapping $A : X \longrightarrow Z$ where the elements of the subset are those $x \in X$ such that $Ax = 1$.

Cantor. Cantor gave a proof that *there is no surjective mapping from* X *to* $[X, Z]$. His proof goes as follows. Let $F : X \longrightarrow [X, Z]$ be any mapping. Define a subset C of X by the formula

$$Cx = \sim F(x)x$$

where it is understood that $\sim 0 = 1$ and $\sim 1 = 0$. C cannot be of the form $F(a)$ for any $a \in X$. For if $C = F(a)$, then $F(a)x = \sim F(x)x$ for all $x \in X$. Hence $F(a)a = \sim F(a)a$. This is a contradiction since the negation $\sim$ has no fixed points. From this Cantor concludes that for X infinite we have a higher infinity for $[X, Z]$ and so a heirarchy of infinities:

$$X < [X, Z] < [[X, Z], Z] < \cdots .$$

Lawvere. Lawvere turns this scenario on its head by considering a more general case where Z (the subobject classifier) could be other than the set of two elements. Then let

$$\alpha : Z \longrightarrow Z$$

be any mapping from Z to itself. Suppose that there exists a function $F : X \longrightarrow [X, Z]$ that is surjective. Define a subset C of X by the formula

$$Cx = \alpha(F(x)x).$$

Then by surjectivity, we have $C = F(a)$ for some a and consequently

$$F(a)a = \alpha(F(a)a).$$

Hence any mapping $\alpha : Z \longrightarrow Z$ must have a fixed point. This is the Lawvere Fixed Point Theorem as it is expressed in the category of sets. In fact the Theorem is easily seen to work in any category with products and a terminal object.

Lawvere's Fixed Point Theorem can be used to place Cantor's orginal argument in different contexts. For example, if we use a set theory with a subobject classifier that is not Boolean, then there can be fixed points for negation, and Cantor's argument does not prove that the mapping F is not surjective. Of course, in individual cases it may turn out that a mapping is indeed not surjective, but the Cantor proof will not show it. The diagonal set $F(a)x = \sim (F(x)x)$ is shown to be contradictory by the element $F(a)a$ because $F(a)a$ is a fixed point for negation. With a non-boolean subobject classifier, it is possible to have an element J in Z such that $\sim J = J$. With this expanded logical point of view the heirarchy of infinities does not get started.

The simplest instance of this situation occurs with the Russell Paradox where the Russell set is defined by the equation

$$Rx = \sim xx.$$

Substituting R for x we have the logical contradication

$$RR = \sim RR,$$

a contradiction because negation has no fixed point for the subobject classifier

$$Z = \{0, 1\}.$$

If we extend the subobject classifier to $Z' = \{0, 1, J\}$ where $\sim J = J$, then the Russell Set simply achieves a certain notoriety, and there is no contradiction.

Return to Self-Reference Now return to our First Self-Reference Theorem. In this context, for the Russell Set, we would generalize to a reference arrow

$$R \longrightarrow \sim \sharp.$$

Applying the shift, we obtain

$$\sharp R \longrightarrow \sim \sharp R.$$

Instead of a contradiction, we obtain a referential arrow from the $\sharp R$ to its negation. By changing equality to reference we have avoided the paradox. This is exactly how such paradox is resolved in computer languages where the referential step is often interpreted as a step in a recursive process.

Now consider the Third Self-Reference Theorem. There we have, instead of

$$F(a)x = \alpha(F(x)x)$$

the reference arrow

$$Fa \longrightarrow \alpha F\sharp.$$

If $\sharp x = xx$ and the reference arrows in C' are collapsed to equality, then we have

$$Fa = \alpha F\sharp,$$

whence

$$Fax = \alpha F \sharp x$$

so that

$$Faa = \alpha F \sharp a = \alpha Faa.$$

Thus

$$Faa = \alpha Faa,$$

and we obtain a fixed point theorem as a special case of the Third Self-Reference Theorem. The Third Self-Reference Theorem is a categorification of the Lawvere Fixed Point Theorem. In the Third Self-Reference Theorem we have assumed that the composition of morphisms in the base category C is associative, and so we have not made the same sorts of distinctions in associating compositions that are naturally present in the Lawvere Theorem. More work remains to be done in comparing these formalisms.

We end this section with a discussion of the treatment of Goedel's Incompletness Theorem in the Lawvere context and its relationship with our treatement of Goedel in the context of the First Self-Reference Theorem for categorical pairs.

Goedel Revisited. Here is how Goedel's Theorem is related to the Lawvere Fixed Point Theorem. Let $\{\phi(n,x)|n = 1,2,3,...\}$ denote a list of all syntactically valid formulas involving a single variable x in the formal system S (as described in Section 2). Suppose that S is strong enough to be able (by proving or invalidating) to determine the truth or falsehood of each particular formula $\phi(n,m)$ for all natural numbers n and m. We define a new formula by

$$Cx = \sim \phi(x,x).$$

Assuming that the list of all formulas and the ability of the formal system to determine their truth or falsity is complete, we then have $Cx = \phi(N,x)$ for some natural number N. Thus we have

$$\phi(N,x) = \sim \phi(x,x)$$

for each natural number x and hence

$$\phi(N,N) = \sim \phi(N,N).$$

Since negation has no fixed point in the standard logic of S, we conclude that any list that we make of statements for the system will be of necessity incomplete with respect to the notion of truth within the system in terms of provability. Provability within, and truth from outside the system are distinct under the assumption that the system S is consistent.

When we describe Goedel's Theorem this way it is clear that it can be seen as an application of the Lawvere Fixed Point Theorem. We simply take $F(x)y = \phi(x,y)$ and the patterns match. Note that in this form of Goedel's Theorem we did not encode directly a statement that asserts its own unprovability. The Cantorian approach to Goedel sidesteps this issue and instead shows the contradictory nature of completeness. This is the difference between our approach to Goedel via the First and Second Self-Reference Theorems and the Lawvere Fixed Point Theorem. Using the Self-Reference Theorems we construct an abstract framework for the Goedel numbering and the epistemology of indirect self-reference that is in back of the incompleteness phenomenon, and we show that this phenomenon is directly related to the higher categorical step of shifting from equality to arrow.

5 Ordinary Language

In this section we consider an interpretation for the First Self-Reference Theorem in terms of ordinary language. In this interpretation the morphisms of category C are all texts in ordinary discourse and all referents for these texts. Thus we regard perceptions and objects in the world as instances of texts in a language that encompasses the written and spoken languages that are commonly used. In this way, if I meet another person, that other person would be regarded as a text whose name I come to learn in the course of meeting him or her. Behind the aspect of a person as a text may well be the actuality of the person, but we will confine ourselves to the textual aspect. Then if I meet person P (a text) and learn his or her name N then at the beginning of that process there is indicated an arrow from N to P.

$$N \longrightarrow P$$

but shortly thereafter, when the naming process is more complete, the text that is P has become modified to contain its name in a prominent place and the name has been shifted to indicate that it is a name of *that person*. In actual practice this process is the one that includes our ability to recognise a person P as *that person with the name N.* We indicate this shift of reference by the indicative shift of Section 2.

$$\sharp N \longrightarrow PN.$$

In terms of our perception, a text P that has undergone this shift is now known by us to have the name N. The name N appears in our cognitive (representational) space along with the (text representing) the person P.

Thus we see that the notion of categorical pair and indicative shift is a model of the referential shift inherent in the naming and referring of texts in ordinary language and in language in a very general context.

The First Self-Reference Theorem then becomes a model for how self-reference occurs in language. For we see that the simplest instance of the Theorem is the act of naming the shift operation $\sharp$.

$$M \longrightarrow \sharp$$

Let M denote the name of the shift operation $\sharp$. Then M is the name of the linguistic ability to combine a name with the text to which that name refers. And we see that once that name of the shift is itself shifted, then a self-reference occurs.

$$\sharp M \longrightarrow \sharp M.$$

The completion of the naming process for the process of naming is self-referential. When we refer to ourselves in language we refer to our own ability to make and complete the act of naming.

Note how the rest of the First Self-Reference Theorem works in this context. If we have a reference

$$G \longrightarrow F\sharp,$$

this is a reference to a text $F\sharp$ that talks about the naming process. Shifting this reference we obtain

$$\sharp G \longrightarrow F\sharp G,$$

a naming of a text that discusses its own name.

We see that in the context of ordinary language a correct modeling must be flexible enough to allow even more heirarchies of reference and, at the same time to allow all these heirarchies to work at the same level since in language the name of a name is still a name. We see therefore that the splitting into two categories C and C' can lead to higher splittings (higher categories) and if these categories are all to be seen at a level, one may need to consider categories of infinite height where every object is a morphism and every morphism is an object. We call such categories *reflexive* and hope that they will be useful in an extension of this work to problems in mathematics, linguistics and philosophy.

To clarify these last remarks, consider sequences of categories

$$C\,C'\,C''\,\ldots C^{(n)}\,C^{(n+1)}\,\ldots$$

where the objects in $C^{(n+1)}$ are the morphisms in $C^{(n+1)}$. All the constructions of this paper will apply in the transition between $C^{(n)}$ and $C^{(n+1)}$. A reflexive category would be at level C^{∞} where any finite descent from morphism to object will reveal only further morphisms.

It might seem that a reflexive category would be a huge undertaking, requiring some sort of limiting construction from a heirarchy of categories. That this is not so is illustrated in Figure 1. Here we show a diagram T of a trefoil knot and take the oriented arcs of that diagram to be morphisms in a category that we shall call the *Trefoil Category*. We also illustrate a diagram T' that is not quite a knot diagram that has the same formal characteristics of morphisms pointing to morphims. Each arc is seen to be an arrow originating on one of the arcs and terminating on another. If the reader examines the Figure, it will be apparent that we have a category with objects $\{A, B, C\}$ and each of these objects is a morphism with

1. $A : C \longrightarrow B$,
2. $B : A \longrightarrow C$,
3. $C : B \longrightarrow A$.

Compositions of these morphisms are available, so this category has more morphisms than it has objects, but it is certainly reflexive in that all its objects are morphisms. Reflexive categories of this sort can be associated with knots and links. We shall study them in a separate paper. A second example is shown in Figure 1 with the link diagram L. Here the associated reflexive category has two objects A and B that are also generating morphisms for the category. We have

1. $A : B \longrightarrow B$,
2. $B : A \longrightarrow A$.

The distinct morphism/objects A and B are "linked" categorically in that each plays the role of a morphism for the other. It is clear that this notion of linking is close to the way we speak of linking in ordinary language where a linkage of plans, ideas or persons involves how each is a process for the other. One reason for bringing in this example of a reflexive category in a section on ordinary language is that we see that the Trefoil Category and the Link Category (and indeed the diagrams as mathematical structures) arise from the ordinary language of sketching of three dimensional forms. But also, when we translate these diagrammatic forms into the corresponding reflexive categories we see that the categories themselves contain patterns of language of a more general character. There is a rich mine of such categorical material in the languages of geometry, topology, art, gesture, dance and music. All these are topics to be pursued in other papers.

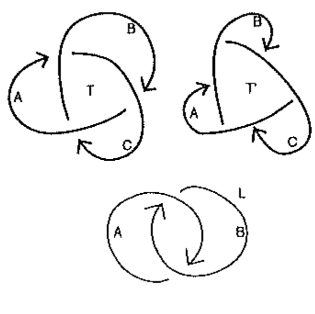


Figure 1: **Trefoil Category and Link Category**

6 Epilogue

Both the Self-Reference Theorems of this paper and the Lawvere Fixed Point Theorem come from generalizing the Cantor diagonal process, and both can also be seen as ways to generalize the Church-Curry Fixed Point Theorem. Let's look at this pattern again. In the Church-Curry Theorem we are given an algebra with a binary operation that is not associative and an *axiom of reflexivity* that states that *functions of a single variable* *expressed in that algebra can be named and regarded as elements of the algebra.* Thus in such an algebra Λ one might define $G[x] = a((bx)x)$ as a function from the algebra to itself. One is then guaranteed that there exists an element g such that for all x in the algebra, $gx = a((bx)x)$. This reflexive assumption of a correspondence between elements of the algebra and mappings of the algebra to itself is very strong.

The simplest instance of this strength is the Church-Curry Fixed Point Theorem which states that every element F of Λ has a fixed point in the sense that there is an a in the algebra such that $Fa = a$. The proof goes as follows. Define $G[x] = F(xx)$ for all x in Λ. Then, by the axiom of reflexivity there exists g in Λ such that $gx = F(xx)$ for all x. Letting $x = g$ we obtain $gg = F(gg)$. So gg is the fixed point for F.

At the formal level, the Lawvere Fixed Point Theorem can be seen as a categorical generalization of the Λ algebra formalism $C[x] = \alpha(F(x)x)$ where it is known that such a C must be represented algebraically by an element of the form $F(a)$ (the surjectivity hypothesis for F). Then we have $F(a)x = \alpha(F(x)x)$ and consequently $F(a)a = \alpha(F(a)a)$, giving α a fixed point with a specific structure. The generality of the pattern allows it to be applied to many situations beyond the original Cantor argument. The application of the Fixed Point Theorem to Goedel's Theorem works best when we do not think of Goedel's Theorem as depending on indirect self-reference.

The First and Second Self-Reference Theorems are generalizations of the Church Curry Fixed Point Theorem where we replace equality signs by arrows of reference and we correspondingly generalize the operator $\sharp x = xx$ to an arrow of reference

$$\sharp x \longrightarrow xx.$$

We then generalize the fundamental repetition operator $\sharp$ a notch further to the indicative shift where, if

$$a \longrightarrow b$$

then

$$\sharp a \longrightarrow ba$$

and the Church-Curry Fixed Point Theorem is transformed into our First Self-Reference Theorem. In fact we could take the initial category C to have one object and its morphisms the elements of the lambda algebra having either no free variable or a single free variable. Composition ab of morphisms a and b is defined whenever a has a free variable. Then ab stands for the substitution of b into the free variable in a. With this we have both the indirect reference given by the First Self-Reference Theorem (and/or the Second Self-Reference Theorem) and the fixed point results of the lambda algebra available in the one categorical pair (C, C').

7 Self-Reference

Finally we return to self-reference in the form of the expansion of a name. Recall the expansion operator as described in Section 1. We have an operation E on names that *expands* a name in the sense that if "A" is the name of A then the expansion E"A" refers to A"A", the result of appending the contents of the name to the name. Thus if we regard the name as pointing to its contents as in

$$``A" \longrightarrow A$$

then

$$E``A" \longrightarrow A``A".$$

Self-reference results when one expands the name of the expansion operator.

$$E``E" \longrightarrow E``E".$$

How is this self-reference related to the self-reference we are all familiar with in our personal experience?

To begin to see an answer to this question, consider the use of the pronoun "I". When I say I then I refer to myself. I alone does not refer to itself. It is required that there be a contents related to the one who uses the word I. I am the one who says I, and this can be said by anyone. So in a sense we can say that I is really the expansion operator and the self-reference associated with I occurs when we apply I to "I", forming I"I" which is self-referent. In other words, we each make a personal identification

$$I = I``I",$$

that says " I am the operation of expanding myself to my content (which is myself)." This was said more eloquently by Heinz von Foerster [17] : "I am the observed relation between myself and observing myself." We encourage the reader to expand further on these themes.

References

[1] H. P. Barendregt, "The Lambda Calculus - Its Syntax and Semantics", North Holland, Elsevier, 1984.

[2] L.H. Kauffman, Self-reference and recursive forms, Vol. 10, Journal of Social and Biological Structures (1987), 53-72.

[3] L.H. Kauffman, Knot Logic. In Knots and Applications ed. by L. Kauffman, World Scientific Pub. (1994), pp. 1-110.

[4] L.H. Kauffman, Space and time in computation and discrete physics, Intl. J. Gen. Systems , Vol. 27, Nos 1-3 (1998), pp. 249 -273.

[5] L.H. Kauffman, Virtual logic, Systems Research, Vol. 13 (Festschrift for Heinz von Foerster), No. 3, pp. 283-310 (1996).

[6] L.H. Kauffman, Time imaginary value, paradox sign and space. in Computing Anticipatory Systems, CASYS - Fifth International Conference, Liege, Belgium (2001) ed. by Daniel Dubois, AIP Conference Proceedings Volume 627 (2002).

[7] L.H. Kauffman, Biologic. AMS Contemporary Mathematics Series, Vol. 304, (2002), pp. 313 - 340.

[8] L.H. Kauffman, Biologic II, in "Woods Hole Mathematics" edited by Nils tongring and R. C. Penner, World Scientific Series on Knots and Everything Vo.34 (2004), p. 94-132.

[9] L.H. Kauffman, Fibonacci form and beyond. *Forma* 19 (special issue edited by Jay Kappraff) (2004), no. 4, 315–334.

[10] L.H. Kauffman, Eigenform, *Kybernetes - The Intl J. of Systems and Cybernetics* 34, No. 1/2 (2005), Emerald Group Publishing Ltd, p. 129-150.

[11] L.H. Kauffman, Reflexivity and Eigenform – The Shape of Process. Constructivist Foundations, Vol 4. No. 3, July 2009, p. 121-137.

[12] L.H. Kauffman, Reflexivity, Eigenform and Foundations of Physics. In Reflexivity, Proceedins of ANPA 30, Arleta D. Ford, Editor , Published by ANPA, June 2010, pp. 158-222.

[13] F. W. Lawvere and S. H. Shanuel, "Conceptual Mathematics - A First Introduction to Categories", Cambridge University Press 1997,2009.

[14] F. W. Lawvere, Diagonal arguments and cartesian closed categories, Reprints in Theory and Applications of Categories, No. 15, 2006, pp. 1-13.

[15] T. Leinster, "Higher Operads, Higher Categories", London Mathematical Society Lecture Notes Series No. 298, Cambridge University Press, 2004.

[16] R. Smullyan, "Goedel's Incompleteness Theorems", Oxford University Press, 1992.

[17] H. von Foerster, "Understanding Understanding - Essays on Cybernetics and Cognition," Springer-Verlag (2003), p. 257.

The Urgent Need for an Academic Revolution From Knowledge to Wisdom

Nicholas Maxwell
(Emeritus Reader at University College London)

Abstract
At present the basic intellectual aim of academic inquiry is to improve knowledge. Much of the structure, the whole character, of academic inquiry, in universities all over the world, is shaped by the adoption of this as the basic intellectual aim. But, judged from the standpoint of making a contribution to human welfare, academic inquiry of this type is damagingly irrational. Three of four of the most elementary rules of rational problem-solving are violated. A revolution in the aims and methods of academic inquiry is needed so that the basic aim becomes to promote wisdom, conceived of as the capacity to realize what is of value, for oneself and others, thus including knowledge and technological know-how, but much else besides. This urgently needed revolution would affect every branch and aspect of the academic enterprise.

Introduction

Humanity is confronted by grave global problems. Most serious of all, perhaps, there is the impending problem of global warming. There is the problem of the progressive destruction of tropical rain forests and other natural habitats, with its concomitant devastating extinction of species. There is the problem of war, over 100 million people having died in countless wars in the 20th century (which compares unfavourably with the 12 million or so killed in wars during the 19th century). There is the arms trade, the massive stockpiling of armaments, even by poor countries, and the ever-present threat of their use by terrorists or in war, whether the arms be conventional, chemical, biological or nuclear. There is the sustained and profound injustice of immense differences of wealth across the globe, the industrially advanced first world of North America, Europe and elsewhere experiencing unprecedented wealth while something like a third of humanity live in conditions of poverty in the developing world, hungry, unemployed, without proper housing, health care, education, or even access to safe water. There is the long-standing problem of the rapid growth of the world's population, especially pronounced in the poorest parts of the world, and adversely affecting efforts at development. And there is the horror of the Aids epidemic, again far more terrible in the poorest parts of the world, devastating millions of lives, destroying families, and crippling economies.

From Knowledge to Wisdom

What can be done in response to global problems such as these? There are a multitude of things that can be done, and *are* being done, with varying amounts of success. Here, I wish to concentrate on just *one* thing that could be done, which would go to the heart of the above global problems, and to the heart of our apparent current incapacity to respond adequately to these problems.

We need to bring about a wholesale, structural revolution in the aims and methods, the entire intellectual and institutional character of academic inquiry. At present academic inquiry is devoted to acquiring *knowledge*. The idea is to acquire knowledge, and then apply it to help solve social problems. This needs to change, so that the basic aim becomes to seek and promote *wisdom* – wisdom being understood to be the capacity to realize what is of value in life for oneself and others (and thus including knowledge, know-how and understanding). Instead of devoting itself primarily to solving problems of knowledge, academic inquiry needs to give intellectual priority to the task of discovering possible solutions to problems of living.

I have two arguments in support of this contention. The first appeals to problem-solving rationality, the second to aim-pursuing rationality.

The Crisis of Science without Wisdom

It may seem surprising that I should suggest that changing the aims and methods of academic inquiry would help us tackle the above global problems. It is, however, of decisive importance to appreciate that *all* the above global problems have arisen because of a massive increase in scientific knowledge and technology without a concomitant increase in global wisdom. Degradation of the environment due to industrialization and modern agriculture, global warming, the horrific number of people killed in war, the arms trade and the stockpiling of modern armaments, the immense differences in the wealth of populations across the globe, rapid population growth: all these have been made possible by the rapid growth of science and technology since the birth of modern science in the 17th century. Modern science and technology are even implicated in the rapid spread of Aids, Aids being spread by modern travel.

That the rapid growth of scientific knowledge and technological know-how should have these kinds of consequence is all but inevitable. Scientific and technological progress massively increase our power to act: in the absence of wisdom, this will have beneficial consequences, but will also have harmful ones, whether intended, as in war, or unforeseen and unintended (initially at least), as in environmental degradation. As long as we lacked modern science, lack of wisdom did not matter too much: our power to wreak havoc on the planet and each other was limited. Now that our power to act has been so massively enhanced by modern science and technology, global wisdom has become, not a luxury, but a necessity.

The crisis of our times, in short – the crisis behind all the others – is the crisis of science without wisdom. Having a kind of academic inquiry that is, by and large, restricted to acquiring knowledge can only serve to *intensify* this crisis. Changing the nature of science, and of academic inquiry more generally, is the key intellectual and institutional change that we need to make in order to come to grips with our global problems – above all, the global problem behind all the others, the crisis of ever-increasing technological power in the absence of wisdom. We urgently need a new kind of academic inquiry that gives intellectual priority to promoting the growth of global wisdom.

The Damaging Irrationality of Knowledge-Inquiry

There are those who blame scientific rationality for our problems, but that profoundly misses the point. What we are suffering from is not too much reason, but not enough. Scientific rationality, so-called, is actually a species of damaging *irrationality* masquerading as rationality. Academic inquiry as it mostly exists at present, devoted to the growth of knowledge and technological know-how – *knowledge-inquiry* I shall call it – is actually profoundly irrational when judged from the standpoint of contributing to human welfare. Judged from this all-important standpoint, knowledge-inquiry violates three of the four most elementary, uncontroversial rules of reason that one can conceive of (to be indicated in a moment). And that knowledge-inquiry is grossly irrational in this way has everything to do with its tendency to generate the kind of global problems considered above. Instead of false simulacra of reason, what we so urgently need is authentic reason devoted to the growth of wisdom.

Knowledge-inquiry demands that a sharp split be made between the social or humanitarian aims of inquiry and the *intellectual* aim. The intellectual aim is to acquire knowledge of truth, nothing being presupposed about the truth. Only those considerations may enter into the intellectual domain of inquiry relevant to the determination of truth – claims to knowledge, results of observation and experiment, arguments designed to establish truth or falsity. Feelings and desires, values, ideals, political and religious views, expressions of hopes and fears, cries of pain, articulation of problems of living: all these must be ruthlessly excluded from the intellectual domain of inquiry as having no relevance to the pursuit of knowledge – although of course inquiry can seek to develop factual knowledge about these things, within psychology, sociology or anthropology. Within natural science, an even more severe censorship system operates: an idea, in order to enter into the intellectual domain of science, must be an empirically testable claim to factual knowledge.

The basic idea of knowledge-inquiry, then, is this. First, knowledge is to be acquired; then it can be applied to help solve social problems. For this to work, authentic objective knowledge must be acquired. Almost paradoxically, human values and aspirations must be excluded from the intellectual domain of inquiry so that genuine factual knowledge is acquired and inquiry can be of genuine human value, and can be capable of helping us realize our human aspirations.[1]

[1] For a much more detailed exposition of knowledge-inquiry, or "the philosophy of knowledge", see Maxwell (1984 or 2007, chapter 2). For evidence that knowledge-

This is the conception of inquiry which, I claim, violates reason in a wholesale, structural and damaging manner.

What do I mean by "reason"? As I use the term here, rationality appeals to the idea that there are general methods, rules or strategies which, if put into practice, give us our best chance, other things being equal, of solving our problems, realizing our aims. Rationality is an aid to success, but does not guarantee success, and does not determine what needs to be done.

Four elementary rules of reason, alluded to above, are:

1. Articulate and seek to improve the articulation of the basic problem(s) to be solved.
2. Propose and critically assess alternative possible solutions.
3. When necessary, break up the basic problem to be solved into a number of *specialized* problems – preliminary, simpler, analogous, subordinate problems – (to be tackled in accordance with rules (1) and (2)), in an attempt to work gradually toward a solution to the basic problem to be solved.
4. Inter-connect attempts to solve the basic problem and specialized problems, so that basic problem-solving may guide, and be guided by, specialized problem-solving.

Two preliminary points now need to be made.

First, granted that academic inquiry has, as its fundamental aim, to help promote human welfare by intellectual and educational means,[2]

---

inquiry prevails in academia, see Maxwell (1984 or 2007, chapter 6; 2000). I do not claim that everything in academia accords with the edicts of knowledge-inquiry. My claim is, rather, that this is the only candidate for rational inquiry in the public arena; it is the dominant view, exercising an all-pervasive influence over academe. Work that does not conform to its edicts has to struggle to survive.

[2] This assumption may be challenged. Does not academic inquiry seek knowledge for its own sake – it may be asked – whether it helps promote human welfare or not? Elsewhere (Maxwell, 2007, pp. 17-19, 70-5, 205-13) I have argued that wisdom-inquiry does better justice than knowledge-inquiry to *both* aspects of inquiry, pure and applied. The basic aim of inquiry, according to wisdom-inquiry, is to help us realize what is of value in life, "realize" meaning both "apprehend" and "make real". "Realize" thus accommodates both aspects of inquiry, "pure" research or "knowledge pursued for its own sake" on the one hand, and technological or "mission-oriented" research on the other – both, ideally, seeking to contribute to what is of value in human life. Wisdom-inquiry, like sight, is

then the *problems* that inquiry fundamentally ought to try to help solve are problems of living, problems of action. From the standpoint of achieving what is of value in life, it is what we *do*, or refrain from doing, that ultimately matters. Even where new knowledge and technological know-how are relevant to the achievement of what is of value – as it is in medicine or agriculture, for example – it is always what this new knowledge or technological know-how enables us to *do* that matters.

All the global problems discussed above require, for their resolution, not merely new knowledge, but rather new policies, new institutions, new ways of living. Scientific knowledge, and associated technological know-how have, if anything, as we have seen, contributed to the creation of these problems in the first place. Thus problems of living – problems of poverty, ill-health, injustice, deprivation – are solved by what we do, or refrain from doing; they are not solved by the mere provision of knowledge (except when a problem of living *is* a problem of knowledge).

Second, in order to achieve what is of value in life more successfully than we do at present, we need to discover how to resolve conflicts and problems of living in more *cooperatively rational* ways than we do at present. There is a spectrum of ways in which conflicts can be resolved, from murder or all out war at the violent end of the spectrum, via enslavement, threat of murder or war, threats of a less extreme kind, manipulation, bargaining, voting, to cooperative rationality at the other end of the spectrum, those involved seeking, by rational means, to arrive at that course of action which does the best justice to the interests of all those involved. A basic task for a kind of academic inquiry that seeks to help promote human welfare must be to discover how conflict resolution can be moved away from the violent end of the spectrum towards the cooperatively rational end.

Granted this, and granted that the above four rules of reason are put into practice then, at the most fundamental level, academic inquiry needs to:

---

there to help us find our way around. And like sight, wisdom-inquiry is of value to us in two ways: for its intrinsic value, and for practical purposes. The first is almost more precious than the second.

1. Articulate, and seek to improve the articulation of, personal, social and global problems of living that need to be solved if the quality of human life is to be enhanced (including those indicated above);
2. Propose and critically assess alternative possible solutions – alternative possible *actions*, *policies*, *political programmes*, *legislative proposals*, *ideologies*, *philosophies of life*.

In addition, of course, academic inquiry must:

3. Break up the basic problems of living into subordinate, specialized problems – in particular, specialized problems of knowledge and technology.
4. Inter-connect basic and specialized problem-solving.

Academic inquiry as it mostly exists at present puts (3) into practice to splendid effect. The intricate maze of specialized disciplines devoted to improving knowledge and technological know-how that go to make up current academic inquiry is the result. But, disastrously, what we have at present, academic inquiry devoted primarily to improving knowledge, fails to put (1), (2) and (4) into practice. In pursuing knowledge, academic inquiry may articulate problems of knowledge, and propose and critically assess possible solutions, possible claims to knowledge – factual theses, observational and experimental results, theories. But, as we have seen, problems of *knowledge* are not (in general) problems of *living*; and solutions to problems of *knowledge* are not (in general) solutions to problems of *living*. Insofar as academia does at present put (1) and (2) into practice, in departments of social science and policy studies, it does so only at the periphery, and not as its central, fundamental intellectual task.

In short, academic inquiry devoted primarily to the pursuit of knowledge, when construed as having the basic humanitarian aim of helping to enhance the quality of human life by intellectual means, fails to put the two most elementary rules of reason into practice (rules (1) and (2)). Academic inquiry fails to do (at a fundamental level) what it most needs to do, namely (1) articulate problems of living, and

(2) propose and critically assess possible solutions. And furthermore, as a result of failing to explore the basic problems that need to be solved, academic inquiry cannot put the fourth rule of rational problem-solving into practice either, namely (4) inter-connect basic and specialized problem-solving. As I have remarked, *three* of the four most elementary rules of rational problem-solving are violated. (For a more detailed development of this argument see Maxwell, 1980, 1984 or 2007, 2004, 2010.)

This gross structural irrationality of contemporary academic inquiry has profoundly damaging consequences for humanity. As I have pointed out above, granted that our aim is to contribute to human welfare by intellectual means, the basic problems we need to solve are problems of living, problems of action, not problems of knowledge. In failing to give intellectual priority to problems of living, knowledge-inquiry fails to tackle what most needs to be tackled in order to contribute to human welfare. In devoting itself to acquiring knowledge in a way that is unrelated to sustained concern about what humanity's most urgent problems are, as a result of failing to put (1) and (2) into practice, and thus failing to put (4) into practice as well, the danger is that scientific and technological research will respond to the interests of the powerful and the wealthy, rather than to the interests of the poor, of those most in need. Scientists, officially seeking knowledge of truth *per se*, have no official grounds for objecting if those who fund research – governments and industry – decide that the truth to be sought will reflect their interests, rather than the interests of the world's poor. And priorities of scientific research, globally, do indeed reflect the interests of the first world, rather than those of the third world.[3]

Knowledge and technology successfully pursued in a way that is not rationally subordinated to the tackling of more fundamental problems of living, through the failure to put (1), (2) and (4) into practice, is bound to lead to the kind of global problems discussed above, problems that arise as a result of newly acquired powers to act being divorced from the ability to act wisely. The creation of our current global problems, and our inability to respond adequately to

[3] Funds devoted, in the USA, UK and some other wealthy countries, to military research are especially disturbing: see Langley (2005) and Smith (2003).

these problems, has much to do, in other words, with the long-standing, rarely noticed, structural *irrationality* of our institutions and traditions of learning, devoted as they are to acquiring knowledge dissociated from learning how to tackle our problems of living in more cooperatively rational ways. Knowledge-inquiry, because of its irrationality, is designed to *intensify*, not help *solve*, our current global problems.[4]

Wisdom-Inquiry

At once the question arises: What would a kind of inquiry be like that *is* devoted, in a genuinely rational way, to promoting human welfare by intellectual means? I shall call such a hypothetical kind of inquiry *wisdom-inquiry*, to stand in contrast to knowledge-inquiry.

As a first step at characterizing wisdom-inquiry, we may take knowledge-inquiry (at its best) and modify it just sufficiently to ensure that all four elementary rules of rational problem-solving, indicated above, are built into its intellectual and institutional structure: see Figure 1.

The primary change that needs to be made is to ensure that academic inquiry implements rules (1) and (2). It becomes the fundamental task of social inquiry and the humanities (1) to articulate, and seek to improve the articulation of, our problems of living, and (2) to propose and critically assess possible solutions, from the standpoint of their practicality and desirability. In particular, social inquiry has the task of discovering how conflicts may be resolved in less violent, more cooperatively rational ways. It also has the task of promoting such tackling of problems of living in the social world beyond academe. Social inquiry is, thus, not primarily social *science*, nor, primarily, concerned to acquire knowledge of the social world; its primary task is to promote more cooperatively rational tackling of problems of living in the social world. Pursued in this way, social inquiry is intellectually more fundamental than the natural and technological

[4] See Maxwell (1984 or 2007, chapter 3) for a much more detailed discussion of the damaging social repercussions of knowledge-inquiry.

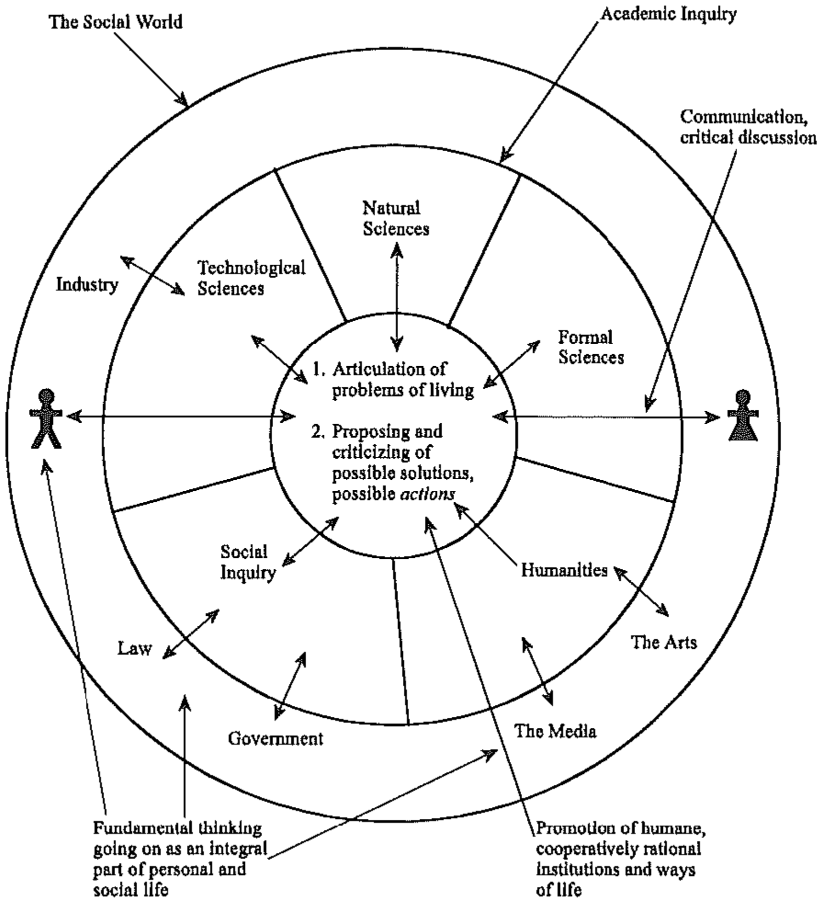


Figure 1: Wisdom-Inquiry Implementing Problem-Solving Rationality

sciences, which tackle subordinate problems of knowledge, understanding and technology, in accordance with rule (3). In Figure 1, implementation of rule (3) is represented by the specialized problem-solving of the natural, technological and formal sciences, and more specialized aspects of social inquiry and the humanities. Rule (4) is represented by the two-way arrows linking fundamental and specialized problem-solving, each influencing the other.

One can go further. According to this view, the thinking that we engage in as we live, in seeking to realize what is of value to us, is intellectually more fundamental than the whole of academic inquiry (which has, as its basic purpose, to help cooperatively rational thinking and problem-solving in life to flourish). Academic thought emerges as a kind of specialization of personal and social thinking in life, the result of implementing rule (3); this means there needs to be a two-way interplay of ideas, arguments and experiences between the social world and academia, in accordance with rule (4). This is represented, in figure 1, by the two-way arrows linking academic inquiry and the social world.

The natural and technological sciences need to recognize three domains of discussion: evidence, theory, and aims. Discussion of aims seeks to identify that highly problematic region of overlap between that which is discoverable, and that which it is of value to discover. Discussion of what it is of value to discover interacts with social inquiry, in accordance with rule (4).

The Enlightenment Programme

So much for my first argument in support of wisdom-inquiry. I come now to my second argument, which appeals to, and modifies, the Enlightenment programme of learning from scientific progress how to achieve social progress towards an enlightened world.

In order to implement this programme properly, it is essential to get the following three steps right.

1. The progress-achieving methods of science need to be correctly identified.
2. These methods need to be correctly generalized so that they become fruitfully applicable to any human endeavour, whatever the aims may be, and not just applicable to the endeavour of improving knowledge.
3. The correctly generalized progress-achieving methods then need to be exploited correctly in the great human endeavour of trying to make social progress towards an enlightened, wise, civilized world.

Unfortunately, the *philosophes* of the Enlightenment got all three points wrong. And as a result these blunders, undetected and uncorrected, are built into the intellectual-institutional structure of academia as it exists today.[5]

First, the *philosophes* failed to capture correctly the progress-achieving methods of natural science. From D'Alembert in the 18th century to Popper in the 20th (Popper, 1959, 1963), the widely held view, amongst both scientists and philosophers, has been (and continues to be) that science proceeds by assessing theories impartially in the light of evidence, *no permanent assumption being accepted by science about the universe independently of evidence.* But this standard empiricist view is untenable. If taken literally, it would instantly bring science to a standstill. For, given any accepted theory of physics, T, Newtonian theory say, or quantum theory, endlessly many empirically more successful rivals can be concocted which agree with T about observed phenomena but disagree arbitrarily about some unobserved phenomena. Physics would be drowned in an ocean of such empirically more successful rival theories.

In practice, these rivals are excluded because they are disastrously disunified. *Two* considerations govern acceptance of theories in physics: empirical success and unity. But in persistently accepting unified theories, to the extent of rejecting disunified rivals that are just as, or even more, empirically successful, physics makes a big persistent assumption about the universe. The universe is such that all disunified theories are false. It has some kind of unified dynamic structure. It is physically comprehensible in the sense that explanations for phenomena exist to be discovered.

But this untestable (and thus metaphysical) assumption that the universe is comprehensible is profoundly problematic. Science is obliged to assume, but does not know, that the universe is comprehensible. Much less does it know that the universe is comprehensible in this or that way. A glance at the history of physics

[5] The blunders of the *philosophes* are not entirely undetected. Karl Popper, in his first four works, makes substantial improvements to the traditional Enlightenment programme (although Popper does not himself present his work in this fashion). Popper first improves traditional conceptions of the progress-achieving methods of science (Popper, 1959). This conception, *falsificationism*, is then generalized to become *critical rationalism*. This is then applied to social, political and philosophical problems (Popper, 1961, 1962, 1963). The version of the Enlightenment programme about to be outlined here can be regarded as a radical improvement of Popper's version: see Maxwell (2004, chapter 3).

reveals that ideas have changed dramatically over time. In the 17th century there was the idea that the universe consists of corpuscles, minute billiard balls, which interact only by contact. This gave way to the idea that the universe consists of point-particles surrounded by rigid, spherically symmetrical fields of force, which in turn gave way to the idea that there is one unified self-interacting field, varying smoothly throughout space and time. Nowadays we have the idea that everything is made up of minute quantum strings embedded in ten or eleven dimensions of space-time. Some kind of assumption along these lines must be made but, given the historical record, and given that any such assumption concerns the ultimate nature of the universe, that of which we are most ignorant, it is only reasonable to conclude that it is almost bound to be false.

The way to overcome this fundamental dilemma inherent in the scientific enterprise is to construe physics as making a hierarchy of metaphysical assumptions concerning the comprehensibility and knowability of the universe, these assumptions asserting less and less as one goes up the hierarchy, and thus becoming more and more likely to be true: see figure 2.

In this way a framework of relatively insubstantial, unproblematic, fixed assumptions and associated methods is created within which much more substantial and problematic assumptions and associated methods can be changed, and indeed improved, as scientific knowledge improves.

Put another way, a framework of relatively unspecific, unproblematic, fixed *aims* and methods is created within which much more specific and problematic aims and methods evolve as scientific knowledge evolves. (A basic aim of science is to discover in what precise way the universe is comprehensible, this aim evolving as assumptions about comprehensibility evolve.) There is positive feedback between improving knowledge, and improving aims-and-methods, improving knowledge-about-how-to-improve-knowledge. This is the nub of scientific rationality, the methodological key to the unprecedented success of science.[6] Science adapts its nature to what it discovers

[6] Natural science has made such astonishing progress in improving knowledge and understanding of nature because it has put something like the hierarchical methodology,

about the nature of the universe (see Maxwell, 1974, 1976, 1984 or 2007, 1998, 2004, 2005).

So much for the first blunder of the traditional Enlightenment, and how to put it right.

Second, having failed to identify the methods of science correctly, the *philosophes* naturally failed to generalize these methods properly. They failed to appreciate that the idea of representing the problematic aims (and associated methods) of science in the form of a hierarchy can be generalized and applied fruitfully to other worthwhile enterprises besides science. Many other enterprises have problematic aims – problematic because aims conflict, and because what we seek may be unrealizable, undesirable, or *both*. Such enterprises, with problematic aims, would benefit from employing a hierarchical methodology, generalized from that of science, thus making it possible to improve aims and methods as the enterprise proceeds. There is the hope that, as a result of exploiting in life methods generalized from those employed with such success in science, some of the astonishing success of science might be exported into other worthwhile human endeavours, with problematic aims quite different from those of science.

Third, and most disastrously of all, the *philosophes* failed completely to try to apply such generalized, hierarchical progress-achieving methods to the immense, and profoundly problematic enterprise of making social progress towards an enlightened, wise world. The aim of such an enterprise is notoriously problematic. For all sorts of reasons, what constitutes a good world, an enlightened, wise or civilized world, attainable

indicated here, into scientific practice. Officially, however, scientists continue to hold the standard empiricist view that no untestable metaphysical theses concerning the comprehensibility and knowability of the universe are accepted as a part of scientific knowledge. As I have argued elsewhere (Maxwell, 2004, chapter 2), science would be even more successful, in a number of ways, if scientists adopted and explicitly implemented the hierarchical methodology indicated here.

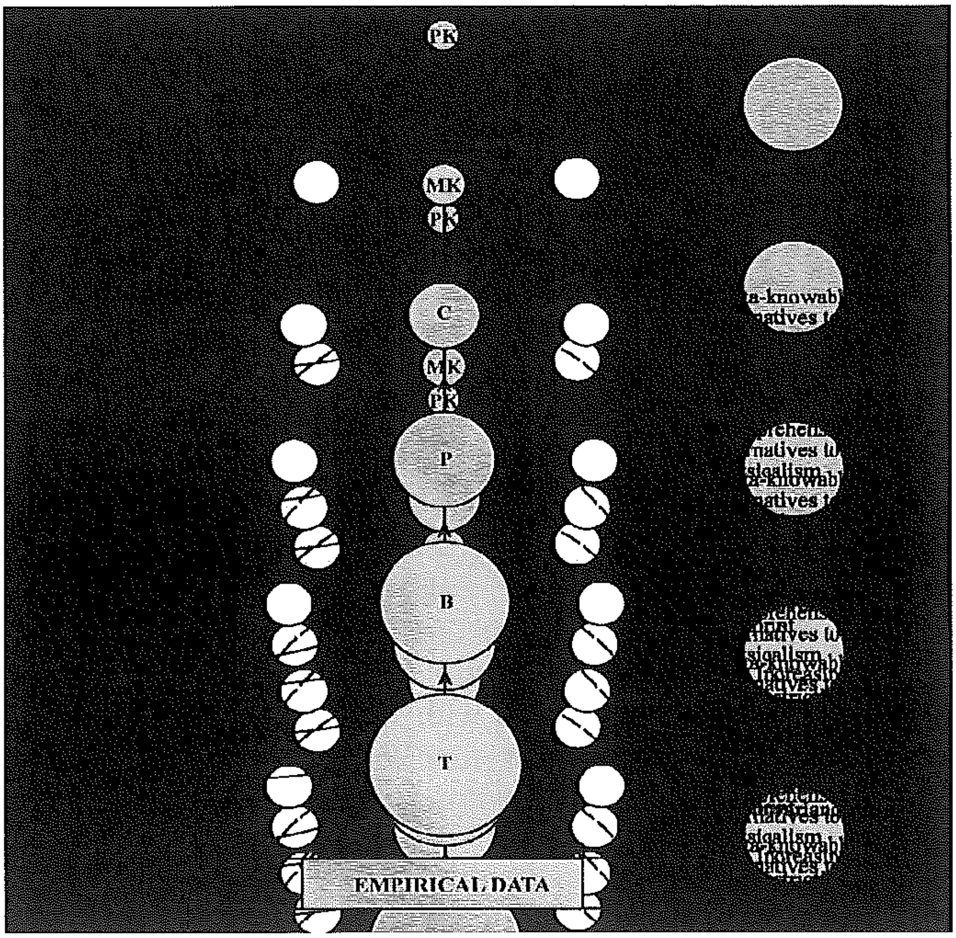


Figure 2: Hierarchical Conception of Science

and genuinely desirable, must be inherently and permanently problematic.[7] Here, above all, it is essential to employ the generalized

[7] There are a number of ways of highlighting the inherently problematic character of the aim of creating civilization. People have very different ideas as to what does constitute civilization. Most views about what constitutes Utopia, an ideally civilized society, have been unrealizable *and* profoundly undesirable. People's interests, values and ideals clash. Even values that, one may hold, ought to be a part of civilization may clash. Thus freedom and equality, even though inter-related, may nevertheless clash. It would be an odd notion of individual freedom which held that freedom was for some, and not for others; and yet if equality is pursued too single-mindedly this will undermine individual freedom, and will even undermine equality, in that a privileged class will be required to enforce equality on the rest, as in the old Soviet Union. A basic aim of legislation for civilization, we may well hold, ought to be increase freedom by restricting it: this brings

version of the hierarchical, progress-achieving methods of science, designed specifically to facilitate progress when basic aims are problematic: see Figure 3. It is just this that the *philosophes* failed to do. Instead of applying the hierarchical methodology to *social life*, the *philosophes* sought to apply a seriously defective conception of scientific method to *social science*, to the task of making progress towards, not a *better world*, but to better *knowledge* of social phenomena. And this ancient blunder is still built into the institutional and intellectual structure of academia today, inherent in the current character of social science (Maxwell, 1984 or 2007, chapters 3, 6 and 7).

Properly implemented, in short, the Enlightenment idea of learning from scientific progress how to achieve social progress towards an enlightened world would involve developing social inquiry, not as social *science*, but as social *methodology*, or social *philosophy*. A basic task would be to get into personal and social life, and into other institutions besides that of science – into government, industry, agriculture, commerce, the media, law, education, international relations – hierarchical, progress-achieving methods (designed to improve problematic aims) arrived at by generalizing the methods of science.

A basic task for academic inquiry as a whole would be to help humanity learn how to resolve its conflicts and problems of living in more just, cooperatively rational ways than at present. This task would be intellectually more fundamental than the scientific task of acquiring knowledge. Social inquiry would be intellectually more fundamental than physics. Academia would be a kind of people's civil service, doing openly for the public what actual civil services are supposed to do in secret for governments. Academia would have just sufficient power (but no more) to retain its independence from government, industry, the press, public opinion, and other centres of

---

out the inherently problematic, paradoxical character of the aim of achieving civilization. One thinker who has stressed the inherently problematic, contradictory character of the idea of civilization is Isaiah Berlin; see, for example, Berlin (1980, pp. 74-9). Berlin thought the problem could not be solved; I, on the contrary, hold that the hierarchical methodology indicated here provides us with the means to learn how to improve our solution to it in real life.

power and influence in the social world. It would seek to learn from, educate, and argue with the great social world beyond, but would not dictate. Academic thought would be pursued as a specialized, subordinate part of what is really important and fundamental: the thinking that goes on, individually, socially and institutionally, in the social world, guiding individual, social and institutional actions and life. The fundamental intellectual and humanitarian aim of inquiry would be to help humanity acquire wisdom – wisdom being the capacity to realize (apprehend and create) what is of value in life, for oneself and others, wisdom thus including knowledge and technological know-how but much else besides.

One outcome of getting into social and institutional life the kind of aim-evolving, hierarchical methodology indicated above, generalized from science, is that it becomes possible for us to develop and assess rival philosophies of life as a part of social life, somewhat as theories are developed and assessed within science. Such a hierarchical methodology provides a framework within which competing views about what our aims and methods in life should be – competing religious, political and moral views – may be cooperatively assessed and tested against broadly agreed, unspecific aims (high up in the hierarchy of aims) and the experience of personal and social life.

There is the possibility of cooperatively and progressively improving *such philosophies of life* (views about what is of value in life and how it is to be achieved) much as *theories* are cooperatively and progressively improved in science. In science, ideally, theories are critically assessed with respect to each other, with respect to metaphysical ideas concerning the comprehensibility of the universe, and with respect to *experience* (observational and experimental results).

In a somewhat analogous way, diverse philosophies of life may be critically assessed with respect to each other, with respect to relatively uncontroversial, agreed ideas about aims and what is of value, and with respect to *experience* – what we do, achieve, fail to achieve, enjoy and suffer – the aim being to improve philosophies of life (and more specific philosophies of more specific enterprises within life such as government, education or art) so that they offer greater help with the realization of what is of value in life. This hierarchical methodology is especially relevant to the task of resolving conflicts about aims and ideals, as it helps disentangle agreement (high up in the hierarchy) and disagreement (more likely to be low down in the hierarchy).

Wisdom-inquiry, because of its greater rigour, has intellectual standards that are, in important respects, different from those of knowledge-inquiry. Whereas knowledge-inquiry demands that emotions and desires, values, human ideals and aspirations, philosophies of life be excluded from the intellectual domain of inquiry, wisdom-inquiry requires that they be included. In order to discover what is of value in life it is essential that we attend to our feelings and desires. But not everything we desire is desirable, and not everything that feels good is good. Feelings, desires and values need to be subjected to critical scrutiny. And of course feelings, desires and values must not be permitted to influence judgements of factual truth and falsity.

Wisdom-inquiry embodies a synthesis of traditional Rationalism and Romanticism. It includes elements from both, and it improves on both. It incorporates Romantic ideals of integrity, having to do with motivational and emotional honesty, honesty about desires and aims; and at the same time it incorporates traditional Rationalist ideals of integrity, having to do with respect for objective fact, knowledge, and valid argument. Traditional Rationalism takes its inspiration from science and method; Romanticism takes its inspiration from art, from imagination, and from passion. Wisdom-inquiry holds art to have a fundamental rational role in inquiry, in revealing what is of value, and unmasking false values; but science, too, is of fundamental importance. What we need, for wisdom, is an interplay of sceptical rationality and emotion, an interplay of mind and heart, so that we may develop

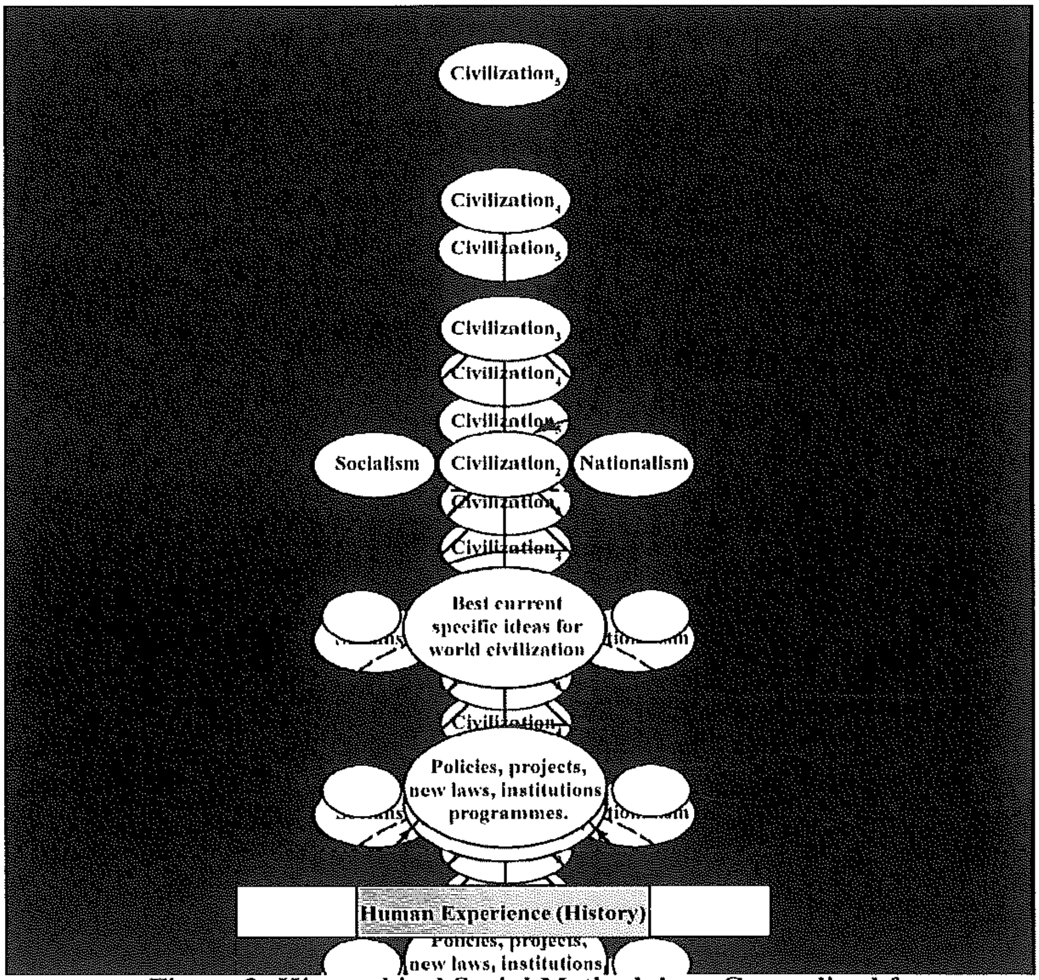


Figure 3: Hierarchical Social Methodology Generalized from Science

mindful hearts and heartfelt minds. It is time we healed the great rift in our culture, so graphically depicted by Snow (1986).

All in all, if the Enlightenment revolution had been carried through properly, the three steps indicated above being correctly implemented, the outcome would have been a kind of academic inquiry very different from what we have at present, inquiry devoted primarily to the intellectual aim of acquiring knowledge.

Conclusion

Humanity is in deep trouble. We urgently need to learn how to make progress towards a wiser, more civilized world. This in turn requires that we possess traditions and institutions of learning rationally designed – *well designed* – to help us achieve this end. It is just this that we do not have at present. What we have instead is natural science and, more broadly, inquiry devoted to acquiring knowledge. Judged from the standpoint of helping us create a better world, knowledge-inquiry of this type is dangerously and damagingly irrational. We need to bring about a major intellectual and institutional revolution in the aims and methods of inquiry, from knowledge-inquiry to wisdom-inquiry. Almost every branch and aspect of academic inquiry needs to change.

This revolution – intellectual, institutional and cultural – if it ever comes about, will be comparable in its long-term impact to that of the Renaissance, the scientific revolution, or the Enlightenment. The outcome will be traditions and institutions of learning rationally designed to help us acquire wisdom. There are a few scattered signs that this intellectual revolution, from knowledge to wisdom, is already under way. It will need, however, much wider cooperative support – from scientists, scholars, students, research councils, university administrators, vice chancellors, teachers, the media and the general public – if it is to become anything more than what it is at present, a fragmentary and often impotent movement of protest and opposition, often at odds with itself, exercising little influence on the main body of academic work. I can hardly imagine any more important work for anyone associated with academia than, in teaching, learning and research, to help promote this revolution.

References

Berlin, I., 1980, *Against the Current*, Hogarth Press, London.
Langley, C., 2005, *Soldiers in the Laboratory*, Scientists for Global Responsibility, Folkstone.
Maxwell, N., 1974, 'The Rationality of Scientific Discovery', *Philosophy of Science 41*, 1974, pp. 123-53 and 247-95.
_______, 1976, *What's Wrong With Science?*, Bran's Head Books, Frome, England (2nd edition, 2009, Pentire Press, London).
_______, 1980, 'Science, Reason, Knowledge and Wisdom: A Critique of Specialism', *Inquiry 23*, pp. 19-81.
_______, 1984, *From Knowledge to Wisdom*, Blackwell, Oxford.
_______, 1998, *The Comprehensibility of the Universe* (Oxford University Press, Oxford, pbk. 2003).
_______, 2004, *Is Science Neurotic?*, Imperial College Press, London.
_______, 2005, 'Popper, Kuhn, Lakatos and Aim-Oriented Empiricism', *Philosophia 32*, pp. 181-239.
_______, 2007, *From Knowledge to Wisdom: A Revolution for Science and the Humanities*, Pentire Press, London (2nd edition of Maxwell, 1984, revised and enlarged).
_______, 2010, *Cutting God in Half – And Putting the Pieces Together Again: A New Approach to Philosophy*, Pentire Press, London.
Popper, K. R., 1959, *The Logic of Scientific Discovery*, Hutchinson, London.
_______, 1961, *The Poverty of Historicism*, Routledge and Kegan Paul, London.
_______, 1962, *The Open Society and Its Enemies*, Routledge and Kegan Paul, London.
_______, 1963, *Conjectures and Refutations*, Routledge and Kegan Paul, London.
Smith, D., 2003, *The Atlas of War and Peace*, Earthscan, London.
Snow, C. P., 1986, *The Two Cultures: And a Second Look*, Cambridge University Press, Cambridge.

Quantum Mechanics and Physical Meaning

Peter Rowlands

Department of Physics, University of Liverpool, Oliver Lodge Laboratory, Oxford St, Liverpool, L69 7ZE, UK. email p.rowlands@liverpool.ac.uk

Abstract. It is proposed that quantum mechanics requires two vector spaces which are dual to each other, but of which only one can be observed directly. The two spaces can be seen to be responsible, through topology, for the existence of fundamental particle singularities. The dualities can be seen in a number of different forms in many aspects of fundamental physics.

Two spaces

Quantum mechanics is ultimately connected with the way that we define dynamic physical objects ('particles') as existing 'in space'. The concept of space is inconceivable without matter, yet 3-D space has no mechanism within itself for constructing the physical singularities that make up material particles. Here I am going to assume that a physical 'singularity' (e.g. a fundamental particle) can only be constructed from a commutative combination of two vector spaces. From this stems a whole range of physical properties and effects – spin ½ for fermions, *zitterbewegung*, Berry phase, etc.

Let us take a vector space (a Clifford or geometrical algebra of 3-D) described by + and – versions of 8 units:

i	**j**	**k**	vector	
*i***i**	*i***j**	*i***k**	bivector	pseudovector (quaternion)
i			trivector	pseudoscalar
1			scalar	

We now take a vector space which is commutative to the first, and can be considered dual to it (though not a 'vector space dual'):

i	**j**	**k**	vector	
*i***I**	*i***J**	*i***K**	bivector	pseudovector (quaternion)
i			trivector	pseudoscalar
1			scalar	

If we take a commutative combination of the two spaces (a tensor product), allowing for + and – signs, we obtain an algebra structured on 64 units:

i	**j**	**k**	*i***i**	*i***j**	*i***k**	*i*	1
I	**J**	**K**	*i***I**	*i***I**	*i***K**		
Ii	**Ij**	**Ik**	*i***Ii**	*i***Ij**	*i***Ik**		
Ji	**Jj**	**Jk**	*i***Ji**	*i***Jj**	*i***Jk**		
Ki	**Kj**	**Kk**	*i***Ki**	*i***Kj**	*i***Ki**		

This is recognisably the algebra of the Dirac equation, or the γ matrices, just as the algebra of a single vector space is recognisably that of the Pauli or σ matrices. In fact, all possible versions of the γ matrices can be derived from a commutative combination of two sets of σ matrices, say σ_1, σ_2, σ_3 and Σ_1, Σ, Σ_3. The units of the Dirac algebra, of course, can, of course (like the γ matrix products) be represented as a group of order 64, requiring only a *pentad* of 5 generators. Typically, they could be expressed as

K *i***Ii** *i***Ij** *i***Ik** *i***J**

There are many other ways of representing a generating pentad, but all have exactly the same structure, with the rotation symmetry of one vector space preserved (here, the lower case symbols, **i**, **j**, **k**) and with that of the other broken (here, the upper case symbols, **I**, **J**, **K**) in precisely the same way. Essentially, the space with the unbroken symmetry (**i**, **j**, **k**) is that of space as we know it, the space of observation. The one with the broken symmetry (**I**, **J**, **K**) is a dual space, which is unobservable, but which determines what happens in the observed space. This is what we can call the vacuum space.

Singularity and nilpotency

A key fact is that the act of creating a singularity using the two spaces determines that they are precisely dual and that each contains the same information as the other, though in a different form as regards observation. The reason is that physical singularities constructed from units of the form

K *i***Ii** *i***Ij** *i***Ik** *i***J**

are nilpotent or square to zero (or are constructed from nilpotents). In the case of the fundamental singularity of physics, the fermion, if we supply scalar coefficients E, p_x, p_y, p_z and m, to the algebraic units, we find that

$$(\mathbf{K}E + i\mathbf{I}\mathbf{i}p_x + i\mathbf{I}\mathbf{j}p_y + i\mathbf{I}\mathbf{k}p_z + i\mathbf{J}m)^2 = 0$$

and this can be interpreted either classically or quantum mechanically to produce the Einstein or Dirac equations.

It is this additional zero-condition which seems to create the 'singularity' state. We can interpret this as indicating that the totality of the universe is zero, and that the act of creating a fermion (with all its special energy conditions, etc.) as a singularity simultaneously creates a kind of 'hole in nothing', which we call vacuum, or the rest of the universe. We create source and sink simultaneously, one point-like and localised, and the other delocalised. Only the space related to the localised state can be observed. The asymmetry between the two spaces – unbroken and broken symmetries – is the origin of mass in two senses, which must be related, both through *zitterbewegung* and through the vacuum asymmetry involved in the Higgs mechanism.

A possible analogy between the two spaces is, if we create a knot out of two pieces of string, say coloured red and blue, but imagine that each doesn't know that the other exists (which is effectively the meaning of commutativity). We then imagine seeing the universe from the point of view of one of them, say, the blue. From the blue perspective ('blue space'), the blue string is straight, so we must devise some special contortion to create the state of the red string from the blue's perspective. The spatial 'double twist' becomes equivalent to a singularity, an additional structure within the space. Penrose has examined something similar from the point of view of twistor theory, which has a family resemblance to the algebra of the dual space in that it is constructed of four real units and four imaginary. Visually, the effect can be represented in the Robinson congruence (Penrose, 2004, p. 982, Fig. 33.15).

The nilpotent Dirac equation

A quite simple way to derive the nilpotent version of the Dirac equation is to take the Einstein relation

$$E^2 - p^2 - m^2 = 0$$

and factorise to give

$$(\pm ikE \pm i\mathbf{p} + jm)\,(\pm ikE \pm i\mathbf{p} + jm) = 0.$$

Here, the four terms in the first bracket could be considered as arranged as a row vector and the four terms in the second bracket as a column vector. We have also used the convention that $ik = \mathbf{K}$, etc. Applying a canonical quantization to the first bracket, we obtain, for a free particle, the Dirac equation

$$\left(\mp k\frac{\partial}{\partial t} \mp ii\nabla + jm\right)(\mp ikE \pm i\mathbf{p} + jm)e^{-i(Et-\mathbf{p}.\mathbf{r})} = 0\,.$$

We can also use the convention that E and $\mathbf{p}$ represent operators as well as amplitudes to express it as

$$(\pm ikE \pm i\mathbf{p} + jm)\,(\pm ikE \pm i\mathbf{p} + jm)\; e^{-i(Et-\mathbf{p}.\mathbf{r})}.$$

The significant fact now is that the amplitude is always nilpotent. Using this convention, we can replace E and $\mathbf{p}$ with covariant derivatives (e.g. $\partial / \partial t + e\phi + ...$, $- \nabla - e\mathbf{A} + ...$), or incorporate any number of field terms after the differential operators. The expression

$$(\pm ikE \pm i\mathbf{p} + jm)\,(\pm ikE \pm i\mathbf{p} + jm) \to 0$$

will still apply. The phase term will no longer be a simple exponential, but we can assume that the amplitude will continue to be nilpotent.

In principle, this means that we can do relativistic quantum mechanics for a fermion in any state, subject to any number of interactions, simply by defining an operator of the form

$$(\pm ikE \pm i\mathbf{p} + jm)$$

This will then uniquely determine the phase factor which makes the amplitude nilpotent. There is no need to define any equation at all.

$$(\text{operator acting on phase factor})^2 = \text{amplitude}^2 = 0$$

Adopting an appropriate sign convention, we can now, for example, specify the four amplitude states as derived from four creation operators:

$(ikE + i\mathbf{p} + jm)$	fermion spin up
$(ikE - i\mathbf{p} + jm)$	fermion spin down
$(-ikE + i\mathbf{p} + jm)$	antifermion spin down
$(-ikE - i\mathbf{p} + jm)$	antifermion spin up

With this convention, $i\mathbf{p} / ikE$ represents the same helicity or handedness as $(-i\mathbf{p}) / (-ikE)$, but the opposite helicity to $(i\mathbf{p}) / (-ikE)$ or $(-i\mathbf{p}) / (ikE)$.

Defining a dual space spinor

A set of primitive idempotents constructing a spinor can be defined in terms of the two spaces:

$$\begin{array}{l} 1 - \mathbf{Ii} - \mathbf{Jj} - \mathbf{Kk} \\ 1 - \mathbf{Ii} + \mathbf{Jj} + \mathbf{Kk} \\ 1 + \mathbf{Ii} - \mathbf{Jj} + \mathbf{Kk} \\ 1 + \mathbf{Ii} + \mathbf{Jj} - \mathbf{Kk} \end{array} \qquad (1)$$

This formulation arose out of discussions with J. B. Almeida.

The complete duality of the two spaces is immediately apparent. The 4 terms are orthogonal as well as idempotent, multiplying to 0. Another aspect of the construction is that **Ii**, **Jj**, **Kk** multiply commutatively, although producing a change of sign in products such as $\mathbf{IiJj} = -\mathbf{Kk}$. The idempotents defined in (1) also imply that the system is asymmetric or chiral, since a change of all the signs does not produce idempotents. We can only change the sign of two of **Ii**, **Jj**, **Kk**, say **Ii** and **Kk**, the result being a simple switching of the *order* of the idempotents:

$$\begin{array}{l} 1 - \mathbf{Ii} - \mathbf{Jj} - \mathbf{Kk} \\ 1 - \mathbf{Ii} + \mathbf{Jj} + \mathbf{Kk} \\ 1 + \mathbf{Ii} - \mathbf{Jj} + \mathbf{Kk} \\ 1 + \mathbf{Ii} + \mathbf{Jj} - \mathbf{Kk} \end{array} \qquad (2)$$

If we introduce the alternative definition of the vectors $\mathbb{I}$, $\mathbf{J}$, $\mathbb{K}$ as complexified quaternions, say, by making $\mathbb{I}$, $\mathbf{J}$, $\mathbb{K}$ become *ii*, *ij*, *ik*, then we can introduce these new definitions into (1) as

$$\begin{array}{l} 1 - ii\mathbf{i} - ij\mathbf{j} - ik\mathbf{k} \\ 1 - ii\mathbf{i} + ij\mathbf{j} + ik\mathbf{k} \\ 1 + ii\mathbf{i} - ij\mathbf{j} + ik\mathbf{k} \\ 1 + ii\mathbf{i} + ij\mathbf{j} - ik\mathbf{k} \end{array} \qquad (3)$$

while still retaining the primitive idempotent properties. We can do the same to (2), which will now become:

$$\begin{array}{l} 1 + ii\mathbf{i} - ij\mathbf{j} + ik\mathbf{k} \\ 1 + ii\mathbf{i} + ij\mathbf{j} - ik\mathbf{k} \\ 1 - ii\mathbf{i} - ij\mathbf{j} - ik\mathbf{k} \\ 1 - ii\mathbf{i} + ij\mathbf{j} + ik\mathbf{k} \end{array} \qquad (4)$$

In both (2) and (4) we see the chirality in a particularly distinctive form which will manifest itself as a physical chirality, and which ultimately stems from the fact that one of the two vector spaces **I**, **J**, **K** and **i**, **j**, **k** is observed while the other is not.

Now we may imagine the spinors (3) and (4) applied to the 'pre-spinor' nilpotent wavefunction, say (ikE + $\mathbf{ip}$ + jm) to produce the 4 distinctive terms:

$$\begin{pmatrix} ikE + \mathbf{ip} + jm \\ ikE - \mathbf{ip} + jm \\ -ikE + \mathbf{ip} + jm \\ -ikE - \mathbf{ip} + jm \end{pmatrix}$$

or even the 4 distinct 4-component wavefunctions:

$$\begin{pmatrix} ikE + \mathbf{ip} + jm \\ ikE - \mathbf{ip} + jm \\ -ikE + \mathbf{ip} + jm \\ -ikE - \mathbf{ip} + jm \end{pmatrix} \begin{pmatrix} ikE - \mathbf{ip} + jm \\ ikE + \mathbf{ip} + jm \\ -ikE - \mathbf{ip} + jm \\ -ikE + \mathbf{ip} + jm \end{pmatrix} \begin{pmatrix} -ikE + \mathbf{ip} + jm \\ -ikE - \mathbf{ip} + jm \\ ikE + \mathbf{ip} + jm \\ ikE - \mathbf{ip} + jm \end{pmatrix} \begin{pmatrix} -ikE - \mathbf{ip} + jm \\ -ikE + \mathbf{ip} + jm \\ ikE - \mathbf{ip} + jm \\ ikE + \mathbf{ip} + jm \end{pmatrix} \qquad (5)$$

which may be abbreviated as:

$$\begin{pmatrix} \pm ikE \pm \mathbf{ip} + jm \\ \pm ikE \mp \mathbf{ip} + jm \\ \mp ikE \pm \mathbf{ip} + jm \\ \mp ikE \mp \mathbf{ip} + jm \end{pmatrix}. \qquad (6)$$

Here we may associate one of the dual vector spaces, say, **i**, **j**, **k**, as locating the position in the wavefunction's column vector to which the other will apply the transitions. Alternatively, we may apply them to the 4-component wavefunctions (6) determined by the lead terms in (5).

Suppose we apply the first terms in (3) and (4) to pre- and post-multiply the 4 identical terms in a column vector representing the 'pre-spinor' wavefunction:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -ik\mathbf{k} & 0 & 0 \\ 0 & 0 & -ii\mathbf{i} & 0 \\ 0 & 0 & 0 & -ij\mathbf{j} \end{pmatrix} \begin{pmatrix} ikE+\mathbf{i}\mathbf{p}+jm \\ ikE+\mathbf{i}\mathbf{p}+jm \\ ikE+\mathbf{i}\mathbf{p}+jm \\ ikE+\mathbf{i}\mathbf{p}+jm \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & ik\mathbf{k} & 0 & 0 \\ 0 & 0 & ii\mathbf{i} & 0 \\ 0 & 0 & 0 & -ij\mathbf{j} \end{pmatrix}$$

We may effectively assume that the columns in the first 4 × 4 matrix bear the coefficients 1, **k**, **i**, **j** and the rows *i*, *j*, *k*, the position being reversed in the second 4 × 4 matrix. The coefficients in the rows of the column vector (not shown explicitly) are 1, **k**, **i**, **j**, and the operation results in a row vector with the coefficients

$$(ikE+\mathbf{i}\mathbf{p}+jm \quad ikE-\mathbf{i}\mathbf{p}+jm \quad -ikE+\mathbf{i}\mathbf{p}+jm \quad -ikE-\mathbf{i}\mathbf{p}+jm).$$

which is the full 'spinor' form of the nilpotent wavefunction.

We can now imagine the second terms in (3) and (4) applied to, say, $(ikE - \mathbf{i}\mathbf{p} + jm)$ to give

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & ik\mathbf{k} & 0 & 0 \\ 0 & 0 & -ii\mathbf{i} & 0 \\ 0 & 0 & 0 & ij\mathbf{j} \end{pmatrix} \begin{pmatrix} ikE+\mathbf{i}\mathbf{p}+jm \\ ikE+\mathbf{i}\mathbf{p}+jm \\ ikE+\mathbf{i}\mathbf{p}+jm \\ ikE+\mathbf{i}\mathbf{p}+jm \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -ik\mathbf{k} & 0 & 0 \\ 0 & 0 & ii\mathbf{i} & 0 \\ 0 & 0 & 0 & ij\mathbf{j} \end{pmatrix} =$$

$$(-ikE+\mathbf{i}\mathbf{p}+jm \quad -ikE-\mathbf{i}\mathbf{p}+jm \quad ikE-\mathbf{i}\mathbf{p}+jm \quad ikE-\mathbf{i}\mathbf{p}+jm).$$

and so on. The chirality of the m term becomes apparent in this process. (It is significant to the full duality of the spaces that we could have obtained the same results using the complexified quaternion specification for **i**, **j**, **k** rather than **I**, **J**, **K**.)

Pre-multiplication by the term (3) alone shows the creation of the three distinct idempotents that become the vacuum terms that accompany the lead term in the nilpotent formulation. In principle, the nilpotent formulation (for a free particle)

$$[(ik\partial / \partial t + \boldsymbol{i}\nabla + \boldsymbol{j}m)]\ [(ikE + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ e^{-i(Et - \mathbf{p.r})}] = 0$$
$$\textit{operator} \qquad\qquad \textit{wavefunction}$$

can be expressed in three idempotent forms

$$[(ik\partial / \partial t + \boldsymbol{i}\nabla + \boldsymbol{j}m)\boldsymbol{k}]\ [\boldsymbol{k}(ikE + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ e^{-i(Et - \mathbf{p.r})}] = 0$$

$$[(ik\partial / \partial t + \boldsymbol{i}\nabla + \boldsymbol{j}m)\boldsymbol{i}]\ [\boldsymbol{i}(ikE + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ e^{-i(Et - \mathbf{p.r})}] = 0$$

$$[(ik\partial / \partial t + \boldsymbol{i}\nabla + \boldsymbol{j}m)\boldsymbol{j}]\ [\boldsymbol{j}(ikE + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ e^{-i(Et - \mathbf{p.r})}] = 0$$

that each supply a product zero pairing between their components. These alternative representations of the same object are a direct expression of the *zitterbewegung*, which means that the fermion switches continually between its real and vacuum representations, producing the fixed mass term which expresses the chirality of the spinor process.

Spinors and the Berwald-Moor metric

If we now take the zero product of the four terms in (1):

$$(1 - \mathbf{ii} - \mathbf{jj} - \mathbf{kk})(1 - \mathbf{ii} + \mathbf{jj} + \mathbf{kk})(1 + \mathbf{ii} - \mathbf{jj} + \mathbf{kk})(1 + \mathbf{ii} + \mathbf{jj} - \mathbf{kk}) = 0 \qquad (7)$$

we see that it has the form of a singularity in a Berwald-Moor type metric of a Finsler space

$$(x_1 - x_2 - x_3 - x_4)(x_1 - x_2 + x_3 + x_4)(x_1 + x_2 - x_3 + x_4)(x_1 + x_2 + x_3 - x_4) \qquad (8)$$

Here x_1, x_2, x_3, x_4 can be regarded as base units of the dual vector ('spinor') space formed by 1, **ii**, **jj**, **kk** (with volume unit –1). The line element is the fourth root of the metric (8), and, as (7) can be resolved into three cases of nontrivial pairing of zero products, in addition to the trivial case of location at the origin, it would appear that, in these cases, the line element becomes a nilpotent. The more classic type of Berwald-Moor metric

$$(x_1 + x_2 + x_3 + x_4)(x_1 - x_2 + x_3 + x_4)(x_1 + x_2 - x_3 + x_4)(x_1 + x_2 + x_3 - x_4)$$

would also produce a singularity in this spinor space, but in this case the *squared* line element would be a nilpotent, rather than the line element itself.

True physical singularities (fermions or their products) appear to require a dual vector space which nevertheless produces an asymmetry or chirality in the space of observation because of its combination in the asymmetric nilpotent structure with the unobserved dual vacuum space. This has many manifestations including spin ½, *zitterbewegung*, Berry phase and nonzero rest mass for fermionic states. Because the spin chirality of fermions emerges through exactly the same process as the chirality producing mass via *zitterbewegung*, essentially because the spinor process produces an *observed* asymmetry between vector spaces that are dual in their original formulation, it becomes apparent that the *zitterbewegung* mass is exactly that produced by the chirality of vacuum space in the Higgs mechanism.

The fermion and the universe

A fermion can be specified as a set of space and time variations. The phase factor in the wavefunction is then, simply and precisely, the expression of all possible variations in space and time which are encoded in the creation operator, and it is uniquely defined as soon as the operator is specified. The mass term can be seen as purely passive, convenient, rather than necessary information. The nilpotent structure, howwver, ensures that we can no longer define an isolated system. Defining a fermion 'disturbs the universe'. The energy conservation principle only works over the entire universe, since the E and $\mathbf{p}$ terms contain the entire range of the fermion's interactions. The nilpotent fermion defines an *intrinsically dissipative* system; the first, second and third laws of thermodynamics become a necessary consequence of the definition.

The nilpotent method allows an immediate transformation to quantum field theory, without the cumbersome apparatus of quantum field integrals, etc. For complete specification, it requires the interaction between the fermion and the rest of the universe. The four components of the spinor wavefunction are creation / annihilation operators for the four possible fermion / antifermion states, and these act on the entire universe to create the states. The single phase factor makes calculation relatively easy. It also corresponds with Feynman's interpretation of negative energy states requiring reversed time.

The symmetry between E, **p** and m is determined by the coefficients k, i, j, thus making clear the significance of the second vector space. A particle with a nilpotent wavefunction will be automatically Pauli exclusive, as the combination state with an identical particle $\psi_1 \psi_1$ will be zero. Such wavefunctions are also Pauli exclusive in the conventional sense of being automatically antisymmetric ($\psi_1 \psi_2 - \psi_2 \psi_1$):

$$\begin{aligned}&(\pm ikE_1 \pm i\mathbf{p}_1 + jm_1)(\pm ikE_2 \pm i\mathbf{p}_2 + jm_2)\\ &-(\pm ikE_2 \pm i\mathbf{p}_2 + jm_2)(\pm ikE_1 \pm i\mathbf{p}_1 + jm_1)\\ &= 4\mathbf{p}_1\mathbf{p}_2 - 4\mathbf{p}_2\mathbf{p}_1 = 8\, i\, \mathbf{p}_1 \times \mathbf{p}_2 = -8\, i\, \mathbf{p}_2 \times \mathbf{p}_1\end{aligned}$$

We see immediately that

$$(\psi_1 \psi_2 - \psi_2 \psi_1) = -(\psi_2 \psi_1 - \psi_1 \psi_2)$$

This result has a quite astonishing consequence. It implies that, instantaneously, any nilpotent wavefunction must have a **p** vector in 'spin space', or *vacuum space*, at a *different orientation* to any other. We can now imagine the wavefunctions of all nilpotent fermions instantaneously correlating because the planes of their **p** vector directions must all intersect. The intersections then actually create the *meaning* of Euclidean space, with an intrinsic spherical symmetry generated by the fermions themselves.

Here, we see the nilpotent condition defined in two different spaces. In the case of the vacuum space we might imagine nilpotency as creating a unique direction on a set of axes k, i, j or **K**, **I**, **J** defined by the values of E, **p** and m, while the real space condition is defined by the axes **i**, **j**, **k**. In the real space, the **p** vector carries *all the information* available to a fermionic state, its direction also determining its E and **p** values uniquely. In the vacuum representation, half of the possibilities on one axis (those with $-m$) would be eliminated automatically (as being in the same direction as those with m); also those with zero m (since the directions would all be along the line $E = p$). Such hypothetical massless particles would be impossible, in addition, for fermions and antifermions with the same helicity, as E, p has the same direction as $-E, -p$.

Using discrete differentiation

We have previously (Rowlands, 2007) used a 'discrete' or anticommutative process of differentiation, as defined by Lou Kauffman (2004), where

$$\frac{\partial F}{\partial t} = [F, \mathcal{H}] = [F, E] \quad \text{and} \quad \frac{\partial F}{\partial X_i} = [F, P_i],$$

to remove both the phase factor from the amplitude and the mass term from the operator. The amplitude is now

$$\psi = ikE + i\mathbf{i}P_1 + i\mathbf{j}P_2 + i\mathbf{k}P_3 + jm$$

and the operator

$$\mathcal{D} = ik\frac{\partial}{\partial t} + i\mathbf{i}\frac{\partial}{\partial X_1} + i\mathbf{j}\frac{\partial}{\partial X_2} + i\mathbf{k}\frac{\partial}{\partial X_3}$$

where $\frac{\partial \psi}{\partial t} = [\psi, \mathcal{H}] = [\psi, E]$ and $\frac{\partial \psi}{\partial X_i} = [\psi, P_i]$.

After some algebraic manipulation, we find that

$$-\mathcal{D}\psi = i\psi(ikE + i\mathbf{i}P_1 + i\mathbf{j}P_2 + i\mathbf{k}P_3 + jm) + i(ikE + i\mathbf{i}P_1 + i\mathbf{j}P_2 + i\mathbf{k}P_3 + jm)\psi - 2i(E^2 - P_1^2 - P_2^2 - P_3^2 - m^2).$$

For nilpotent ψ, we obtain

$$\mathcal{D}\psi = \left(ik\frac{\partial}{\partial t} + i\nabla\right)\psi = 0.$$

as a Dirac equation using discrete differentials. Generalised to four states, with $\mathcal{D}$ and ψ representing 4-spinors, this creates the equivalent to the Dirac equation in this calculus:

$$\mathcal{D}\psi = \left(\pm ik\frac{\partial}{\partial t} \pm i\nabla\right)(\pm ikE \pm i\mathbf{i}P_1 \pm i\mathbf{j}P_2 \pm i\mathbf{k}P_3 + jm) = 0$$

While canonical quantization usually requires i or $i\hbar$ in defining the differentials, the lack of a mass term in the operator here means that this is no longer an absolute requirement, and we are able to use an operator that does not distinguish between quantum and discrete classical contexts. The four creation and annihilation operators (fermion / antifermion, spin up / spin down) are now exact negatives of each other, thus emphasizing the fact already established that the 'active' parts in the process are the space and time variations. If we replace the differentials with covariant derivatives, we can even introduce a distorted space-time without a mass term.

What is vacuum?

One of the key aspects of the nilpotent formalism is its intrinsic holism. Our understanding of the universe cannot be separated from the way we define a fermion, and it gives a specific meaning to the concept of *vacuum*. We imagine creating a fermion *ab initio*, that is, from entirely nothing, but with certain interaction potentials embedded in its operator. If this fermion is assigned a 'wavefunction' ψ_f, then the 'hole' in 'nothing' left by its creation can be described as $-\psi_f$; and both the superposition of the states of fermion and 'hole' ($\psi_f - \psi_f$) and their combination state ($-\psi_f \psi_f$) vanish for a nilpotent wavefunction.

'Vacuum' is simply the 'hole' in nothing left by the creation of a fermion. However, the interaction potentials that define the particular fermion state created mean that the vacuum has to supply an 'environment' with makes this possible (essentially, a system of other fermions which create those potentials). So, if the entire universe adds up to nothing, the vacuum state for a fermion does not. It becomes *the rest of the universe* for that state, and it is 'object' that the fermionic operator is actually acting on. However, both the nilpotent structure and Pauli exclusion ensure that this 'rest of the universe' must be so constructed as to appear in total as the mirror image of the fermion state.

The same condition must apply at the same time for each of the other fermions which collectively create the vacuum state for the first fermion. So, Pauli exclusion principle also means that no two fermions share the same vacuum (or the same phase factor). The zero condition for the entire universe required by the nilpotent formalism is logically satisfying because it is necessarily *incapable* of further explanation. It also provides the strongest possible constraint for defining the laws of physics, with vacuum now an active component of the theory.

Nilpotency can thus be seen as a statement of a *physical* principle, rather than a purely mathematical operation. A fermion 'constructs' its own vacuum, or the entire 'universe' in which it operates, and we are able to make a sharp distinction between the local and nonlocal, which sees them as alternate sources of the same information. The vacuum is the region of nonlocality, just as the fermion is the region of locality. Mathematically, the 'local' is whatever happens inside the nilpotent bracket structure ($\pm\, ikE \pm i\mathbf{p} + jm$), and the 'nonlocal' whatever happens outside it. Superposition and combination are nonlocal processes, but potential terms attached to E or $\mathbf{p}$ are local.

A Wheeler-type 'one fermion' theory of the universe might well make sense in such a context (Feynman, 1972). As we shall show in the section on 'Supersymmetry and renormalization', the objections involving the annihilation of matter with antimatter that were once thought to hold against it are no longer relevant. However, a single fermion cannot be defined as isolated. It must be interacting, construct, in effect, a 'space', so that its vacuum is not localised on itself. A point-like fermion must be accompanied by a dispersed vacuum. In this sense, a single (noninteracting) fermion cannot exist. We can only define if we also define its vacuum. We can only define the local if we also define the nonlocal.

Zitterbewegung and Berry phase

The spin ½ for the fermion is a routine derivation (Rowlands, 2007) from the commutator $[\sigma, H]$ for the Hamiltonian $H = (i\mathbf{p} + jm)$, and the defined pseudovector quantity $\sigma = -\mathbf{1}$, the factor 2 coming from the anticommutativity of the components of $\mathbf{p}$. If we regard a fermion as only being created simultaneously with its mirror image vacuum state, then we can regard the spin ½ term as an indication that taking the fermion alone only gives us half of the knowledge we require to specify the system.

As is well known, an unexpected additional term emerges in the solution for the equation of motion of the Dirac velocity operator, $c\mathbf{a} = -ij\mathbf{1}$. This term predicts a violent oscillatory motion or high-frequency vibration of the fermion at its Compton frequency and directly determined by the particle's rest mass. This *zitterbwegung* always been interpreted as a switching between the fermion's four spin states. It is certainly a vacuum effect, and can be seen as a switching by the fermion between real and vacuum spaces.

Dirac (1958) has interpreted *zitterbewegung* as implying that a fermion (or any massive particle) actually propagates along the light cone, oscillating between $+c$ and $-c$ at a frequency which determines its measured mass and momentum. This is because a measured value of velocity can only be found by knowing positions at two different times. To find the instantaneous velocity, the time interval must be reduced to zero, thus fixing the positions with exact precision, and hence making the momentum value completely indeterminate. The ultimate significance of *zitterbewegung* in this context may be that it locates rest mass as the result of defining a singularity.

Zitterbewegung can thus be seen as an intrinsic aspect of defining a fermion as a point-singularity through the nilpotent structure created by dual vector spaces. A related effect can be seen in those Berry phase phenomena which involve a fermion with half-integral spin subjected to a cyclic adiabatic process becoming single valued in the presence of either another fermionic state, for example, an electron (Cooper pairing) or nucleus (Jahn-Teller effect), or an 'environment' whose origin is ultimately fermionic. This could be, for example a vector potential (Aharonov-Bohm effect) or a flux line (quantum Hall effect). In each of these cases the Berry phase can be interpreted topologically, with the initial fermion travelling in a space that has changed from being simply- to multiply-connected by incorporating the other fermionic state or environment as a 'singularity.

In fact, if the electron is a singularity, it exists in its own multiply-connected space, and creates its own Berry phase. Now, a physical singularity can only be defined with reference to a nonlocalised phase. It is the nonlocalised phase which enables two such singularities to interact, and which allows us to describe such interactions in terms of a quantum field. Information from the dual spaces of one system (potentials or even distortions of its space-time structure) creates changes in the dual spaces of the other, via changes in the E and $\mathbf{p}$ terms of its operator, and, through the phase factor, of its amplitude. Even a pure vector potential (as in the Aharonov-Bohm effect) will alter the $\mathbf{p}$ term and so produce these changes. Under cyclic adiabatic conditions, we can consider the E and p magnitudes of the combination to be equalised as in the formation of a bosonic-type state.

Spherical symmetry and the Coulomb force

One of the most important aspects of the nilpotent structure (with its pseudoscalar, vector and scalar components) is that it *already incorporates the fundamental interactions*. By defining a nilpotent fermion using this mathematical formalism we are *necessarily* incorporating the fundamental interactions as part of the structure. Essentially, the Coulomb interaction terms arise from defining fermions as point sources with spherical symmetry; the characteristic strong interaction behaviour comes from defining a combination state to accommodate the explicit vector behaviour **p**; while the characteristic weak interaction harmonic oscillator can be seen in the 4-component spinor structure.

To define an operator for a fermion in the field of point source with intrinsic spherical symmetry, we may use polar coordinates for the ∇ term, using the standard prescription provided by Dirac (1958):

$$(ikE - ii\nabla + jm) \rightarrow \left(ikE - ii\left(\frac{\partial}{\partial r} + \frac{1}{r} \pm i\frac{j + 1/2}{r} \right) + jm \right)$$

The only way in which a nilpotent solution can now be obtained from such an operator is if the $1/r$ term with coefficient $\boldsymbol{i}$ is matched by a similar one with coefficient $\boldsymbol{k}$. So, requiring *spherical symmetry* as an intrinsic property of a point particle source, means that a term of the form A/r must be added to E. This means that the minimum nilpotent operator must be of the form

$$\left(\pm ik\left(E - \frac{A}{r} \right) \mp ii\left(\frac{\partial}{\partial r} + \frac{1}{r} \pm i\frac{j + 1/2}{r} \right) + jm \right)$$

To find a nilpotent solution, we must now find the phase factor to which this operator must apply to create a nilpotent amplitude, and finding the solution for this case will provide a template for all other cases. In fact, as is well known, the phase factor is of the form

$$\phi = e^{-ar_v^{\,\gamma}} \sum_{v=0} a_v r^v$$

Applying this and equating the squared amplitude to zero, we find

$$4\left(E-\frac{A}{r}\right)^2 = -2\left(-a+\frac{\gamma}{r}+\frac{\nu}{r}+\ldots\frac{1}{r}+i\frac{j+½}{r}\right)^2$$

$$-2\left(-a+\frac{\gamma}{r}+\frac{\nu}{r}+\ldots\frac{1}{r}-i\frac{j+½}{r}\right)^2 + 4m^2.$$

The full solution is now found in just a few lines. Equating constant terms, we obtain:

$$a = \sqrt{m^2 - E^2}\ .$$

Equating terms in $1/r^2$, with $n = 0$:

$$\left(\frac{A}{r}\right)^2 = -\left(\frac{\gamma+1}{r}\right)^2 + \left(\frac{j+½}{r}\right)^2.$$

Assuming the power series terminates at n', and equating coefficients of $1/r$ for $n = n'$:

$$2EA = -2\sqrt{m^2 - E^2}\,(\gamma + 1 + n'),$$

and
$$\frac{E}{m} = \frac{1}{\sqrt{1+\dfrac{A^2}{(\gamma+1+n')^2}}} = \frac{1}{\sqrt{1+\dfrac{A^2}{\left(\sqrt{(j+½)^2 - A^2} + n'\right)^2}}}\ .$$

When $A = Ze^2$ we have the relativistic 'hydrogen atom' or Coulomb force solution in just 6 lines! It is derived purely from defining fermions as (discrete) point sources with spherical symmetry. Though it is associated particularly with the electromagnetic force, all known interactions involving fermions (electric, strong, weak, gravitational) have a Coulomb term (with $U(1)$ symmetry), relating to the quantity known as the coupling constant for the interaction. It is a purely scalar term, depending on the scalar values of the terms (iE, **p**, m) in the nilpotent operator.

***P*, *T*, *C* transformations and bosonic states**

The fundamental symmetry transformations of particle physics, parity (P), time-reversal (T) and charge conjugation (C), can be accomplished pre- and post-multiplication of the nilpotent operator by the quaternion units, ***i***, ***k***, ***j***.

P	$\boldsymbol{i}\,(ikE + \boldsymbol{i}\mathbf{p} + jm)\,\boldsymbol{i} = (ikE - \boldsymbol{i}\mathbf{p} + jm)$
T	$\boldsymbol{k}\,(ikE + \boldsymbol{i}\mathbf{p} + jm)\,\boldsymbol{k} = (-ikE + \boldsymbol{i}\mathbf{p} + jm)$
C	$-\boldsymbol{j}\,(ikE + \boldsymbol{i}\mathbf{p} + jm)\,\boldsymbol{j} = (-ikE - \boldsymbol{i}\mathbf{p} + jm)$

From this, it is easy to show that $CP \equiv T$, $PT \equiv C$, $CT \equiv P$, and $CPT \equiv$ identity. The operations can be done in any order.

C is not really an independent operation, and is defined only through P and T. This is because only space and time are active elements, and their variations provide the coded information that uniquely determines the phase factor and so the entire nature of the fermion state. As we have seen, using the commutator method for differentiation, mass (a term which connects with the charge conjugation operator is a purely passive element, which can be excluded from the operator without loss of information.

We can consider the three terms in the nilpotent 4-spinor which follow the lead term or 'real' state of the fermion as the P-, T- and C-transformed versions of this state, or the ones into which it could transform without changing the magnitude of its energy or momentum. They are vacuum 'reflections' of the real state, originating in mathematically defined vacuum operations. While a fermion cannot form a combination state with itself, it can be imagined as forming a combination state with each of these vacuum 'reflections'. Where the 'reflection' exists as a 'real' particle state, then the combined state will become one of the three classes of bosons or boson-like objects. Summed up over the four terms in the component spinors, the wavefunctions of such combination states will be scalars, as required:

Spin 1 boson:

$(ikE + \boldsymbol{i}\mathbf{p} + jm)\,(-ikE + \boldsymbol{i}\mathbf{p} + jm)$ $\qquad T$

Spin 0 boson:

$(ikE + \boldsymbol{i}\mathbf{p} + jm)\,(-ikE - \boldsymbol{i}\mathbf{p} + jm)$ $\qquad C$

Fermion-fermion – Bose-Einstein condensate / Berry phase, etc.:

$(ikE + \boldsymbol{i}\mathbf{p} + jm)\,(ikE - \boldsymbol{i}\mathbf{p} + jm)$ $\qquad P$

The fermion-fermion combination can be used to demonstrate how one of these bosonic terms becomes a scalar when the product is summed over the 4 component terms:

$$(ikE + \mathbf{i}\mathbf{p} + jm)\,(ikE - \mathbf{i}\mathbf{p} + jm) = (ikE + \mathbf{i}\mathbf{p} + jm)\,(-2\mathbf{i}\mathbf{p})$$
$$(ikE - \mathbf{i}\mathbf{p} + jm)\,(ikE + \mathbf{i}\mathbf{p} + jm) = (ikE - \mathbf{i}\mathbf{p} + jm)\,(2\mathbf{i}\mathbf{p})$$
$$(-ikE + \mathbf{i}\mathbf{p} + jm)\,(-ikE - \mathbf{i}\mathbf{p} + jm) = (-ikE + \mathbf{i}\mathbf{p} + jm)\,(-2\mathbf{i}\mathbf{p})$$
$$(-ikE - \mathbf{i}\mathbf{p} + jm)\,(-ikE + \mathbf{i}\mathbf{p} + jm) = (-ikE - \mathbf{i}\mathbf{p} + jm)\,(2\mathbf{i}\mathbf{p})$$

Here, the terms on the right-hand side in ikE and jm total to 0, which leaves only $4 \times (-\mathbf{i}\mathbf{p})(2\mathbf{i}\mathbf{p}) = 8p^2$, which is a scalar (normalizable to 1).

Spin 1 bosons (for example, photons and gluons) can be massless because

$$(ikE + \mathbf{i}\mathbf{p})\,(-ikE + \mathbf{i}\mathbf{p}) \neq 0.$$

However, a spin 0 boson must have a mass because, otherwise,

$$(ikE + \mathbf{i}\mathbf{p})\,(-ikE - \mathbf{i}\mathbf{p}) = 0, \text{ etc.}$$

This is why the massless Goldstone boson must be transformed into the massive Higgs. A related fact is that fermion and antifermion cannot have the same handedness (determined by the sign of the E / $\mathbf{p}$ ratio). The fermion-fermion state, however, can be effectively massless, as in Cooper pairs, because, this time,

$$(ikE + \mathbf{i}\mathbf{p})\,(ikE - \mathbf{i}\mathbf{p}) \neq 0.$$

The weak interaction

While the Coulomb-type interaction is a necessary consequence of defining a point source with spherically symmetry, the other two main interactions, weak and strong, can be considered as the respective local consequences of the nonlocal superposition and combination states which result from defining a nilpotent fermion wavefunction in the form $(\pm ikE \pm \mathbf{i}\mathbf{p} + jm)$.

The weak interaction is essentially a harmonic oscillator, in which fermions and antifermions are mutually annihilated while bosons are created, or bosons are annihilated while fermions and antifermions are mutually created. The bosons appear at fermion-antifermion vertices, and are products of actions with at least a weak amplitude. *Zitterbewegung* is a result of defining fermions by a spinor wavefunction, which is a superposition of four fermion / antifermion states. These states must exist simultaneously or be regarded as the result of a process of switching between them. Their combinations produce all the main boson types, as do the allowed switchings.

In principle, the fermion, by its *very existence as a 4-spinor*, is always acting weakly, even if only with vacuum, and the fermion is necessarily a weak *dipole* (fermion / antifermion) in relation to its vacuum states. The single-handedness of the weak interaction can be regarded as the result of a weak dipole moment connected with fermionic ½-integral spin. The sources of weak interactions are always at least dipolar, and, in addition to the Coulomb (inverse linear) potential required for spherically symmetry, there must be a dipole or multipole potential (say, $\propto 1 / r^{-n}$, where $n \geq 2$). A combined potential of this form, when applied to a nilpotent operator, produces the energy levels of a harmonic oscillator, exactly as required (Rowlands, 2007).

The interaction has an $SU(2)$ symmetry, which, like the *zitterbewegung*, is, ultimately, related to the ± sign ambiguity attributable to the pseudoscalar term iE, and can be realised only through a switch in the helicity state. For nilpotent-nilpotent structures taken as defining the vertices for boson production via the weak interaction, it appears that the pure weak interaction requires left-handed fermions and right-handed antifermions, which requires both a C violation and a simultaneous P or T violation.

The strong interaction

A final interaction type emerges from another aspect of the nilpotent's internal structure. This is the vector nature of the **p** term, defined in a combination state through its 3-components. It is not obvious that we can write down a 3-component state vector, since

$$(ikE + \mathbf{i}\mathbf{p} + jm)\,(ikE + \mathbf{i}\mathbf{p} + jm)\,(ikE + \mathbf{i}\mathbf{p} + jm) = 0$$

However, we can write down terms of the form:

$$(ikE + \mathbf{i}\mathbf{p} + jm)\,(ikE + jm)\,(ikE + jm) \rightarrow (ikE + \mathbf{i}\mathbf{p} + jm)$$
$$(ikE + jm)\,(ikE + \mathbf{i}\mathbf{p} + jm)\,(ikE + jm) \rightarrow (ikE - \mathbf{i}\mathbf{p} + jm)$$
$$(ikE + jm)\,(ikE + jm)\,(ikE + \mathbf{i}\mathbf{p} + jm) \rightarrow (ikE + \mathbf{i}\mathbf{p} + jm)$$

This means that we can have a nonzero 3-component combination state vector if we use the vector properties of **p** and the arbitrary + or – nature of its sign. A state vector of the form, separating the **p** components:

$$(ikE \pm i\mathbf{i}p_x + jm)\ (ikE \pm i\mathbf{j}p_y + jm)\ (ikE \pm i\mathbf{k}p_z + jm)$$

will have six independent allowed phases, i.e. when

$$\mathbf{p} = \pm\ \mathbf{i}p_x\ ,\ \mathbf{p} = \pm\ \mathbf{j}p_y\ ,\ \mathbf{p} = \pm\ \mathbf{k}p_z\ ,$$

and these will be *gauge invariant*, or indistinguishable, and all present at once. The E, $\mathbf{p}$, m symbols here belong to a totally entangled state, rather than the subcomponents.

The 6 possible phases arc related by a gauge invariant $SU(3)$ symmetry, with 8 generators, exactly like that attributed to coloured 'quarks':

$(ikE + i\ \mathbf{i}p_x + j\ m)\ (ikE + j\ m)\ (ikE + j\ m)$	$+RGB$
$(ikE - i\ \mathbf{i}p_x + j\ m)\ (ikE + j\ m)\ (ikE + j\ m)$	$-RBG$
$(ikE + j\ m)\ (ikE + i\ \mathbf{j}p_y + j\ m)\ (ikE + j\ m)$	$+BRG$
$(ikE + j\ m)\ (ikE - i\ \mathbf{j}p_y + j\ m)\ (ikE + j\ m)$	$-GRB$
$(ikE + j\ m)\ (ikE + j\ m)\ (ikE + i\ \mathbf{k}p_z + j\ m)$	$+GBR$
$(ikE + j\ m)\ (ikE + j\ m)\ (ikE - i\ \mathbf{k}p_z + j\ m)$	$-BGR$

The simultaneous existence of $\pm\ \mathbf{p}$ or left- and right-handed helicity states here is a sure indicator that the state described has a nonzero mass.

Because the phases are gauge invariant, however, the mediators of the transitions will be six massless colour-anticolour bosons of the form:

$$(ikE - i\mathbf{i}p_x)\ (-\ ikE - i\mathbf{j}p_y)$$

and two combinations of three 'colourless' bosons of the form:

$$(ikE - i\mathbf{i}p_x)\ (-\ ikE - i\mathbf{i}p_x)$$

These *gluon* structures will be identical to an equivalent set in which both brackets undergo a complete sign reversal.

As a superposition of combination states, these structures will generate an entirely nonlocal local interaction. That is, the $SU(3)$ exchange of momentum $\mathbf{p}$ involved must be entirely independent of any spatial position of the 3 components of the baryon. This, in turn, generates a local consequence in the form now known as the strong interaction. Since the rate of change of momentum, or 'force', is constant with respect to spatial positioning or separation, this becomes equivalent to a potential which is linear with distance, exactly as we require for the conventional strong interaction.

Applying this strong linear potential to the energy term of the nilpotent, together with the Coulomb potential required for a point source, now gives the local consequence of the initial nonlocal combination. Potentials of this form applied to a nilpotent operator gives an analytic solutions which has the well-known strong interaction characteristics of infrared slavery and asymptotic freedom (Rowlands, 2007). It would seem, then, that all the well-known interaction types involved with fermions can be seen to be characteristic consequences of aspects of their nilpotent structure alone. Significantly, the linear potential of the strong interaction is the only one that is optional for a fermion, the nilpotency not being dependent directly on the vector nature of **p**.

In fact, there are only three analytic interaction solutions for a fermionic nilpotent operator: the Coulomb inverse linear potential, which provides the 'hydrogen atom solution' associated with the electric interaction; Coulomb plus linear potential, which produces strong confinement; and Coulomb plus any other potential or sum of potentials, which provides the harmonic oscillator solution characteristic of the weak interaction.

These interactions can be seen as products of the three types of quantity (pseudoscalar, multivariate vector and scalar) contained within the fermionic nilpotent structure. In principle, they arise from the need for a discrete (point) source to preserve spherical symmetry and hence to conserve angular momentum. The three interactions and their associated symmetries can be identified as being connected with the three separately conserved aspects of angular momentum, and the three things from which spherical symmetry is intrinsically independent:

magnitude	scalar	$U(1)$	length of radius vector
direction	vector	$SU(3)$	choice of axes
handedness	pseudoscalar	$SU(2)$	left- or right-handed rotation

Partitioning the vacuum

We have seen that the entire structure of quantum mechanics (and of quantum field theory) follows as soon as we define the creation operator for a single fermion ($\pm\, ikE \pm \mathbf{i}\mathbf{p} + jm$) as a nilpotent. Really, it is a simultaneous definition of four creation (or annihilation) operators (or of two of each):

$(ikE + \mathbf{i}\mathbf{p} + jm)$	fermion spin up
$(ikE - \mathbf{i}\mathbf{p} + jm)$	fermion spin down
$(-ikE + \mathbf{i}\mathbf{p} + jm)$	antifermion spin down
$(-ikE - \mathbf{i}\mathbf{p} + jm)$	antifermion spin up

Only one of these terms will define the real fermionic state, the rest being vacuum 'reflections', or the states into which it could transform. If we regard the mass terms as passive (and capable of being eliminated), the sum total of all four terms is zero, precisely as we would expect from a zero totality universe defined by a single fermion. The vacuum reflections can be understood more directly is we take $(ikE + \mathbf{i}\mathbf{p} + jm)$ and postmultiply it by any of the *idempotent* terms

$$k(ikE + \mathbf{i}\mathbf{p} + jm)$$

or

$$i(ikE + \mathbf{i}\mathbf{p} + jm)$$

or

$$j(ikE + \mathbf{i}\mathbf{p} + jm)$$

The result in each case is $(ikE + \mathbf{i}\mathbf{p} + jm)$, multiplied by a scalar, which may be removed by normalization, and the operation can be done an indefinite number of times. (In the case of $j(ikE + \mathbf{i}\mathbf{p} + jm)$, there will also be a vector term, which is removed on alternate multiplications.) In effect, the three idempotent terms behave as a vacuum operators.

The three vacuum coefficients k, i, j can be seen as being responsible for the concept of discrete or point-like charge, performing as the sources of weak, strong and electric charges. They act to partition the *continuous* (gravitational) vacuum, represented by $-(ikE + \mathbf{i}\mathbf{p} + jm)$ and responsible for zero-point energy, into discrete components, whose special characteristics are determined by the respective pseudoscalar, vector and scalar natures of the terms iE, $\mathbf{p}$ and m, with which they are associated. As vacuum coefficients, though delocalized, they are related to 'real' weak, strong and electric localized charges, and we have already seen how this nonlocal / local connection works, particularly in the case of weak and strong interactions.

The three partitions can be described as strong, weak and electric 'vacua', and they can be seen to have very particular roles within existing physics:

$\boldsymbol{k}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$	weak vacuum	fermion creation
$\boldsymbol{i}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$	strong vacuum	gluon plasma
$\boldsymbol{j}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$	electric vacuum	isospin / hypercharge

(The electric vacuum creates the $SU(2)$ symmetry of weak isospin in its alternate states of being empty and filled (Rowlands, 2007).

The idea also connects strongly with the gravity-gauge theory correspondence now appearing in string theory, and also apparent in previous work on particle structures. The 3 bosonic states can also be related to the vacua produced by the 3 quaternionic operators:

weak *spin 1*

$$(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{k}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{k}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{k}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ \ldots$$
$$(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(-i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(-i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ \ldots$$

electric *spin 0*

$$(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{j}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{j}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{j}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ \ldots$$
$$(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(-i\boldsymbol{k}E - \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(-i\boldsymbol{k}E - \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ \ldots$$

strong *paired fermion state*

$$(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{i}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{i}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{i}\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ \ldots$$
$$(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(i\boldsymbol{k}E - \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,(i\boldsymbol{k}E - \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\ \ldots$$

Since all these structures are equivalent to the original fermion state $(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$, at the same time as generating equivalent boson states in vacuum, we see that the nilpotent formalism is the only one in which an intrinsic and exact supersymmetry between fermions and bosons emerges purely from the mathematical formalism. For each real fermion there is a vacuum boson and vice versa. We can use this directly to eliminate the divergences due to self-energy in quantum field theory by adding the fermion and boson loops in equivalent numbers.

Supersymmetry and renormalization

The vacuum energy for a particle of mass m and spin j is given by

$$\frac{1}{2}(-1)^{2j}(2j+1)\int d^3k\sqrt{k^2+m_j^2} = \frac{1}{2}(-1)^{2j}(2j+1)\int d^3k\sqrt{k^2}\left(1+\frac{1}{2}\frac{m_j^2}{k^2}-\left(\frac{m_j^2}{k^2}\right)^2+\ldots\right)$$

To remove the quartic, quadratic and logarithmic divergences, we need to ensure that

$$\sum_j (-1)^{2j}(2j+1)=0$$
$$\sum_j (-1)^{2j}(2j+1)m_j^2=0$$
$$\sum_j (-1)^{2j}(2j+1)m_j^4=0$$

The first of these conditions requires equal numbers of fermionic and bosonic degrees of freedom. If we have $j = \pm \frac{1}{2}$ for the fermionic loops and $j = \pm 1$ for the bosonic loops, then

$(-)^{2j}\,(2j+1)^{2j} = -2$	for	$j = \frac{1}{2}$
$(-)^{2j}\,(2j+1)^{2j} = 3$	for	$j = 1$
$(-)^{2j}\,(2j+1)^{2j} = 0$	for	$j = -\frac{1}{2}$
$(-)^{2j}\,(2j+1)^{2j} = -1$	for	$j = -1$

giving a total of

$$\sum_j (-1)^{2j}(2j+1) = -2+3+0-1=0$$

as required.

The result, which can also be found using perturbation calculations, can be extended to the Higgs boson. The nilpotent formalism also removes the infrared divergence for propagators. he reason is relatively simple. The equal mathematical representation of $+E$ and $-E$ at the same level means that the negative fermion energy always cancels the positive boson energy. At the same time, it also solves the so-called matter-antimatter asymmetry in the universe. Matter and antimatter are actually completely symmetric, but they inhabit dual spaces.

Duality and the nilpotent structure

The nilpotent structure incorporates duality at many levels:

operator and wavefunction	fermion and vacuum
fermion and vacuum boson	operator and amplitude,
nilpotent and idempotent	broken and unbroken symmetries

All originate in the idea that, by defining a fermion state, we are also defining a fundamental singularity. To define a singularity we are forced to use a dualistic structure by simultaneously defining what is not singular. It is the reason why the fermion has half-integral spin – we can only define it by simultaneously splitting the universe into two halves which are mirror images of each other.

The duality manifests itself physically in the phenomenon of *zitterbewegung*. Using either operator or amplitude, we define ($\pm ikE \pm i\mathbf{p} + jm$) as a 4-spinor, with 4 terms (each of which is nilpotent) arranged as a column / row vector. The 'real' state (the one subject to physical observation) is determined by the signs of E and $\mathbf{p}$ in the first term. The other three states are like three 'dimensions' of vacuum, the states into which the real term could transform by respective P, T or C transformations. The duality ensures that fermion and vacuum occupy separate 3-dimensional 'spaces', which are combined in the γ algebra defining the singularity state.

These 'spaces', though seemingly different, are truly dual, each containing the same information, and the duality manifests itself directly in many physical forms.

Pauli exclusion by antisymmetric wavefunctions	uses	**i, j, k**
Pauli exclusion by nilpotency	uses	**I, J, K**

Both sets of coordinates yield information about the same physical quantity: angular momentum.

spin ½ from anticommuting aspects of **p** components	**i, j, k**
spin ½ using Thomas precession (relativity)	**I, J, K**
velocity addition law from 2 D of space	**i, j, k**
same using relativistic space-time	**I, J, K**

holographic principle – bounding 'area' defined by

two spatial coordinates	**i, j, k**
or one of space and one of time	**I, J, K**

Comments on Penrose's twistor representation and on strings

These dualities show that the relativistic connection between space and time exists in a different vector space to the connection between the three spatial components. It is not strictly a 4-dimensional connection at all, though it appears as such for the scalar product of a massless object. The intrinsic 4-D connection is an assumption of general relativity, which is Penrose's starting point (and also a problem, as it is incompatible with quantum mechanics). This is why Penrose's twistors assume a massless world, with the intrinsic motion of the particles at the speed of light.

The twistors derive their dual 4-D vector space from the intrinsic duality of a 3-D vector space, in requiring vectors and pseudovectors. However, quantum mechanics really requires an additional duality (a dual dual space, which does not require an arbitrary extension to 4-D). Mass emerges from this extra duality even if we assume that the intrinsic motion of the particles is at the speed of light. Defining a physical singularity in terms of two vector spaces produces mass, as well as spin and chirality.

The nilpotent operator ($\pm ikE \pm \boldsymbol{i}\mathbf{p} + jm$) can be regarded as a 10-D object (embedded in Hilbert space): 5 for iE, $\mathbf{p}$, m and 5 for $\boldsymbol{k}, \boldsymbol{i}, \boldsymbol{j}$; and six of these (all but iE and $\mathbf{p}$) are compactified. It reduces to 8 or 2×4 if the intrinsic redundancy of m and the scalar 1 are considered. I have it on the authority of a string theorist colleague that: 'Self-duality in phase space determines vacuum selection.' There can be no doubt that ($\pm ikE \pm \boldsymbol{i}\mathbf{p} + jm$) fulfils all these criteria (whatever they mean in specific terms). It is also a mass-shell system and incorporates the right groups.

REFERENCES

Dirac, P. A. M., 1958, *The Principles of Quantum Mechanics*, fourth edition. Clarendon Press.

Feynman, R. P., 1972, The development of the space-time view of quantum electrodynamics, in *Nobel Lectures: Physics 1963-70*. Elsevier Publishing Company, Amsterdam.

Kauffman, L. H., 2004, Non-commutative worlds, *New Journal of Physics.*, **6**, 2-46.

Penrose, R., 2004, *The Road to Reality*, Jonathan Cape.

Rowlands, P., 2007, *Zero to Infinity*, World Scientific, Singapore.

ANPA PROCEEDINGS EDITORIAL POLICY

ANPA has been criticised in the past - in particular by members of its own Advisory Board - for having no formal editorial policy for its Proceedings. This has been balanced by a feeling within ANPA that we should keep ourselves open to all viewpoints. In the last few years as editor I [K.B.] have tried to tighten things up in such a way as I felt would satisfy our critics whilst not compromising our own position. This has been partially successful although for some time I have felt that it is time that there was a formally stated policy. The following has been approved by the Executive Council, although it is open to feedback from all. By "the editor" is meant the Editor or (an) appropriate nominated Referee(s) (note the capital R!)

1. The paper should make a new and original contribution to the fields of ANPA's interest. Survey papers are acceptable.

2. The default use of language for submitted papers in Physics and Philosophy of Physics should be the common language of Physics as usually understood by Physicists, and, in particular, by Philosophers of Physics. Any other use of language should be carefully explained at the start of the paper and all appropriate definitions included there.

3. The editor should be satisfied that the paper is ***presented*** in such a way that the majority of the readership will understand the author's intentions. In particular ***it should be clear*** that the author has a correct understanding of the subject matter. Please avoid forward references.

4. "Verbatim" reports will be accepted subject to the above three conditions only, regardless of whether the final draft is an accurate rendition of what was originally said.

5. Theories of any nature are acceptable material, provided they are compatible with the known facts, and provided they are deemed to be of interest to the readership. Theories of alternative, imaginary worlds are also acceptable, provided their nature is made clear.

GUIDELINES FOR PREPARATION OF MANUSCRIPTS FOR THE ANPA PROCEEDINGS

- FORMAT: Only electronic submission as a PDF file or a WORD file (as a second option) will be accepted
- FONTS SIZE: The body text should be in at least 14 point type
- EMBEDDED FONTS: All fonts must be embedded
- LINE SPACE: The manuscript should be double-spaced (or at least 1.5 spaced) throughout
- JUSTIFIED TEXT: The text must be justified on both sides
- MARGINS: The left, right and the top margins should be the same: 30mm. The bottom margin should be 40mm
- MAX NUMBER OF PAGES: The length of the manuscript must not exceed 30 pages (please, contact the editor if your manuscript is longer than that)
- PAGE SIZE: The basic page size should be the standard European A4 format page (length: 297mm; width: 210mm)
- PAGE NUMBERS: The page number should NOT be typed
- DEADLINE: Contributions must be submitted latest 1 January following the ANPA meeting

ALTERNATIVE NATURAL PHILOSOPHY ASSOCIATION STATEMENT OF PURPOSE

1. The primary purpose of the Association is to consider coherent models based on a minimal number of assumptions, so as to bring together major areas of thought and experience within a Natural Philosophy alternative to the prevailing scientific attitude. The Combinatorial Hierarchy, as such a model, will form an initial focus of our discussions.

2. This purpose will be pursued by research, publications and any other appropriate means including the foundation of subsidiary organisations and the support of individuals and groups with the same objective.

3. The Association will remain open to new ideas and modes of action, however suggested, which might serve the primary purpose.

4. The Association will seek ways to use its knowledge and facilities for the benefit of humanity and will try to prevent such knowledge and facilities being used to the detriment of humanity.

ORGANISATION

1. The Founder of the Association was Pierre Noyes. The Founder Members were Pierre, John Amson, Ted Bastin, Clive Kilmister and Frederick Parker-Rhodes. They will be known herein as the Founders. The Executive Council is the governing body of the Association. It consists of: (a) The Founders and all past Presidents of the Association, the President, the Co-ordinator and the Treasurer, (b) Ordinary members nominated by classes (a) and (b), who serve for three years, with the possibility of re-nomination.

2. The Members of the Association are (a) the members of the Executive Council and (b) others nominated by the Members and approved by the Executive Council.

3. The President has the responsibility for calling meetings of the Executive Council, at least annually, for the determination of overall policy.

4. The Treasurer is the responsible financial officer of the Association for the receipt and disbursement of funds and shall maintain and make available appropriate records, including annual accounts.

5. The President and the Co-ordinator services will include the organisation of meetings and the editing of the Proceedings of such meetings for publication, co-ordination of, and participation in, the research activities of the Association, preparation when appropriate of research reports and publication of such reports, and other such duties as may be assigned.

6. Members of the Executive Council may as appropriate receive funds for travel, expenses, etc.

7. The Executive Council has selected an independent Advisory Board. It may adopt its own rules for the operation and replacement of members. The Executive Council may nominate candidates to the Board. Any member of the Board, or the Board collectively, may make recommendations to the Executive Council, or directly to the Membership. Action taken on such recommendations must be promptly reported by the Executive Council to the Board in writing.

8. No one of the general public who has booked and paid the appropriate fees should be excluded from attending ANPA meetings except as decided by a majority of the Executive Council in a formal meeting. Notification of ANPA meetings will be sent to all members.

Executive Council *(active members only):* Dr. Keith Bowden, Arleta D. Ford, Mike Horner, Prof. Louis Kauffman
President: Dr. Keith Bowden, TPRU, Birkbeck College, University of London
Co-ordinator: Arleta D. Ford, TPRU, Birkbeck College, University of London
Treasurer: David Roscoe, Department of Applied Mathematics, Sheffield University
Advisory Board: Mike Horner (Chairman), Profs. G.F. Chew (Berkeley), C. Isham (Imperial College), M. Redhead (Cambridge and LSE), N. Cartwright (LSE), H. Pierre Noyes (Stanford), Rev. Dr. Philip Luscombe (Wesley House, Cambridge).

ISBN 978-0-9562148-1-2