Participations

Proceedings of ANPA 21

Keith G. Bowden, *Editor*

Proceedings of the 21st Annual International Meeting of the

Alternative Natural Philosophy Association

Wesley House, Jesus Lane, Cambridge,
September 1999

Published by **ANPA**
c/o Dr. K. Bowden,
Theoretical Physics Research Unit,
Birkbeck College, Malet St, London WC1E 7HX.

July 2000

Participations: Proceedings of ANPA 20/Keith G. Bowden, *Editor*

ISBN 0 9526215 5 X

published by ANPA c/o Dr. Keith G. Bowden,
139 Sandringham Rd, Barking,
Essex IG11 9AH, UK

Contents

My Beautiful Launderette

As there is no television, today we watched the clothes going round and around in the washing machine. Unlike with television, one tends to follow the clothes around with one's eyes leading to some discomfort in the neck and shoulders the next day. Eventually came the inevitable sound of tearing rubber and a flood of water onto the kitchen floor. I wondered if anyone ever reads these editorials. (I wish they would read the Notes for Authors.) We spent some time clearing up the mess. It is normal.

Arleta has decided not to publish her algebra in this volume. She tells me that she is expanding it for next year and wants to concentrate on displacement activities for the time being. This is something I can empathise with. Unexpectedly, however, we do have some fascinating trigonometrical notes from John Amson to compensate, and additionally Arleta has promised to entertain everyone for an afternoon at our meeting. Mike cannot come (despite bribes), but sends his greetings. (I have also included my IJGS paper, which was accidentally omitted last year.) Last night I dreamed that the washing machine was chasing after me. It had changed somewhat, now being constructed from transparent blue plastic. It called itself "Mac" and affected unusual airs when challenged.

Today the washing machine repairman came. He was drunk as usual. He said that he would do what he could but that he did not believe that he could stop its nighttime activities and indeed that even a three month guarantee did not cover such eventualities. Various people were debating the correct usage of quaternions and ideals. We asked him for what we could trade the machine in. The debate got louder and it became difficult even to hear the repairman. He was apparently saying that he no longer did washing machines and was moving into the cable television business. "That is where the future lies," he continued, somewhat tangentially, leaving us at a loss.

Remember, What the dormouse said.

KEITH BOWDEN, TPRU, JULY 2000

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 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.
{* added by KGB}

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.

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. Other such reports are better submitted to the Newsletter.

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.

ANPA Proceedings Notes for Authors

I would like to try to continue conformity of *style* for future issues of the Proceedings. Ideally I would like contributions to be submitted in International Journal of General Systems format (I have some copies of their Notes for Authors) or similar - **LOOK AT MY PAPER IN THIS ISSUE OF THE PROCEEDINGS FOR AN EXAMPLE.**

At least, Times Roman, 12 point, ***single sided, two copies (HARD COPY),*** is preferred. **10 point is TOO SMALL to be reduced to A5; 14 point is better for short papers.** Main heading 20 point capitalised and centred, other headings 16 point capitalised to the left. Author's name(s) capitalised and centred. Address italicised and centred. No underlining. At least a one inch bottom margin for footers; page numbers NOT bottom right. *Only copy in good English will be considered, and remember, this is a formal Proceedings.* **Remember also to include your name (surprising how many people omit this!), affiliation and full address, email address and the version number (even if it is 1.0) or date of the draft, centred below the main heading.** I often get sent more than one version of a paper and invariably mix them up! Send copy **to *KEITH BOWDEN, 139 SANDRINGHAM RD, BARKING, ESSEX IG11 9AH.***

The copy date for the ANPA2000 Proceedings is January 1st 2001. The issue will go to print on April 1st 2001. This will be adhered to rigidly this year.

Keith Bowden,
Theoretical Physics Research Unit,
Birkbeck College,
Malet St,
London WC1E 7HX, UK.
k.bowden@physics.bbk.ac.uk
0208-594-5064
FOR EDITORIAL ADDRESS SEE THE PENULTIMATE PARAGRAPH

Many thanks to Patrick and David in the Print Unit at the School of Oriental and African Studies for printing and binding this volume and for making such an excellent job of ANPA20.

THE PARTICIPANT OBSERVER

TED BASTIN
Maesllwyn, Tan-y-Groes,
Cardiganshire SA43 2JF

and C.W.KILMISTER
Red Tiles Cottage, High St,
Barcombe, Lewes,
E Sussex BN8 SDH

1. THE PARTICIPANT OBSERVER IS INEVITABLE IN A PROCESS THEORY

In this paper we give predominance to the philosophical argument about the participant observer. Although we do not avoid the mathematical consequences, and think that it is vital to handle them somehow, our treatment must be speculative in part at this stage, and much more work is needed.

The phrase "participant observer" was coined by someone at the time of Program Universe. It presented a real insight into what was entailed but was not followed up - being sidestepped by a Bridgman-type operationalism. We have been sitting on the fence ever since - protecting ourselves by saying that, of all the things that might occur in our process account of the hierarchy, effects associated with observation would appear automatically, given long enough. In the past couple of years it has become increasingly clear that further elaboration and analysis of the concept of discrimination can never avoid the epistemological questions which are usually lumped under the concepts of the observer and of observation. Indeed recent studies by Clive amount to mathematical proof of this.

Through a long investigation of the hierarchy foundations we find it necessary to postulate a 'pre-geometry' (Zimmerman's expression) which has the following characteristics:

1. It is sequential and refers to experience which is sequential (and therefore recursive if it is to have any power of construction or 'memory' (reference back).

2. It is at once mental and physical - this distinction not having yet appeared.

3. The basic operation of discrimination already used in the hierarchy is postulated without a distinction between knowing and constructing.

4. These three characteristics can be summarised in the notion of process, in which the immediacy of the current step in each sequence is primary and logically prior to any objective or universal background 'space'.

To these principles I add some which explicitly introduce the participant observer (PO).

1. The PO is a sequence of which only the last term is immediately present or, as we may say consciously presented.

2. The entire sequence acts in the build-up of a world picture, and there is no previously existing 'objective' world picture.

3. If a sequence has a limit point (for example, a Parker-Rhodes bound) then it defines an *observation*. Since enumeration of the steps of the sequence furnishes a countable number, we may equivalently call an observation a *measurement* and the number is the *result* of the measurement.

How is a sparse, discrete set of groups as we have in CH able to give an adequate description of physics? We have always claimed that structure could be treated "from the filling". However it is necessary to fill in between and we now propose the sequences of the PO as the right way to do this and emphatically reject the mathematician's automatic response of generalising the groups to algebras over some field (say the rationals).

2. LIMIT POINTS

The essential PO feature here is the determination that the sequence has been *completed* and so has given rise to a limit point (which may or may not be a term of the sequence) which is then the entity in the CH. It is important to realise that we deal in whole sequences with limit points. Such a sequence may be recapitulated to give the detail, which we experience as the particularity of an observation. It is only with very special sequences that we advance directly to an atomic event, which incorporates none of the special detail. Clearly we have here something

which might be called memory, though this term would lead us into a distinction between a 'present' element and those that are past, and this we wish to avoid. If the background is ideal and leads straight to the limit then he/it gets no information about spatial configuration and the world is completely isotropic and featureless. We seem still to have to appeal to the ergodic principle to say that a limit point will be reached if we wait long enough, and we seem also to need the varying lengths that we have to wait to incorporate empirical measurement. These will be fitted into the isotropic space picture.

Mathematically, the limit points arise through the exhaustion of algebraic possibilities. Physically, they appear as a discrete or atomic event of some sort, which gives a definite piece of information. This definition is not as technically restrictive as it may appear, since all macroscopic descriptions depend ultimately on quantised effects - photons if not particles. One is reminded of Bohr's belief that macroscopic description must always be involved in measurement. In the famous paper of Bohr and Rosenfeld use is made of springs and weights to make a formal argument, but the mechanical details are left to the imagination of the physicist and do not matter much.

Because the PO delineates entities in the CH clearly as limit points, the notion of *signal* as something in the CH, which is not, an entity, becomes clearer. Such signals are needed to show when two different sequences have reached the same limit point. Those of a Frege-like turn of mind would see the sense/reference distinction here - the two sequences are different senses with a common reference, the physical object.

3. LEVELS

The link between the combinatorial hierarchy and experience of the physical world has been maintained through the concept of the *level*, and it is this concept whose meaning has to be re-examined to clarify the physical interpretation. In the thinking which went on and culminated in the Parker-Rhodes algebra the levels were stages of construction and we jumped into a direct physical identification of them without going into the questions about whether they were there in the world or whether they were constructions of our own making. With the deepening in foundations incident upon our taking a process view of the hierarchy it becomes the more urgent to remedy this vagueness. In the modern *process* form of the hierarchy construction, entities become accessible, but nothing is known about these, so at the initial stage the first two are taken as at the first level and must be discriminated giving a third. This is a dcs (discriminately closed subset) and this is the first level, and there is no trouble until later.

It may be that this first dcs is represented by a function, and so by an element at the next level: or else it may be that instead a third entity becomes accessible.... and so on.... until a situation arises in which there are only entities which have become accessible. There are also elements which represent dcs's of the first, and there may be elements which represent dcs's of the second kind of element and so on. When a new element comes along does it have first to be tried against various others to see if it can be compared or not before plugging into the discrimination machine? Or can one generalise the discrimination machine in some way so as to absorb this first decision into the general one? It seems inevitable that this complex situation can be reduced to the simple one of the combinatorial hierarchy (with or without aspect) only by invoking something that mirrors the role of measurement in picking out from the welter of events just those that are needed. In other word the constructive or process account does not tell us why we stop at the construction of a level, as we need to do to get a physical interpretation - in terms of coupling constants for example.

We had always in effect assumed that no reasonable person would go beyond the direct constructions and so the levels would be definite in extent. This was originally just an act of faith. Now it seems that something entirely different is needed. Let us recall the Concept of Order work. The basis of that was the distinction between "possibility of ordering" and "absence of possibility of ordering". This distinction constituted a principle which enabled us to introduce spatial intuition where otherwise there would be only sequentiality and composition in Etter's sense. Even the discriminately closed subset is subsidiary to this principle - being a way of handling it in a conventional mathematics - and arises out of the need for a formal tool which gives the same numbers as the order principle. Thus the intuition which started things off was that the essentially *spatial* element in experience was the provision of an alternative to sequentiality. This alternative route has to be the impossibility of defining an order. We drew attention to the unique structure of the quadratic group in which all the formal relationships of the form $c = a \times b$ were true under any re-ordering of a, b, c. The importance of the quadratic group was that it was the minimal presentation of the particular kind of symmetry which we called 'similarity of position' and linked with simultaneity (the impossibility of defining time). The group was still not completely minimal since consistent thinking demanded a unit element by relating one element with itself, and one could proceed further with more complicated structures - notably the quaternion group. The mathematician was liable to be misled however and to mistake mathematical complexity for the essentials. We tried to explain that we

were defining dimensionality itself in a combinatorial way and with no reference at that stage to algebra in the first place.

Begin with Frederick's bit-strings as vectors. One particular bit-string, (0, 0, 0....) is marked out as special and so excluded from dcs's and so on. Since sequential process is the only way we have found to understand the CH we (have to) enshrine this special role of one element from the beginning, so we call it a *signal*, z. Other things in the system are *elements*. We label elements by a set of symbols, using discrimination*. It is done by a modification of Conway's construction. Note carefully just what comes:

1. Element. Call it a. $a + a = z$.
2. Eventually a different one b . $a + b = c$.
3. $c + a \neq z, c, a$ so could be b, $= a + c$ also.
4. $b + c \neq b, c, z$ so could be $a = c + b$ also.

We used to say that this gives the quadratic group,

	z	a	b	c
z	z	a	b	c
a	a	z	c	b
b	b	c	z	a
c	c	b	a	z

but we were jumping the gun. In fact *all* that our argument gives is

	z	a	b	c
z				
a		z	c	b
b		c	z	a
c		b	a	z

Things like $z + z$, $z + a$ etc. can never arise in the physical process. Accordingly we argued (probably correctly) that they can be assigned any convenient values and it is fairly obvious that the *most* convenient such assigning will be $z + u = u$ for all u. This produces an associative system i.e. the quadratic group. [If you want to put it the other way round; i.e. put the cart before the horse as mathematicians often like to do, you can arbitrarily assume associativity and so prove, e.g. $z + a = (b + b) + a = b + c = a$]. Assigning these seven values in the table is the primitive entry of

mental (or "theoretical") as compared to the other nine which arise physically.

4. THE CONSTRUCTION OF SEQUENCES AND THE PARTICIPANT OBSERVER

The next stage is in our argument is to identify the passage to a new level, or rather the being forced by experience to a new level, as the formation of a new term in the participant observer sequence. We see the PO as a sequence of steps at each of which a piece of empirical knowledge is acquired by advancing to the two-level condition and then with reversion to the primitive level. Thus at each stage in the PO sequence there is a primitive intuition which breaks the bounds of the level structure followed by an 'understanding' of the situation involving a two stage process with memory, followed again by a reversion to the primitive intuition. We are trying to encapsulate the jump backwards and forwards from spatial intuition to the ordered sequence or process, that is at the heart of experience. In Etter's words we are trying to bridge the gap between composition and extension by inventing a process in which we jump backwards and forwards between them, pushing the construction forward each step of the way. All this happens in the case of any awareness of the physical world but we only get clarity in the physicists' sense about it after a long period of abstraction which is roughly the history of physics.

5. 'PHYSICAL' AND 'THEORY'

Coming back now to the Participant Observer, we find him or her having this immediate perception or intuition of dimensional structure which is not represented in current theory. Clive likes my illustration of the batsman facing the fast bowler, which I should be embarrassed to trot out otherwise given my cricket-bore propensity. As a first approximation you move your bat in the vertical plane which includes your eye and the motion of the ball *before and after* it hits the ground. This is universally called 'playing a straight bat', though no cricket writer has so defined it nor shown understanding of the variational principle, which makes it, necessary. This principle is needed because the large uncertainty in the height of the ball after bounce (the distance from the batsman of the point of contact of the ball with the ground is very hard to estimate) has to be covered by the bat, which is long and thin. The point here is that the action is all at the limits of perception, and you do not see much except this plane. Then the activity of the bowler is concentrated on producing variation in the motion to

defeat the batsman and so necessarily has to introduce a new dimension - either by aerodynamic action or by reaction off the pitch. The batsman is doing well if he has time to observe this change to the first approximation picture. The direction (towards or away from him) is a second approximation. Anyway he begins to build up another dimension, and so in the combinatorial way he has 3D. One can build progressively an understanding of fine batsmanship, which involves further subsidiary variational principles, though needless to say the batsman only through haptic experience knows these.

I realise that my example will seem bizarre. However it is only an example and one could find quite different circumstances to illustrate the same point wherever one looked. More importantly one should try to put oneself outside the conventional representation of dimensionality, from which vantage point one will see that it is at least equally unconvincing to see up/down, forwards/backwards, left/right as underlying our analysis of each bit of world-structure. And that leaves out the most vital point from which we began long ago that our way of doing things explains the 3D whereas you will not find this arising anywhere in the conventional view. If we postulate an isotropic space its dimension number can be anything and one needs a reason for putting in 3.

We now have two interlocking forms of experience. We shall call the one that gives rise to dimensionality directly from experience *physical*, and the one that follows mathematical procedures *theory*. (I know this terminology is grammatically awkward, but let it stand for the moment.) One can lead up to this dichotomy by considering the mathematics of the hierarchy. Consider the first two stages in the construction of the hierarchy.

1. Two entities (vectors for Frederick) give rise by discrimination to a group G_2 and 3 dcs's.

2. Each dcs is represented by a single entity (matrix for F).

3. The naturally induced binary operation (matrix addition) is a discrimination, so the 3 second level entities (matrices) generate a group $G^*_2 \cong G_3$ and 7 dcs's.

4. Because this is just like stage 1 with G_3 instead of G_2 the process can repeat, until the Parker-Rhodes limit. Stage 4 is important in the way it is used to ensure repetition, though one might perhaps have done that differently. This is the stage, which raises questions when aspect is taken into account.

There is also a difference between G_2 and G_3 in the original CH, which has been glossed over. G_2 arises directly from the process, and G^*_2 comes from it also in a direct way. Thinking in terms of the PO and of observation one could say they come directly from it. We call such structures direct. However G_3 imports an additional theoretical element, as it were, because, to have two entities and their discrimination and then to introduce a third is a process which MAY introduce a G_3 in one step, MAY do it in two steps,..... but the expected number of steps to be sure to generate a G_3 is not finite.

The point is that if the potentially new element is discriminated against the three (a, b, ab) already in play, and if in fact it is not new, the chance of establishing this is obviously 1/3. In the other 2/3 outcomes, the discrimination produces another element so at the next discrimination, even in the most favourable case when no further new elements have entered, there are 5 new elements in play, and the chance of success after at most 2 steps is $2/3(1 + 2/5) = 7/15$, and so on. Proof of divergence is easy (Raabe's test). We call such a construction, for which there is only a probability $p < 1$ for it to arise at any particular point in the process, *indirect.*

How then is it possible for $G^*_2 \cong G_3$ when one side is direct and one is indirect? The answer seems to lie in the structure. Every sub-group of G_3 generated by 2 of the 3 generating elements is isomorphic to G_2 and so could be directly generated. The whole isomorphism $G^*_2 \cong G_3$ is a composite of isomorphisms between the different subgroups and that it is the putting together of these sub-isomorphisms which makes the whole thing which is then indirect.

However direct/indirect does not mean quite the same as physical/theory. It is when one asks the question why we should attach great importance to the direct structures when the mathematician would see no case for doing that, that we are forced to look deeper for reasons which are not, perhaps, in the normal sense mathematical that we contemplate the physical/theory distinction.

6. ASPECT

It was an argument that purported to explain spin in terms of the physical development which led to the discovery of what we call 'aspect' within the corpus of theory. The aspect generalisation led to the realisation that discrimination is not commutative. The process is one in which the incoming new element, and not the already established one, has to be re-labelled if they should turn out to be equal. Because $a + b \neq b + a$ would

make mathematicians choke we go over to (what was already in CO1 [The Concept of Order 1]) the notation ab for discriminating b against a, so now a b ≠ b a. None-the-less we have to recognise that a b, b a are not independent, so there arises another signal, call it y, so that (a b)(b a) = y. The basis structure now has 6 elements instead of 3, say a, b, c, d, e, f but proceeding with Conway it turns out that a f = f a = y, b e =e b =y c d = d c = y. So the six elements group into 3 pairs (a,f), (b,e), (c, d). The reason is obvious. Just begin

	a	b
a	z	→ c
b	d	z

and c → d

a b must be different from a, b, z so call it c. b a must be different from c as well as from a, b, z so call it d and then c d = y because c = a b and d = b a, etc. Introduce the notion of dual, * , a* = f, f* = a etc. Then inspection of the table shows that, if u v = w, then u v* = u* v = w* and u* v* = w. So we don't need such a huge table but can summarise it as:

	y	z	a	b	c
y					
z					
a			z	c	b*
b			c*	z	a
c			b	a*	z

where the blank squares can be assigned. Because we have really got columns for a* etc. there are 6 × 6 = 36 physically determined squares, 28 to be assigned by "theory".

Look at the little 2 × 2 square at the top. It can really only contain signals, no elements. To escape triviality the lines must be different and so must be the two columns. There seems little plausibility in other than one of:

	y	z
y	y	z
z	z	y

	y	z
y	z	y
z	y	z

and both of these are in fact tables for the only group of order 2, C_2.

The first has y as identity and z as the other element: the second has them interchanged. So these two assignings are the two *different* theoretical constructions, since y, z is not any old symbols but the signals for weak equivalence and equality respectively.

Thus, unlike the theoretical assigning in the old CH, there is a choice of two "reasonable enough" theories. Each involves the group C_2 and so has to distinguish the two cases. I use the notation $(1, \phi)$ for this group, 1 being the identity, and $\phi^2 = 1$. Then the two theories would be respectively:

$$y = 1, \quad z = \phi \qquad \text{and} \qquad z = 1, \quad y = \phi.$$

Moreover it is obviously very useful (though it may not be necessary) to define the dual operation by $u^* = \phi\, u$ in each case. Then one system is

	ϕ	a	b	c
ϕ=z	1	a*	b*	c*
a	a*	ϕ	c	b*
b	b*	c*	ϕ	a
c	c*	b	a*	ϕ

where $u^* = \phi\, u$

and the other is

	ϕ	a	b	c
ϕ	1	a*	b*	c*
a	a*	1	c	b*
b	b*	c*	1	a
c	c*	b	a*	1

where $u^* = \phi\, u$.

What seems important here, shown up by aspect, is that these two systems are different theoretical constructions of the same physical things i.e. of the part enclosed by the dotted lines. The second one fails to be associative because

$$a(ab) = ac = \phi b$$

but $$(aa)b = b.$$

The first one is associative. The situation at the next level with aspect is more complex with one associative and seven non-associative systems. But there, as here, the physical development of each is the same.

7. PHYSICAL AND THEORY - ELECTROMAGNETIC LANGUAGE AND SPIN

To get further along the road of direct spatial intuition (and I deliberately use the Kantian term rather than 'perception' since that term carries with it all sorts of physiological baggage which I do not want) a two stage procedure is needed. This takes the form of adding into the intuition an assumption about where the effects come from. In effect forces appear. If we continue with the cricket example for a moment we postulate some background which gives stability to the motion so that we can ask about the relationship of the deviation from the plane to whatever stabilises the plane dynamics. It is time I came clean. I am talking about electrodynamic interactions, and it doesn't matter that electromagnetism is not used in normal cricket.

The point to make is very primitive. When we have in mind that – say - a particle motion is attributable to a force in a particular dimension and that an acceleration on it is produced by a force which we say is at right-angles to the plane containing the original motion then we say that the force is due to a magnetic field. All the terms in this statement can be changed but the basic message remains a fundamental characteristic of our world and it is the distinctive aspect of electrodynamics which is stated in combinatoric terms similar to those in the purely mechanical description of the batsman. However the situation has passed beyond what was contemplated in the case of the batsman. By postulating knowledge of the initial forces, which is beyond the batsman's remit, we force a new and more complicated picture. We are using a perception and corresponding language, which is not provided by the first level, and have passed beyond that. This way of understanding the origination of electromagnetism explains the failure of Weyl to formulate a geometric account of it that

could extend Einstein's work on gravitation to give a unified picture. The gauge theory still left mechanics and electromagnetism as two obstinately different things, and so failed in its primary purpose.

At this point some readers may say "All this is a replacement for what common physiology does perfectly well in observation: are we really impelled to such flights of fancy?" Well: in a recent communication Lou Kauffman reminds us of (William) James's criticism of the commonsense notion that we may contemplate an unbroken chain of causation which leads from the mental act to the physical raising of an arm. He insisted (and he had good cause from personal experience to insist) that there would always be an unbridgeable gap. As Lou says (and James would agree) you might as well invoke psychokinesis *tout court.* For our case the argument has to be put round the other way because we are concerned with perception rather than with volition, but the same issues are involved.

Two results are assumed here which need careful scrutiny. One is the assertion that the principles of discrimination and of discriminate closure are really an effective elaboration of the order/absence-of-order principle. Here I mean that though they are not at all the same thing because the discrimination algebra will necessarily contain further structure, everything in the order principle appears in the discrimination closure argument. The second result (really a corollary of the first) is that the numbers arising from the levels in the old hierarchy interpretation already appear in the more primitive order/absence of order logic EXCEPT the use of cumulative sums which is characteristic of theory.

I do not know how to describe the order logic. It is not mathematics in the usual sense because the mathematician wants to ask questions which can only be answered by extending it (for example to a group). Donkey's years ago we thought it a great advance to incorporate it into a conventional algebra, which became the discrimination algebra, and of course it was. It was made possible by Frederick. However now we have to backtrack because of the urgent need to understand observation and the physical concomitants of level change.

Etter's thinking has been found invaluable already through his basic distinction of composition and extension. Now we find another point of contact in his relational as opposed to functional view of mathematical development. He says in describing Link Theory that one must replace functional parts by relational parts, and functional composition by relational composition. Our theory/physical distinction is surely comparable. In view of Etter's vagueness in applying his logical work to actual physics we should consider whether we can make a fundamental connection by identifying our 'physical' with his 'relational' and our theory with his 'functional'. We leave this suggestion open for discussion.

The 'aspect' extension of the algebra arose because of the need to understand the two quantum states, which give the characteristic spin algebra. It was clear long ago that spin was an (perhaps *the*) quintessentially combinatorial concept and that it needed to be given a combinatorial as distinct from a geometrical basis. Clive found quite recently that this was not possible without the aspect extension. The two spin states form part of the observer intuition.

The reader may worry at this point that the action of the participant observer requires space, whereas the spin is combinatorial. However this worry overlooks a real success. The sequential development in the PO is what gives algebraic or compositional form to the spatial intuition.

To summarise, we argue that the choice inherent in CO1 gave rise to the quadratic group but was itself the origin of dimension structure as experienced by the observer. The batsman is to be taken literally, but so is the spinor 'microscopic' theory. The spin account bridges the gap, which ought not to be seen as a matter of scale. The isometric space has to be filled in by sequences of level changes as is assumed in the construction, and this filling in presupposes the limit point since there is nothing left over from the volition of the participant observer to control the sequence. We seem to have a dispersal of the consciousness over whole sequences, which conflicts at first sight with the apparent concentration of one step, which is the commonsense notion. The participant observer consists of sequences of jumps back and forth between two levels, and the activity of the observer (1) to preserve the levels against indefinite proliferation of the sort which Clive is unable to prevent. Consciousness is only of the last stage at any moment but the whole sequence is present in some form since otherwise there would be no measure attached. We do not need to provide for storage of these numbers separately as we had supposed we had to do but could find no way of doing.

We see ourselves as providing an understanding of electromagnetism itself, since the observational concomitant of each stage in the measurement is the discovery that a 3D structure is needed to represent the 'field' responsible for the acceleration of the test particle. As we normally say, the particle is accelerated at right angles both to its motion and to the applied field - that being the definition of the e.m. field. Having seen this, we do the jump into purely mechanical picture with the measurement step now incorporated into the measurement sequence. This principle takes us back to the period after the publication of Weyl's gauge theory of electromagnetism when it slowly became apparent that the attempt to incorporate the e.m. effect into a relativity type extended geometry - could never succeed. We have had carefully to consider why it could never succeed, since the motivation of Weyl - his finding it

ultimately unsatisfactory that mechanics and electromagnetism admitted two kinds of stuff in the world - retains its force and provides a push in our direction.

8. SOME PERPLEXITIES REMOVED BY THE PHYSICAL/THEORY DISTINCTION

Our separation of the 'physical' from the 'theoretical' throws light upon three places where there has been a perplexing overdetermination. In two of these we seem to have arrived at an important result in two ways. The perplexity has been due to the apparent independence, or difficulty in seeing a relation between, the two arguments. These are:

1. The dimensionality of space (3D). From the physical direction we get a necessary level jump at 3 depending on our primitive order argument. At the same time it was necessary to identify the number of constructive levels before the Parker-Rhodes cut-off as the origin of this dimension number. The essential step here is that by our sequential view of the PO we can accommodate the first step in the construction in a realist way. Was the second deduction of 3D more than filling an explanatory gap? Does it still have any persuasive force? Is there still a mystery about the two appearances of the number 3? We could argue that it is a happy accident, but it may be that we should consider what we would do if the accident were unhappy. (A bit like considering what we would do if there were five primes between 0 and seven.)

2. Let us consider the correspondence between the CWK and the McG accounts of the non-integral part of the reciprocal of the fine-structure constant. The new theory enables us to take McG seriously without schizophrenia, in which case the correspondence becomes even more important than 1. suggests. May I remind you of the heady days perhaps 6 or 8 years back when a combination of David and Pierre would arrive with some fresh revelation of numerical successes in the domain of particle constants. All that work waits for a proper assessment. On the face of it, it is awe-inspiring. David saw it in terms of his ordering operator calculus which is very different from any current view of physical measurement, and one dare not say the calculations were not in some way real, but on the other hand the slow integration of the different viewpoints has not been done. The separation of 'physical' from theory casts a new light on this perplexity since McGoveran's calculations - once given the cumulative sums - use the physical numbers and so are separate from those of Kilmister which use theory.

3. Spin. The original spin argument can now be used and the whole of the aspect theory comes as a consequence. Clive's observation that discriminations do not have to be symmetric flows from the 'dimensional' account. The physical account, which was impossible in 'theory' and necessitated aspect, is now allowable. Hence we see aspect as required by our new separation of physical and theory. This strengthens our whole position since we can now use the aspect The connection of the macroscopic physics with quantisation is also strong. At the limit of each measurement sequence we reach a point at which no complexity other than the combinatorial structure can be presented; there is no room any more for any spatial configuration, and this is quantisation itself. Appropriately the coupling constant which defines the relationship between the quantised units is the measure of the coupling of the electromagnetic field specifically - namely $1/\alpha$. This limit to the observation sequence is the point at which we depend on having a combinatorial account of spin, and it was here that the 'aspect' generalisation of the hierarchy began (see ANPA 1998 Joint paper, CWK and TB). I think it is clear that we have here in principle the whole account of how the quantised electromagnetic quantities grade over into classical or 'continuum' ones, though it would be desirable to go into more detail. It should also be quite easy to fit in CPT and other basic quantum principles of that sort given that they no longer trail with them all sorts of implausible geometrical allusions.

INFINITESIMALS, ZERO NUMBERS AND THE DISCRETE ORDERED CALCULUS

LOUIS H. KAUFFMAN

Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 851 South Morgan Street, Chicago, IL 60607-7045, USA

1. INTRODUCTION

This paper is an expanded version of column number 11 of my series on Virtual Logic for the Journal "Cybernetics and Human Knowing." In that article I discuss the mathematical subject of Newton's calculus using the notion of infinitesimals as an imaginary form of number. A form that allows us to reason towards real answers and compute limits that would otherwise be ineluctable. I also discuss "zero numbers", a concept grounded in ordinary numbers that opens up the powers of zero so that 0, 0×0, 0×0×0, ... are all distinct forms of the void. We shall see that these subjects are related to one another and that they are both related to the perennial topic of this column - paradox resolution and the relation of imagination/imaginary values to properties of a system as a whole.

In this article, I begin with the material in the column for the cybernetics journal, and then continue with a discussion of the structure of the discrete ordered calculus of me and Pierre Noyes from the point of view of infinitesimal. My thesis in this article is that the theory of infinitesimal calculus can be regarded as a form of discretisation that nevertheless incorporates concepts of continuity.

First, I shall explain what sort of thing is an infinitesimal and how we will work with it. The idea of an infinitesimal was conceived by Newton and Leibniz in their invention of the calculus. The concept was that of a process (or calculation) whose results were getting ever-smaller so that, if inspected, the value of the infinitesimal would be smaller than any named positive real number, and yet the infinitesimal is not zero! Newton called them "fluxions", tiny vanishing particles of flux, particles of the action that lies at the basis of

the dynamic world. These vanishing particles are always in motion, elusive motion that avoids evaluation.

How can you imagine an infinitesimal? You can recognise yourself as a slow creature and think of the infinitesimal as a fantastically quick process. By the time you reach into your bag of numbers and pull one out for comparison, the value of the infinitesimal has disappeared way below the value that you have chosen. In fact, you cannot take a reading of the value of the infinitesimal, it moves too fast to allow any reading to take place. And yet there it is vanishing away before your very eyes! Bishop Berkeley, in a fit of criticism, called the infinitesimals "ghosts of departed quantities" and wondered how Newton, Leibniz and their descendants could muster the faith to believe in them.

In this article I shall present the idea of an infinitesimal as a kind of imaginary value related to ideal mathematical elements such as the square root of minus one or infinity.

In mathematics, almost every construction is "ideal" in the sense that it is an abstraction or idealisation of experience. But some constructions, such as the "vanishing point" on the horizon where parallel lines meet, are particularly worth emphasising as ideal or imaginary because they connote a place where we extend a given structure by adding new elements to it that inform that very structure. In self-observing systems of the human kind such an addition is the notion of a self. You will not find the self inside the system. Nor is it outside the system. It is not quite illusory either! The self is imaginary and because it is authentically imaginary, the self is real. In this way the study of infinitesimals is directly related to cybernetic concerns of emergence of values that relate to the structure of a system or a language as a whole.

Along with infinitesimals, we shall study the powers of zero in a system that says that zero zeros are not the same as zero! The concept of zero numbers is due independently to George Spencer-Brown [1] and to Frederick Joseph Staley [2]. Apparently, Spencer-Brown discovered the zero numbers in the course of extending Laws of Form to encompass ordinary arithmetic, while Joe Staley found them in contemplating the void as an underpinning of Mathematics. There is a curious affinity between infinitesimals and zero numbers, even though they are quite different. (Infinitesimals are not zero, but they adhere to zero. Powers of zero are not zero, but they are not at any distance from zero.)

Finally, we bring together the infinitesimal calculus based on invertible infinitesimals and the discrete ordered calculus (DOC) of the author and Pierre Noyes.

In the infinitesimal calculus the derivative is defined without limits by the equation

$$\mathbf{df/dt = Re[(f(t + Z) - f(t)/Z]}$$

where Z is an infinitesimal time (t) displacement, and Re[A] denotes the real (non-infinitesimal, non-infinite) part of an extended real number A.

In DOC (using infinitesimals) the derivative is defined by the equation

$$\mathbf{Df(t) = J[(f(t+Z) -f(t)//Z]}$$

where J is the time-shift operator with the property

$$\mathbf{Jf(t+Z) = f(t)J.}$$

Both the DOC D and the classical d/dt satisfy the Leibniz rule (see the section on DOC for the definition of this rule) making their calculi suitable for many applications. Our fundamental result in this paper is the equation

$$\mathbf{Re[Df(t)] = Jdf/dt.}$$

This result binds the continuous, commutative world with the non-commutative, discrete micro-world. This is a beginning that we shall explore further in work to come.

2. THE PARADOX OF DIVISION BY ZERO

ONE/ZERO
= (ZERO × ONE) / (ZERO × ZERO)
= ZERO / (ZERO × ZERO)
(since ZERO × ZERO = ZERO)
= ZERO / ZERO = ONE

Hence ONE / ZERO = ONE

Hence ONE = ZERO × ONE = ZERO.

But we cannot have ONE = ZERO!

So we say that you cannot divide by zero.

Here it is assumed that ZERO × ZERO = ZERO.

If you have the notion of subtraction then this fact is forced upon you, for
$0 = 1 + (-1)$
whence
0×0
$= 0 \times (1 + (-1))$
$= 0 \times 1 + 0 \times (-1)$
$= 0 + 0$
$= 0.$
This calculation
shows that $0 \times 0 = 0$ is a consequence of

$$\mathbf{a(b+c) = ab + ac}$$

(this is called the distributivity of multiplication over addition), and the principle that $0a = 0$ for any non-zero a.

We see then that we could attempt to let go of $00 = 0$ in a context of only positive or zero numbers since in that context there is no proof that $00 = 0$ in the form shown above. In other words, if we work in positive arithmetic then zero is not seen as a combination of two integers. Any time you add two positive integers you get another positive integer. The positive integers including zero is called the "natural numbers" by mathematicians. Up until the middle fifteen hundreds the concept of negative numbers was barely heard of. Even today, most people are not fully comfortable with negative numbers. (After all who wants to have a negative bank balance?!) Elementary counting happens first without subtraction and without the concept of an absence as a something. This will be the context of the zero numbers that we shall discuss.

Here is Joe Staley [2] on the concept of the zero numbers:

"According to arithmetic 00=0, zero times zero is just zero. On the other hand, zero times anything is just zero number of that thing, none of them; so zero times zero says no zeros, none, not one of them. But zero is one zero, not none.

It would have been possible to resolve this contradiction by several arguments. We might say that 00=0 fits into a consistent mathematical system, so there. Or we might argue about the difference between zero as an object of multiplication, zero times as a verb, and '0' as a symbol, and so on. It was possible, however, to take another path, the path of envoidment, of accepting 00 not equal to 0 and then developing the consequences of a further discrimination (00 < 0), the path taken in the following.

The result of this process of envoidment is to produce a substantial extension of the known numbers. We develop here an arithmetic of numbers smaller than zero but larger than any negative number. This number set has a rich and resonant set of strata, an architecture of sudden ellipses, and a new foundation for the natural numbers, wherein '1' is found as the mutual limits of ascending and descending sequences of zero numbers."

We repeat:

Zero times anything
Is just zero number of that thing,
None of them;
So zero times zero says
No zeros,
None,
Not one of them.

Thoughts like this are necessary in order to enter the realm of an observing system, a system that can reflect on its own use of language.

3. INFINITESIMALS

On the other hand, if we replace zero by an infinitesimal number Q that is not equal to zero, then we do not have $Q \times Q = Q$ and so if we were to write

$$Q / Q \times Q$$

this would just be the same as 1/Q.

Of course since Q, Q × Q, Q × Q × Q, ... give smaller and smaller infinitesimals, it follows that their reciprocals 1/Q, 1/ Q × Q , 1/ Q × Q × Q, ... will be larger and larger infinities. We can live with this! Adding such numbers will enlarge and enrich the number system and there will not be any problems related to the division by zero. Infinitesimals are not zero.

We could go into this domain but I am going to postpone it and enter the next section where we work with the simplest possible sort of infinitesimal, one whose square is zero. Dear reader, please bear with me. This next turn will reward your attention.

4. SQUARE ZERO INFINITESIMALS

Now I want to talk about an even stranger infinitesimal.
I will denote it by Z.
The square of Z is zero!

$$\mathbf{Z\,Z = 0.}$$

You can think that Z is so very small that when you multiply Z by itself it just vanishes. (After all a very small number times a very small number is a much smaller number, so Z is just the limiting case of this phenomenon.) But Z itself is distinct from zero, and Z/2 is smaller than Z, Z/4 is smaller than Z/2 and so on. You cannot divide by Z.

The infinitesimal Z is perfectly suited to doing calculus.

Now why should we dwell on this matter of the calculus? The answer is that it speaks to the fundamental nature of the moment. For calculus is mathematics devised to capture in quantitative terms the dynamics of transition from moment to moment to moment. Newton found the calculus in his attempt to fathom gravity and the motions of the moon, the planets, the stars and the objects of our everyday space. The transition from moment to moment is mediated in this model by an infinitesimal and so the next moment

of time T+Z is infinitely close to the present moment T. So close that they are glued together, and yet they are distinct.

In calculus we are concerned with measuring how a function of a variable t changes when the t value changes by an infinitesimal amount. A function f(t) is just a rule that assigns to each value of t a new value f(t). For example, f(t) could be the velocity of a bird flying by your window at the time t.

We are concerned with the difference f(t+Z)-f(t), the amount that f(t) changes in an infinitesimal amount of time. We make the assumption that an infinitesimal change in t will result in an infinitesimal change in f(t). Thus we write

$$\mathbf{f(t+Z) - f(t) = f'(t)Z}$$

and call f'(t) the derivative of f at t. We say that a function is differentiable at x if there is a well-defined new function f'(t) that satisfies this equation for any infinitesimal Z.

A simple example of differentiability is illustrated by the function

$$\mathbf{f(t) = t^2 = tt.}$$

Here we have

$$\begin{aligned} &\mathbf{f(t+Z) - f(t)} \\ &\mathbf{= (t+Z)(t+Z)-tt} \\ &\mathbf{= tt +Zt +tZ +ZZ -tt = tZ+Zt \text{ (since } Z\times Z=0)} \\ &\mathbf{= 2tZ} \end{aligned}$$

Thus

$$\mathbf{f(t+Z) - f(t) = 2tZ.}$$

So by our definition of f', f'(t) = 2t in this example.

We have used this calculation involving the ideal infinitesimal Z to reason to the very real answer that an (accelerating) bird whose position is given by the square of the time will have velocity equal to twice the value of that time. This is innocent enough for the bird at time zero, but I would not want to be in the way of that same bird after one minute. If you replace the bird by a safe

dropped from the tenth story of your apartment building, then the mathematics is essentially the same, and we all know that it is a bad idea to be under that safe. The calculus lets you find the exact values related to your find out about the orbits of the planets and the behaviour of electrons, radio waves and all manner of varying things.

The whole structure of the calculus can be based on this simple system of infinitesimals. Eventually this will be the way calculus is taught in the schools.

To see how efficient this is, the reader already familiar with the calculus will be astonished at the simplicity of the proof of the chain rule:

Let

$$F(x) = f(g(x).$$

Then

$$F(x+Z)= f(g(x+Z))= f(g(x)+g'(x)Z) =f(g(x)) + f'(g(x))g'(x)Z$$

Thus

$$F'(x) = f'(g(x))g'(x).$$

This is the so-called chain rule, a rule for differentiation that is often left unproven in calculus texts due to the complexity of the usual derivation.

We have, in this section, gone very quickly into the basics of the calculus.

The real point of this section is the structure of the infinitesimal change. By allowing an imaginary number Z with the property that $ZZ = 0$, we illuminate the structure of the moment and the movement of time. It remains for the reader and the author to continue this meditation in both mathematical and cybernetic directions.

Now we are going to look more carefully at the previously mentioned concept of zero numbers. Zero numbers arise naturally in the context of Spencer-Brown's Laws of Form [1]. Laws of Form has been a constant topic in this column because it is a cogent mathematical expression of the idea that "a

universe comes into being in the making of a distinction". That universe is the universe of an observing system. The universe itself (in the largest sense) is one turn of the world into a that that sees and a that that is seen. Those thats (in a plethora of reference) are different from each other, and yet they are the same. The difference between the universe that sees and the universe that is seen is infinitesimal.
Or perhaps it is a power of zero!

5. ZERO NUMBERS

Recall that zero numbers are distinct powers of zero.
Here is how zero numbers arise for Spencer-Brown. He begins by letting <> (the mark in his notation) denote the number one. Then <><> denotes two and <><><> denotes three and so on with a row of n marks denoting the number n: n = <><><>...<>, 0= ,1=<>, 2=<><>, 3=<><><>, 4=<><><><>, 5=<><><><><> ad infinitum.

It turns out that

$$0\text{^}n = 0^n = 0\times0\times0\times0\times...\times0 = <<><><>...<>>.$$

That is, 0^n is denoted by a bracket around a row of n marks. To see this we need to do a little arithmetic in the style of Spencer-Brown.

In the Spencer-Brown arithmetic (henceforward referred to as SB arithmetic) it is given that

<<A>> = A for any A,

<<A><B>>C = <<AC><BC>> for any A, B, C

and AB = BA for any A and B.

The reasons for these rules will become quickly apparent.

Addition is defined by A+B = AB.

Multiplication is defined by A×B = <<A><B>>.

Watch the results of 2×3:

$$2\times3 = \langle\langle 2\rangle\langle 3\rangle\rangle =$$
$$\langle\langle \diamond\diamond \rangle \langle 3\rangle \rangle = \langle \langle \langle\langle 3\rangle\rangle \langle\langle 3\rangle\rangle \rangle \rangle$$
$$= \langle\langle 3\rangle\rangle\langle\langle 3\rangle\rangle = 3\ 3 = \diamond\diamond\diamond\ \diamond\diamond\diamond$$
$$= \diamond\diamond\diamond\diamond\diamond\diamond = 6.$$

It should now be clear how

Crossing: $\langle\langle A\rangle\rangle = A$

and

Transfer: $\langle\langle A\rangle\langle B\rangle\rangle C = \langle\langle AC\rangle\langle BC\rangle\rangle$

interact to create the structure of multiplication in SB.

Now look at $0\times0 = \langle\langle 0\rangle\langle 0\rangle\rangle$.
Since 0 = "nothing", we conclude that $0\times0 = \langle\diamond\diamond\rangle$.
This justifies our interpretation of 0^n as $\langle \diamond \diamond \diamond \ldots \diamond\rangle$.
Zero numbers live naturally in the SB arithmetic.

Now consider $\langle 3\rangle A$ for any number A:

$$\langle 3\rangle A = \langle \diamond\diamond\diamond \rangle A = \langle\langle A\rangle\langle A\rangle\langle A\rangle\rangle = A\times A\times A = A^3.$$

By using Transfer, we have found that

$\langle n\rangle A = A^n$ for any numbers n and A.

Exponentiation is a natural consequence of the interaction of zero numbers and ordinary numbers!

Zero numbers are quite analogous to infinitesimals. But they are a realm unto themselves. For the product of a zero number and any ordinary number is simply zero. And the sum of a zero number and any ordinary number is simply that number. Thus it is with some joy that we can point to the exponential relation

$$0^n.A=A^n$$

as a genuine form of communication between the void-realm of the zero numbers and the solid realm of ordinary numerical reality.
Note, by the way, that

$$0^0 = 1.$$

For

$$0^0 = \langle\rangle = 1$$

since zero is void.

An note also that since $0^n.A=A^n$, we have $A^0 = 0^0.A = \langle\rangle.A = \langle\rangle = 1$.
The reason that

$$\langle\rangle.A = \langle\rangle$$

is that the expression $\langle\rangle.A$ is an instruction to transfer A under each mark that is inside $\langle\rangle$. Since $\langle\rangle$ contains no marks, we transfer A into the void!

Note well that a special case of this last result is the equation

$$\langle\rangle.\langle\rangle = \langle\rangle,$$

expressing the fact that $1^0 = 1$. This is not to be confused with $\langle\rangle\langle\rangle$ which represents 2 as 1+1. In general this means that the arithmetic consisting of zero numbers and numbers is non-associative (since $(\langle a\rangle.b)c = b^a + c$ while $\langle a\rangle.(bc) = (b+c)^a$) and has three operations +, × and . where . indicates exponentiation via combination with zero numbers.

Now we will consider negative numbers.

We let [] denote -1 so that

$$\langle\rangle[] = []\langle\rangle = \quad = 0.$$

Then we have

$$1/0 = 0^{(-1)} = 0^{[]} = <[]>$$

(1/0 has a formal existence after we introduce negative numbers!)

and

$$1/a = a^{(-1)} = <[a]>.$$

Now this implies that

$$[a] = <<[a]>> = <1/a> = 0^{(1/a)}.$$

Negative brackets around the number a indicate zero to the reciprocal of a!

For example,

$$[] = 0^{(1/0)}$$

and so we have the wonderful formula

$$0^{(1/0)} = -1.$$

Note what happens when we multiply -1 × -1 :

$$-1 \times -1 = < <[]> <[]> > = < < [<[]>] > > = [<[]>] = 0\text{^}(1/(0^{(-1)}) = 0^{(0)} = 1.$$

$$-1 \times -1 = 1.$$

As we have previously remarked, there is a paradox if we construct zero numbers and negative numbers in the same system. Here is the paradox in this notation.

$$0\times0 = < \diamond \diamond > = < \diamond < \diamond [] > > = < < <\diamond\diamond> [\diamond] > > = <\diamond\diamond> [\diamond]$$

$$= [\diamond] = 0^{(1/1)} = 0^{(1)} = 0.$$

In order to avoid this paradox, we have to limit the distributive law. It is clear that one way to do this is to disallow the transfer

$$a< [b] <c>> = <[ab] <ac> >$$

where there is a mixture of negative and positive brackets. It is clear from the beauty of the results about positive and negative numbers that it is worth pursuing this resolution. There is one more beautiful paradox inherent in the movement to negative and zero numbers. Consider the expression

$$<>.[] = 0^{(-1)} = (-1)^0.$$

By void transfer we could send [] into the void in <> or we could send <> into the void in []. In one case the result is +1. In the other case the result is -1. This means that in this system <>.[] is a new imaginary value, neither zero nor one. The remainder of the resolution of these paradoxes in the neighborhood of the void will appear in another paper.

6. NOW TOGETHER

We end this essay with an injunction to the reader that she think on both infinitesimals and the zero numbers and feel the presence of this realm of number prior to number, prior to geometry. A realm that has a delicate beginning and powerful possibilities for calculation and for articulation of the subtlety of the moment, an articulation of the subtlety of the relation between the observer and the observed who are the same and yet not the same.

I am sure that all that we have said seems tantalising to some readers (as it does to this author). For one begins to feel that here in the presence of infinitesimals and zero numbers one is at the gateway to genuinely new insights into the relationship of mathematics and reality. Surely the infinitesimals have something to tell us about the very small! They should tell us about the Planck scale in quantum physics, about ultimate units of space and of time. In asking these questions we enter into a construction site. Some aspects are well worked out, like all the known uses of the calculus. Other aspects take a new flavour, and some are just beyond the reach of saying. Let me finish with an example.

It is well known to mathematicians and physicists that infinitesimals in different spatial directions should not commute with one another! In particular it is well-known that if dx and dy represent infinitesimals pointing in perpendicular spatial directions then we should have the rule

$$\mathbf{dxdy = -dydx.}$$

Now don't ask why! (Unless you want to take courses for five years.) But from the point of view of square zero infinitesimals this is not a mystery. After all we are given that

$$\mathbf{dxdx = 0 \text{ and } dydy=0.}$$

And surely (dx + dy) should also be a square zero infinitesimal since it just points in the sum of the spatial directions of dx and dy. But this means that

$$\begin{aligned} \mathbf{0} &= \mathbf{(dx+dy)(dx+dy)} \\ &= \mathbf{dxdx + dxdy + dydx + dydy} \\ &= \mathbf{dxdy + dydx.} \end{aligned}$$

So

$$\mathbf{0 = dxdy + dy\,dx.}$$

This is the same thing as saying that

$$\mathbf{dxdy = -dydx.}$$

Down there in the tiniest parts of the world processes do not commute with one another. No way around this, and this non-commuting is where quantum mechanics starts. We can start with the idea that space is non-commutative in the small, understanding that this comes from the gluing of moment to moment in the realm of the observer, and begin to articulate a quantum story of the world. This is work in progress. See References [3] through [8]. Even in these references I had not quite understood the importance of infinitesimals. The invitation is out. Think on these matters. Let the author of this essay know your thoughts. There is a lot to come.

7. AN INFINITESIMAL VERSION OF THE DISCRETE ORDERED CALCULUS

We now turn to the discrete ordered calculus of the author and Pierre Noyes in order to look at it from the point of view of infinitesimals. First let us recall the set-up for this calculus. We assume that an elementary time step of size Z is given and that for a function f(t) we have the difference quotient Df(t)Z =

f(t+Z)-f(t). We then examine the behaviour of the difference quotient of a product function F(t) = f(t)g(t):

$$\mathbf{DF(t)Z = f(t+Z)g(t+Z) - f(t)g(t)}$$

$$\mathbf{=f(t+Z)g(t) -f(t)g(t) +f(t+Z)g(t+Z) - f(t+Z)g(t)}$$

$$\mathbf{=((Df(t))g(t) + f(t+Z)Dg(t))Z.}$$

We see from this calculation that in general D(fg) is not equal to D(f)g + fD(g).

The equality

$$\mathbf{D(fg) = D(f)g + fD(g)}$$

is called the Leibniz rule.

The Leibniz rule is not true for arbitrary difference quotients. If Z is infinitesimal then the calculation above does show that the difference between D(fg) and D(f)g+fD(g) is itself infinitesimal since f(t+Z) differs infinitesimally from f(t). Note however that the assumption that f(t+Z) differs infinitesimally from f(t) is essentially the assumption that f(t) is a continuous function of t. Thus continuity (or differentiability) and the Leibniz rule are closely intertwined.

Furthermore, the Leibniz rule is precisely satisfied in the realm of square zero infinitesimals for here we have

$$\mathbf{D(fg)(t)Z = ((Df(t))g(t) + f(t+Z)Dg(t))Z}$$

$$\mathbf{=((Df(t))g(t) + (f(t) +Df(t)Z)Dg(t))Z}$$

$$\mathbf{=((Df(t))g(t) + f(t)Dg(t))Z}$$

since ZZ=0.

On the other hand lets suppose that Z is not a square zero infinitesimal, but rather that Z is an infinitesimal with an inverse 1/Z so that

$$\mathbf{Df(t) = (f(t+Z) - f(t))/Z.}$$

The usual way that ordinary derivatives are constructed in the infinitesimal calculus (of invertible infinitesimals) is by the definition

$$\mathbf{f'(t) = df/dt = Re[Df(t)]}$$

where **Re[A]** for an extended real number A denotes its real (non infinitesimal, non infinite) part. The real part of A is uniquely defined, and this method gives a definition of derivative that is free from limits.

Our method of discrete ordered calculus can be applied to the infinitesimal calculus to produce a non-commutative differential calculus with values in the extended real line. We simply follow the same procedure that produced DOC. We introduce a time shifting operator J with the property that

$$\mathbf{Jf(t) = f(t+Z)J.}$$

In other words

$$\mathbf{f(t+Z) = Jf(t)J^{-1}}$$

where J^{-1} denotes an inverse operator for J.

Then we redefine D via the equation

$$\mathbf{Df(t) = (Jf(t+Z) - Jf(t))/Z = (f(t)J - J(f(t))/Z = [f(t),J]/Z.}$$

It then follows algebraically that D satisfies the Leibniz rule. Note that since Z is retained in these calculations it takes the role of a discrete time step. We can simultaneously handle both the discrete ordered calculus and the standard infinitesimal calculus in a single formalism!

It may seem a bit puzzling that we have before us in the same formalism two distinct ways to rehabilitate the Leibniz rule. One way leads to classical physics. The other leads to an underpinning for quantum mechanics. This is where more work needs to be done.

Lets examine this formalism in the case of invertible infinitesimals. Then

$$Df(t) = (Jf(t+Z) - Jf(t))/Z = J\,[(f(t+Z) - f(t))/Z].$$

Thus if we define

$$Re(JA) = JRe(A)$$

for an extended real number A, then

$$Re(Df(t)) = JRe\,[(f(t+Z) - f(t))/Z] = Jdf/dt.$$

Thus

$$Re[Df(t)] = Jdf/dt.$$

This is the fundamental relationship between the DOC derivative and the classical derivative. The DOC derivative contains infinitesimal terms but continues to satisfy the Leibniz rule. Its real part is essentially the classical derivative. Thus the DOC derivative is an infinitesimal extension of the classical derivative that preserves the Leibniz rule.

Lets see how this works in a simple case. Let

$$f(t) = t^2.$$

Then

$$Df(t) = J((t+Z)^2 - t^2)/Z) = J((2tZ + z^2)/Z) = J(2t + Z).$$

$$D(t)t + tD(t) = J(1)t + tJ(1) = Jt + tJ = Jt + J(t+z) = J(2t+Z) = D(t^2).$$

Here we see quite explicitly how the time shifting of the operator J keeps the Leibniz rule in place and that this time-shifting occurs at the infinitesimal level that is nevertheless a discrete level in the mathematics. Our combination of the continuous and the discrete promises to prolong to a unification of discrete and continuum physics and mathematics.

Obviously much more work needs to be done in this domain. This paper has been an initial exploration of significant territory intermediate between the continuous and the discrete.

8. POSTSCRIPT

For more about the Spencer-Brown Arithmetic, the reader can consult [1], [9] and [13], [14], [15]. For an extraordinary construction of all numbers great and small (infinite and infinitesimal) read the amazing book On Numbers and Games by John Horton Conway [10]. The book by Henle and Kleinberg [11] is an excellent introduction to modern theories of infinitesimals based on the work of the logician Abraham Robinson.

Finally one of the earliest American proponents of infinitesimals was the philosopher/mathematician Charles Saunders Peirce. We let him have the last word: ([12] Vol. 4, p. 355)

"... the conception of continuity involves no contradiction and cannot be dispensed with. From this discussion flows the irresistible consequence that infinitesimals exist wherever there is continuity. The logic of the differential calculus is set forth from this point of view. The doctrine of limits is not denied. .... But it is shown to be an unnecessarily roundabout way."

Infinitesimals exist wherever there is continuity.
Let us be direct from now on.

9. REFERENCES

1. G. Spencer-Brown. Laws of Form - German Edition -- Gesetze der Form, Bohmeier Verlag (1997) (see included article "An algebra for the natural numbers" original written in 1961).

2. F. J. Staley, Zeros of the Void -- A First Report on Zero Numbers (1987) (unpublished).

3. Kauffman, Louis H. and Sabelli Hector C. The Process Equation. Cybernetics and Systems, Vol. 29. (1998) pp. 345-362.

4. L. Kauffman, Noncommutativity and Discrete Physics, Physica D 120 (1998), pp. 125 - 138.

5. L. Kauffman, Space and time in computation and discrete physics, Intl. J. Gen. Systems, Vol. 27, Nos 1-3 (1998), pp. 249 -273.

6. L. Kauffman, Discrete Physics and the Derivation of Electromagnetism from the Formalism of Quantum Mechanics, (with H. P. Noyes), Proceedings of the Royal Soc. London A, Vol. 452 (1996), pp. 81-95.

7. L. Kauffman, Quantum Electrodynamic Birdtracks. Twistor Newsletter No. 41. (1996)

8. L. Kauffman and P. Noyes, Discrete physics and the Dirac equation, Physics Lett. A, No. 218 (1996), pp. 139-146.

9. L. Kauffman, Arithmetic in the Form, Cybernetics and Systems, vol. 26 (1995), pp. 1-57.

10. J. H. Conway, On Numbers and Games, (1976) Academic Press, New York.

11. James Henle and Eugene Kleinberg, Infinitesimal Calculus, (1980) MIT Press, Cambridge, Mass.

12. Charles S. Peirce, The New Elements of Mathematics, edited by Carolyn Eisle, Volume 4 (1976) Mouton Publishers, The Hague, Paris.

13. Jack Engstrom, Natural Numbers and Finite Sets Derived From G. Spencer-Brown's Laws of Form, Masters Thesis, Maharishi University of Management (1994).

14. Jack Engstrom, G. Spencer-Brown's "Laws of Form" as revolutionary unifying notation. Semiotica (1999). pp. 33-45.

15. Jeffrey James, A Calculus of Number Based on Spatial Forms, Masters Thesis - University of Washington (1993).

HUYGENS' PRINCIPLE, PHYSICS AND COMPUTERS

KEITH BOWDEN

Theoretical Physics Research Unit,
Birkbeck College, Malet St, London WC1E 1HX, UK.
(k.bowden@physics.bbk.ac.uk)

(Reprinted from International Journal of General Systems,
Special Issue on Physical Systems, Vol. 27, Nos.1-3, 1998, pp. 9-32)

We discuss the emergence of topology from a consideration of set extensions in General Systems Theory. Boundaries arise in a natural way, separating independent elements or regions of the system. Our aim is a unification of Etter theory, Kron's method of Tearing and Jessel's formulation of Huygens' Principle. This should make explicit the equivalence between the objective, structural, holographic and the subjective, relative definitions of information, sought in [Bowden, 1994b], reprinted in this Special Issue. It connects the abstract generalisations of Schrodinger's equation and Born's rule derived in probabilistic Etter theory with the real world of electrical and other physical phenomena in General Physical Systems Theory. This paper can be considered as a continuation of [Bowden, 1990] and [Bowden, 1994a] or [Etter, 1996] and as a response to [Bowden, 1994b], reprinted in this issue.

We review the ideas behind Kron's Method of Tearing and Jessel's Principle of Secondary sources (both special cases of the above theory) and their equivalence. We follow Hiley's argument in [Hiley, 1996] to show how Schrodinger's equation can be thought of as specifying the evolution of (a series of) tearings in continuous space. These can be shown on a commutative diagram as a series of similarity transforms. We compare this with Etter's derivation in [Etter 1996]. We describe briefly a recently published derivation of Maxwell's equations from a noncommutative algebra and show how they fit onto a related commutative diagram. Finally we make some comments on applications of the general theory to computer systems. This paper is a series of vignettes of work in progress. It is designed to point the direction of work to come in Constructive Physics.

KEYWORDS: Kron, Tearing, Diakoptics, Networks, Jessel, secondary sources, Huygens' Principle, holography, information, boundaries, Homology, Physics, Schrodinger's equation

1. INTRODUCTION

In the 1950's Gabriel Kron [1963] introduced his Method of Tearing or Diakoptics. The technique enables the solution of Engineering problems, particularly boundary value problems, by tearing structures topologically into subsystems, solving the subproblem on each subsystem and then recombining into an exact overall solution.

It has been recently rediscovered by the parallel processing community as "Domain Decomposition". The problem always took the form of an (electrical) network analogy or discretisation, thus Kron was one of the pioneers of fundamental theory of the Finite Element Method. (Indeed Roth (or Tonti) diagrams are proving to be of basic importance in modern Finite Element theory, e.g. see [Bossavit, 1995]. This will be discussed in a forthcoming paper.) Kron's claims for the generality of the Method of Tearing caused some controversy until its proof by the mathematician J. Paul Roth [1955] in the context of Homology Theory or Algebraic Topology. In a recent paper [Bowden, 1990] in this Journal it was shown how Kron's Method is equivalent to Jessel's Principle of Secondary Sources, a reevaluation of Huygens' Principle prompted by the physicist Louis de Broglie. Jessel's Principle states that any wave source distribution (e.g. charge) can be replaced by a set of secondary sources on any boundary surrounding the original sources, such that from the outside the resulting waveform is indistinguishable from the original. This is the basis of Dennis Gabor's holography.

Both Kron and Jessel saw their work as natural philosophies. They are both information based theories of Engineering. Both are rich, yet different analyses, one in discrete space and one continuous. They are compatible with the hierarchical, structural, systemic definitions of information within physical structures due to Scarrot

"A system is an interdependent assembly of elements and/or systems. Information is that which is exchanged by the components of a system" [Scarrot, 1989]

"The information contained by a system is a function of the linkages binding simpler into more complex units. The universe is organised into a hierarchy of information levels" [Stonier, 1990]

and Stonier. These considerations, in the light of the information based physics of David Bohm and Basil Hiley [1994], and also of new tentative theories of discrete informational physics, such as that of Ted Bastin and Clive Kilmister [1995] (who derive physical structure from the process of observation or innovation) led the author to propose that it would be a major achievement [Bowden, 1994] to relate these hierarchical, holographic, topological ideas of information to the relative, subjective, statistical, and often discrete, traditional concepts of information such as that of Shannon or Parker-Rhodes

"The information content of a statement is the number of YES/NO questions to which the answers are needed in order to identify the statement" [Parker-Rhodes, 198?]

In this issue Etter [1996] appears to have achieved this goal as a side effect of his search for a consistent, statistical interpretation of Quantum Mechanics. He shows how topological information can be reconstructed from data about statistical physical variables. In the process he shows how what are usually thought to be two empirical statements in orthodox physics, Schrodinger's equation and Born's rule, are really general theorems about links and tearing.

We have described three informational theories of the physical world, one discrete, one continuous and one statistical. The main aim of this paper is to provide a unified theory, which we call General Physical Systems Theory, as its structure is based on the work of Mesarovic and Takahara, although clearly the consideration of the emergence of topology puts a very different light on its applicability. We find that we can make very general statements about independence, regions, boundaries, decomposition and tearing without any reference to the kind of system to which we are referring.

The second part of the paper follows an argument by Basil Hiley in Hiley [1996] which starts from Grassman's idea of a simplicial complex as modelling thought processes and ends by deriving Schrodinger's equation (which, of course, is about *observation*) from changes in this structure. We reproduce the results but emphasising our approach, in particular with respect to changes in the structure being due to the evolution of a succession of tearing operations. We use the opportunity to remind the reader of the mathematical structure of Kron's network theory (which indeed is isomorphic to the structure of simplicial complexes), the Method of Tearing, and of Jessel's Principle (which, of course, are two of the important special cases of our general theory).

Finally we note that the lack of commutativity in the variables implied by Schrodinger's equation leads directly, according to Feynman and Dyson [1990] in the continuous case and Kaufman and Noyes [1996] in the discrete case, to Maxwell's equations. We suggest a new approach to their derivation using general noncommutative definitions of grad, div and curl. These should make clearer the structure of the derivation. The noncommutative field equations can be displayed on an extended Roth's diagram identical to the one in [Bowden, 1990]. These ideas will be investigated in more detail in a forthcoming paper.

2. HUYGENS' PRINCIPLE

A physical (or topological) system $\Sigma=(S,X)$ is a set of variables $X=\{x_1,x_2,...,x_n\}$, where each variable x_i takes its value(s) from a set $X_i=\text{span}(x_i)$, where $\text{span}(x_i)$ is

defined as the set of all possible values of x_i, for all possible values of the other variables, and an "extension" or set of conditions on X which can be written S(X)=0. (Often S(X) will be nonzero valued in our examples below; but there is no inconsistency here as the value of S(X) can always be subtracted from the left hand side and S redefined appropriately.) The x_i may be real valued or discrete (eg, Boolean). They can be static or dynamic (indexed on continuous or discrete time and realised as polynomials in s, the Laplace transform or z^{-1}, the backward shift operator). (In [Bowden 1990] we noted that this could be extended to indexing on (continuous or discrete) space.) And they can be deterministic or stochastic (probabilistic), as in the following examples:

Example Mesarovic and Takahara [1975] consider the general deterministic system $\Sigma_d=(S, X)$ with $S\subset X_1\times X_2\times...\times X_n$, where equality would give the trivial system with all the variables independent. However, in their analysis, they do not consider the emergence of topology.

Example Dempster and Schafer [Sudkamp, 1992] consider the general probabilistic system $\Sigma_p=(S, X)$, where the variables are propositions with a probability of being true, $p(x_i)\in[0,1]$, the closed unit interval, and $S:X_1\times X_2\times...\times X_n\rightarrow[0,1]$, giving the joint probability $p(x_1,x_2,...,x_n) = p(x_1\&x_2\&...\&x_n) \in [0,1]$. They define a *basic probability assignment* S over X as a function $p:2^X\rightarrow[0,1]$ that satisfies (i) p(0)=0 and (ii) $\Sigma_{A\subset X}p(A)=1$ (where Σ here is summation). Etter [1996] considers the emergence of topology from a joint probability function; it is on this work that our ideas are based.

Σ_d can be considered as either a special case of Σ or of Σ_p by putting $S:X_1\times X_2\times...\times X_n\rightarrow\{0,1\}\subset[0,1]$ with 1 for $\{x_1, x_2, ..., x_n\}\in S\subset X_1\times X_2\times...\times X_n$ and 0 otherwise. Conversely Σ_p can be considered as a special case of Σ_d by considering the joint probability to be another variable, that is, appending I=[0,1] to the span of the system and putting $S\subset X_1\times X_2\times...\times X_n\times I$.

It is this extension, or set of conditions, that gives rise to the topology of the system as we will show. It should be clear that our general definitions are all categorical statements in that they deal with generic (polymorphic) data types. The crucial concepts are

Independence Two variables are said to be independent if directly changing the value of one has no effect on the value of the other. Another way of saying this is that information from x_i is never received (or, at least, is always ignored) by x_j and vice versa. (Note that this is a stricter definition than that of Etter who defines two variables to be dependent even if they are only both affected by a third, (that is we define all variables to be inputs, see below). This is related to the difference between

$\partial x_i/\partial x_j=0$ and $dx_i/dx_j=0$.)

Example Two variables x_1 and x_2 in Σ_p are said to be independent iff $p(x_1\&x_2)=p(x_1)p(x_2)$. A set of variables x_i, x_2, ..., x_n is said to be independent iff $p(x_1\&x_2\&...\&x_n)=p(x_1)p(x_2)...p(x_n)$.

Example Two variables x_1 and x_2 in Σ_d are said to be independent iff $S = span(x_1,x_2) = span(x_1)\times span(x_2) = X_1\times X_2$. A set of variables x_1, x_2, ..., x_n is said to be independent iff $span(x_1,x_2,...,x_n)=span(x_1)\times span(x_2)\times...\times span(x_n)$, where $span(x_i)$ is defined as the set of all possible values of x_i, for all possible values of the other variables. A special case of Σ is the system with $S = X_i \times T$ where $T\subset X_1\times...\times X_i'\times...\times X_n$ and the prime indicates a missing term. In this case the variable x_i is said to be independent of the rest of the system.

Now consider Σ_d as a either a special case of Σ_p by putting $S:X_1\times X_2\times...\times X_n\rightarrow\{0,1\}\subset[0,1]$ with 1 for $\{x_1, x_2, ..., x_n\}\in S\subset X_1\times X_2\times...\times X_n$ and 0 otherwise as described above. Similarly $S:X_i\rightarrow\{0,1\}$ with 1 for $\{x_i\}\in X_i$ and 0 otherwise, thus $span(x_i)=\{x_i, S(x_i)=1\}$. The values 0 and 1 can be considered to be the *unnormalised* probabilities that the variables can take the associated values. Then two variables x_1 and x_2 in Σ_d are independent iff $S(x_1,x_2) = S(x_1)S(x_2) = S(x_1)\&S(x_2)$. We may write $p(x)=S(x)$. A set of variables x_1, x_2, ..., x_n is said to be independent iff $span(x_1,x_2,...,x_n)=span(x_1)\times span(x_2)\times...\times span(x_n)$ or equivalently $S(x_1,x_2,...,x_n) = S(x_1)S(x_2)...S(x_n)$.

Independence of Sets If a pair of subsets, R_1 and R_2, of X are independent we write $\{R_1|R_2\}$. Note that it is not possible to define the independence of such subsets in terms of the independence of their elements as can be seen from the following. Consider the two sets of Boolean variables $A=\{a_1, a_2\}$ and $C=\{c\}$ with $c = a_1 XOR a_2$ (XOR is exclusive-or), then any pair $\{a_1, a_2\}$, $\{a_1, c\}$ or $\{a_2, c\}$ consists of two independent variables and we can write $\{a_1|C\}$ and $\{a_2|C\}$ but *not* $\{A|C\}$.

Example Two subsets $R=\{r_1, ..., r_n\}$ and $S=\{s_1, ..., s_m\}$ of X in Σ_p are said to be independent iff $p(r_1\&...\&r_n\&s_1\&...\&s_m)=p(r_1\&...\&r_n)p(s_1\&...\&s_m)$.

Example Two subsets $R=\{r_1, ..., r_n\}$ and $S=\{s_1, ..., s_m\}$ of X in Σ_d are said to be independent iff $S = span(r_1, ..., r_n, s_1, ..., s_m) = span(r_1, ..., r_n)\times span(s_1, ..., s_m)$.

Separability A variable y (or set of variables B) is said to separate two variables x and z iff fixing the value of y (or B) makes the values of x and z independent. This is Etter separability [Etter 1996].

Regions and boundaries A set of variables, or boundary, $B\subset X$ is said to separate

two sets of variables $R_i \subset X$ called regions, iff

$R_1 \cap R_2 = 0$, $R_i \cap B = 0$ and
$R_1 \cup R_2 \cup B = X$ (although we might want to relax this condition)

and B separates every pair *of subsets* $r_1 \subset R_1$ and $r_2 \subset R_2$. If B separates A from C then we write {A|B|C}. The extension to $\{R_1, R_2, ..., R_m\}$ is obvious; there may be one or many internal boundaries. (An alternative definition, which Kron would have preferred, is

$R_1 \cap R_2 = B$, (Kron called B the *intersection*) and
$R_1 \cup R_2 = X$ (although we might want to relax this condition)

and B separates every pair *of subsets* $r_1 \subset R_1 - B$ and $r_2 \subset R_2 - B$.)

Etter has pointed out that **every B_i forms a basis for the associated R_i** (and maybe the complement of R_i in the Universe) so that given a network or a field equation we can interpret it as a system for which every boundary is a basis at least for its interior. Two questions then arise both for specific systems and in the general systems approach. 1. Given a set of time series data for the values of the variables in such a system can we reconstruct the associated sets of boundaries and regions from the data alone? 2. Given such a collection of boundaries and regions, under what circumstances can we recover the entire connectivity of the system. That is, under what circumstances can we reconstruct the topology from the data? These ideas will be the basis of a future paper.

Subsystems A subsystem, Σ' of Σ is a subset X' of X and an extension S' on X' such that S'=S on X'. Two important subsets of X will be referred to as the input set U and the output set Y. U is the set of variables that we can control and Y is the set of variables that we can observe. U and Y will often be the same as B, the boundary of the system, otherwise they will usually be larger. The Universe is the distinguished system with no system boundary. A physical system is a region of the Universe, ie a subsystem with a boundary B (or, more generally, an input set U and an output set Y) which separates it from the rest of the Universe. If U=Y=B for all subsystems, the system is said to be local.

Decomposition (Tearing) A system is said to be decomposable into a set of regions $\{R_1, R_2, ..., R_m\}$ iff the values of the variables $\{r_1, r_2, ..., r_{n(i)}\}$, for each R_i, can be calculated from the local boundary conditions $B_i \subset B$ and the local extension. R_i is said to be disconnected, or torn, from the rest of the system. Such a system is a **Physical System**. This is a generalisation of Kron's Method of Tearing (or "diakoptics") [Kron, 1963].

Fundamental Theorem of Disconnection x can be disconnected (torn) from z at y iff x and z are separable at y; ditto for regions. That is if fixing the boundary splits the problem then knowing the boundary splits the problem and vice versa. This is Etter's Fundamental Theorem of Disconnection [Etter, 1996]. The hypothesis that it is possible to tear apart a physical system is known as Kron's Principle [Bowden, 1990] or the "Diakoptical Proposition" [Dweck].

Suppose we have a disconnection of (S,X), ie two regions $R_1=(S,\{X_1,B_1\})$, and $R_2=(S,\{X_2,B_2\})$, such that when they are reconnected by setting the values of $B_1=B_2$, they make the values of $X=X_1\cup X_2$ (as well as the set equality). Then regardless of the values of B_1 and B_2, the values of X_1 and X_2 must be independent, thus fixing $B_1=B_2=B$, a constant, leaves the values of X_1 and X_2 independent, so they are separable by B.

Conversely, suppose X_1 and X_2 are separable by B. Then fixing $B_1=B_2=B$ makes the values of X_1 and X_2 independent. Thus the values of X_1 and X_2 can be calculated from the local boundary conditions only, therefore R_1 can be disconnected from R_2 at B. QED

Etter has pointed out the further theorem that X_i can be torn out at B iff B forms a basis for X_i.

Example The Dean wants to reorganise the Computer Centre into two interacting departments. Most employees lie naturally in one department or the other, but for some the situation is not so clear. How does she simplify her problem? Our theorem says that the problem splits if we know the job specification and conditions of service of the people in the intersection of the two sets. How do we identify these people? They are just the set of people for whom the problem splits if we fix their contracts.

A corollary of Kron's Principle is the

Generalised Huygen's Principle (Jessel's Principle of Secondary Sources) A set of sources (input variables) within {outside} a region of a *Physical* System can be replaced by a new set of sources on the boundary of that region and from the outside {inside} of the region it will not be possible to tell the difference. This is a generalisation of Jessel's formulation of Huygens' Principle [Jessel 1962].

Suppose we have a disconnection or tear, T of Σ = (S,X), into two regions $R_1=(S,\{X_1,B_1\})$, and $R_2=(S,\{X_2,B_2\})$, then fixing $B_1=B_2=B$ makes the values of X_1 and X_2 independent. Thus, assuming that R_1 is the inner region and R_2 the outer

region with no sources, then X_2 can be calculated from B alone and, furthermore, it must be possible to choose a B such that the values of $X=X_1 \cup X_2$, as well as the set equality. Then we refer to B as the "secondary sources" and, if the original sources are denoted $Y=Y_1 \cup Y_2$, we can calculate B from the commutative diagram

$$\begin{array}{ccc} & \overset{S}{\rightarrow} & \\ \{X_1,B_1\}\times\{X_2,B_2\} & & Y_1\times Y_2 \\ T\uparrow & & \uparrow T' = T+[S,T]S^{-1} \\ X & \underset{S}{\rightarrow} & Y \end{array}$$

A (special case of) another corollary is

The Generalised Principle of Optimality For a system in which the variables are indexed by time (or any other independent variable), and on which there is a criterion of optimality, then **subregions of an optimal trajectory are optimal**. That is, when calculating the optimal trajectory of a subregion it can be assumed that the rest of the trajectory is optimal; optimal trajectories need be calculated only in terms of the local boundary conditions. This is a generalisation of Bellman's Principle of Optimality.

Black boxes Mesarovic and Takahara consider the system $S \subset U \times Y$ or $S:U \rightarrow Y$ where U is the input set and Y the output set. For the sort of treatment we are developing we often won't distinguish between U and Y. If a variable has an impressed value it is an input, if it is left floating it is an output. This kind of system has never been treated in the literature before. We will call it a black box. We will generally use the notation above except that we will introduce a new symbol into the input set $X_i=\{\text{span}(x_i), ?\}$ where the ? means that this variable is left floating, ie it's an output. So a typical input to a system would be the string $\{x_1, x_2, ?, x_4, ?, x_6\}$; the corresponding output would be $\{y_1, y_2, y_3, y_4, y_5, y_6\}$ where $y_i=x_i$ iff $x_i \neq ?$.

The advantage of this notation is that it gives more structure to the system. The question of what is an input and what is an output is now part of the structure of the system itself. Kron always claimed that writing down the equations of a system threw away some of the structure of the system. It is this that we wish to avoid. We can permute the example above in order to partition the input and output strings into input and output sets (confusing jargon!) thus $\{x_1,x_2,x_4,x_6,?,?\}$

Topological systems Let $T \subset 2^X$ be the class of subsets of X consisting of all possible disconnected regions (and their boundaries) of X. Then T is a topology on X iff T satisfies the following axioms:

(1) X and {0} belong to T.
(2) The union of any number of sets in T belongs to T.
(3) The intersection of any two sets in T belongs to T.

The members of T are called open sets. (X,T) is a topological space. It would now be natural to look at the Eilenberg-Steenrod axioms for a homology theory.

3. PHYSICS

Using Tom Etter's ideas we have shown how topological structure emerges from (in)dependence in the form of conditional probabilities on information streams in a parallel way to that which can be inferred from dependency properties of a deterministic classical black box, and in a language such that the theorems are general enough to apply to both paradigms, thus establishing a strong connection between two different approaches to "information". We are now going to follow through a train of thought given by Basil Hiley in [Hiley, 1996] but with some different emphases. Hiley starts from the "Algebra of *Process*" (my italics) with Grassman's original but forgotten idea (which led him on to define the Grassman algebra) of a process as an entity $[P_1, P_2]$. Grassman argued that mathematics was about thought, not about material reality. Hiley asks "How [does] one thought become another? Is the new thought *independent* of the old or is there some essential *dependence*?" (my italics). Such a process was generalised by Grassman by considering $[P_1, P_2]$ as a 1-simplex and building simplicial complexes with lines, planes and volumes etc. connecting thoughts and their dependencies. Hiley notes that we can thus construct a complex of "relations of thought, process, activity or movement. The sum total of all such relations constitutes the holomovement." He emphasises *process* (as do Etter and Bastin and Kilmister) but does not mention dependence again. We consider *independence* to be fundamental, but the two approaches have many similarities. Hiley also shows, in a very similar context to [Manthey 1996] in this issue and elsewhere, that Grassman's structure forms a Clifford algebra.

Assume that we have constructed a topological space from dependencies among information streams. Then Etter argues that changes in this space are induced by cutting or tearing the links. Let us start from Kron's idea of tearing and then see if we can reproduce Hiley's results. Into Etter's work we introduce the mathematics of Kron and simplicial complexes; into Hiley's we introduce the concept of tearing or unlinking. We will use the equivalence between Kron's Method of Tearing and Jessel's Principle of Secondary sources to make the jump from a discrete world to a continuous one, rather glossed over in [Hiley 1996].

Kron always used an electrical analogy when building models. An (electrical) network consists of a one dimensional network (and associated 2 dimensional "meshes" and 0-dimensional "nodes") through whose branches flow currents and across whose branches exist potentials (or voltages) responsible for the flows (or vice versa). The flows and potentials are subject to conservation laws (known as Kirchoff's laws) at nodes and around meshes. The branches consist of impedances governing the dynamic relationships (Ohm's law) between the local branch voltages and currents. Typically these impedances are either linear (resistance), first order integrators (capacitance) or first order differentiators (inductance). Also distributed around the network are a set of voltage and current generators or sources responsible for initiating and/or sustaining the dynamic evolution of the system variables, voltages and currents. So a branch will typically consist of an impedance and a voltage source in series. The system equations can be written and solved in two dual forms

$V_1=ZJ^1$ or		$J^1=YV_1$ or	
$E_1+e_1=Z(I^1+i^1)$	(Ohm)	$I^1+i^1=Y(E_1+e_2)$	(1)
$C'e_1=0$	(Kirchoff)	$A'i^1=0$	(2)
$i^1=Ci^2$		$e_1=Ae_0$	(3)
$E_2=C'E_1$		$I^0=A'I^1$	

where i^1 is the vector of b branch currents,
and I^1 of corresponding generators,
i^2 of m mesh currents,
e_1 of b branch voltages,
E_1 of corresponding generators,
e_0 of n node-pair voltages
and Z is the (diagonal) matrix of branch impedances,
Y is its reciprocal of branch admittances
and C is a $b \times m$ matrix of incidence numbers from the set {1,-1,0} depending whether the ith oriented branch is {positively, negatively, not} incident to the jth oriented mesh. In a similar way the A matrix relates mesh and node variables.

The subscripts and superscripts give the dimension of the simplex and define whether it is *covariant* or *contravariant* respectively. The mathematician J. Paul Roth showed how Kron's formulation is isomorphic to a simplicial complex with complex coefficients. In the context in which we are dealing, if covariant vectors represent the dependence between two thoughts then contravariant vectors would represent mappings between those dependencies. Using the elements of homology theory, or algebraic topology, Roth and Sun Ichi Amari showed the general validity of Kron's Method of Tearing described below.

It can easily be seen that A'C=0 and that the following diagram, due to Roth [1955], commutes (gives the same answer whichever way you go round a loop)

$$\begin{array}{ccccccccc} & & & & C & & A' & & \\ 0 & \rightarrow & i^2 & \rightarrow & J^1 & \rightarrow & I^0 & \rightarrow & 0 \\ & & C'ZC\downarrow & & Z\downarrow\uparrow Y & & \uparrow A'YA & & \\ 0 & \leftarrow & E_2 & \leftarrow & V_1 & \leftarrow & e_0 & \leftarrow & 0 \\ & & & & C' & & A & & \end{array}$$

meshes branches nodes

(Actually the commutation is "weak" as $Y \neq C(C'ZC)^{-1}C'$. The mapping Z is Roth's famous "twisted isomorphism". These ideas will be discussed in a forthcoming paper.)

The rows of the diagram are (co)chain complex sequences and the arrows are boundary operators with A'C=0 and C'A=0 (the boundary of a boundary is zero). The upper sequence relates contravariant variables, i.e., currents of successively lower dimension, via boundary operators. The lower sequence relates covariant variables, ie., voltages of increasing dimension via coboundary operators. The vertical mappings (impedances) are isomorphisms. Multiplying (1) through by (2) and substituting (3) gives the branch currents and voltages in terms of the system sources

$$i^1=C(C'ZC)^{-1}C'(E-ZI) \qquad e_1=A(A'YA)^{-1}A'(I-YE) \qquad (4)$$

Application of the method of tearing to an electrical network decomposes (by reordering) the matrix inversions in equations (4) into the form shown below. This is possibly only for sparsely connected networks. For example, for a linear 1-dimensional physical system with four subsystems, the system matrix looks like this (note the new notation!)

$$\begin{vmatrix} Z_1 & & & & C_1 \\ & Z_2 & & & C_2 \\ & & Z_3 & & C_3 \\ & & & Z_4 & C_4 \\ C_1' & C_2' & C_3' & C_4' & Y \end{vmatrix}$$

where the Z_i are the subsystem matrices, Y is the "intersection network" (subsystem boundary) matrix and the C_i are the connection matrices giving the topology of the

interconnections between the subsystems and the intersection. As none of the subsystems have any common boundary the rest of the matrix is null. This can be considered to be no more than a particular permutation, or reordering, of the system matrix induced by the tearing operation.

So we can write now write the system equations

$$Zx + Cy = b \qquad (1)$$

$$C'x + Yy = c$$

where Z is the block diagonal system matrix,
Y is the intersection network system matrix,
C' is the partitioned connection matrix $[C_1' C_2' C_3' C_4']$,
x is the partitioned vector of subsystem solutions,
y is the intersection network solution vector,
b is the partitioned vector of subsystem boundary conditions (voltage and current sources) and
c is the vector of intersection network boundary conditions

The last equation is the system equation of the intersection network which itself is adjacent to all the subsystems, hence the appearance of C_i' in every term. Note that for a two dimensional physical system the subsystem matrices Z_i are themselves of the form

$$\begin{vmatrix} Z & C \\ C' & Y \end{vmatrix}$$

where Z is block diagonal (again we have reused the symbols for different entities in this matrix). This recursive structure extends naturally to multidimensional systems.

Rearranging equations (1) gives the basic equations of "diakoptics" (network tearing) for a linear system

$$y = (Y - C'Z^{-1}C)^{-1}(c - C'Z^{-1}b) \quad (3)$$

which gives the intersection vector and

$$x = Z^{-1}(b - Cy) \qquad (4)$$

which gives the subsystem solutions. Note that the equation for y involves only the inversion of a matrix of the order of Y, which can usually be made quite small, and

the inversion of Z, which is block diagonal and thus involves only the inversion of (all) the Z_i. Equation (3) can be thought of as a projection of the boundary conditions onto the intersection network. The equation for the *n* subsystem solutions naturally splits into *n* subsystem equations

$$x_i = Z_i^{-1}(b_i - C_i y) \qquad (5)$$

which as stated depend only on the intersection vector and the local boundary conditions. The intersection vector is given by

$$y = (Y - \Sigma_{j=1}^{n} C_j' Z_j^{-1} C_j)^{-1}(c - \Sigma_{k=1}^{n} C_k' Z_k^{-1} b_k) \quad (6)$$

Note that although all of the vectors and matrices given above are assumed to be real, the theory works over any suitable field e.g., complex numbers or polynomials. In particular solution of the equations where the field is polynomials in the Laplace transform s, allows a dynamic analysis, a wave theory rather than a field theory.

In [Bowden, 1990] we showed the equivalence of Kron's Method of Tearing and Jessel's Principle of Secondary Sources. The starting point for Jessel was Huygens' Principle which, after researching Huygens' original work (in the original old Dutch), Jessel reformulated as follows, "The perturbation that goes out (or in) through a closed surface S that contains (excludes) a wave or field source is identical to the perturbation that can be obtained by cutting off the source and replacing it by appropriate sources distributed on the surface S". Consider the wave equation

$$OP\ F = S_{or}$$

where OP is a differential operator such as the Laplacian, but possibly with a time derivative, F is a field distribution in space and S_{or} is a set of (original) sources, (eg., charges). It should be noted that this is the continuous (distributed) version of the basic system equation that diakoptics tries to solve.

Kron's aim was a fast solution of the equation

$$F = OP^{-1} S_{or}$$

Jessel was more interested in changing the form of the field F by modifying the sources. He considered the identity

$$OP(sF) = sOPF + [OP,s]F$$

where the last term is the Lie bracket (OPs-sOP)F. The scalar space function s is a field modifier. For instance it can be used to define a Huygens' surface, S of secondary sources by setting its value to 1 outside the surface and 0 inside the surface. Its value on the surface is undefined but in general may be complex (ie., complicated). As s is a constant scalar everywhere else the Lie bracket vanishes everywhere but on the surface. The first term $sOPF = sS_{or}$ simply defines any sources that are not inside the surface. Thus the Lie bracket gives the secondary sources required to reproduce that element of the field not produced by any sources external to the Huygens' surface and we write

$$S^s_{or} = [OP,s]F$$

and $OP(sF) = sS_{or} + S^s_{or}$.

If the sources are all inside the surface then the first term vanishes because either s or S_{or} is always zero. This is Huygens' Principle.

Jessel and Resconi claim that a change in the field F due to the modifier s induces a corresponding "anticausal" change s' in the sources. They show this on the commutative diagram

$$\begin{array}{ccc} & OP & \\ sF & \rightarrow & sS_{or}+S^s_{or} \\ s\uparrow & & \uparrow s'=s+[OP,s]OP^{-1} \\ F & \rightarrow & S_{or} \\ & OP & \end{array}$$

and $s' = OPsOP^{-1} = s+[OP,s]OP^{-1}$.

They call a diagram of this form an ELS or Elementary Logical System. s and s' are said to be similar when there exists a 1:1 relation OP such that the diagram commutes. For instance if OP is a matrix it must not be singular. The domain and codomain of s (and s') must be of the same type. Bertrand Russell [1919] saw similarity as a very basic concept and noted that "when two relations are similar, they share all properties that do not depend on the actual terms in their fields... Even statements involving the actual terms of the field of a relation, though they may not be true as they stand when applied to a similar relation, will always be capable of translation into statements that are analogous." Thus similarity is not unlike isomorphism which in turn is a special case of exactness.

Example Consider the electrostatic equation

$$\mathrm{div}(D) = \rho$$

where the charges, ρ are contained within a surface defined by s then the secondary sources to reproduce the field outside the surface are given by

$$\begin{aligned} S^s_{or} &= [\mathrm{div},s]D \\ &= \mathrm{div}(sD) - s\,\mathrm{div}D \\ &= D.\mathrm{div}(s) \end{aligned}$$

and $sD = \mathrm{div}^{-1}(\mathrm{div}^{-1}\rho).\mathrm{div}(s).$

or

$$\begin{array}{ccc} & \overset{\mathrm{div}}{} & \\ sD & \rightarrow & D.\mathrm{div}(s) \\ s\uparrow & & \uparrow \mathrm{div}^{-1}(.).\mathrm{div}(s) \\ D & \rightarrow & \rho \\ & \underset{\mathrm{div}}{} & \end{array}$$

which gives the field in the region outside the surface in terms of secondary sources on the surface which in their turn are calculated from the original sources inside the surface.

Note how similar this is to Kron's procedure. We claim in [Bowden 1990] that the intersection network of Kron's Method of Tearing is in fact a Huygens' surface. The ELS for Kron's Method looks like this

$$\begin{array}{ccc} & \overset{Z}{} & \\ x & \rightarrow & b\text{-}Cy \\ [I\ 0]\uparrow & & \uparrow [I\ 0]\text{-}C(C'Z^{-1}C\text{-}Y)^{-1}[C'Z^{-1}\text{-}I] \\ |x| & \rightarrow & |b| \\ |y| & & |c| \end{array}$$

$$\begin{vmatrix} Z & C \\ C' & Y \end{vmatrix}$$

If it is required only to solve the *i*th system then s will take the special form

$$[0 \ldots 0\ I\ 0 \ldots 0]$$

where the I is in the *i*th position. The intersection vector is a holographic field containing all the information from the boundary conditions. The holographic form is easier to visualise if a dynamic system is considered.

Sometimes we may be interested in applying a succession of tearing operations. For instance, the "staircase method" tears out a subsystem at a time, (similarly hierarchical tearing [Bowden, 1994] tears the system into subsystems, then subsubsystems, etc.) Considering diakoptics to be a reordering of the system matrix induced by the tearing operation, where P is a reordering operator, usually a permutation matrix, leads us to see more clearly the relationship to Jessel's Principle of Secondary sources. Such a succession of transformations was referred to by David Bohm as an "ordering or enfolding", and by Jessel and Resconi as an Logical System and was represented by them on a commutative diagram. A Logical System is a sequence of ELS as shown below

$$\begin{array}{ccccccc} & OP_1 & & OP_2 & & & \\ F_0 & \rightarrow & F_1 & \rightarrow & F_2 & \ldots & F_n \\ s_0\uparrow & & s_1\uparrow & & s_2\uparrow & & s_n\uparrow \\ F_0 & \rightarrow & F_1 & \rightarrow & F_2 & \ldots & F_n \\ & OP_1 & & OP_2 & & & \end{array}$$

Thus s_0 is similar to s_1 and hence to s_2 and so on. In the example above due to Jessel s is a scalar with value either 0 or 1. If these are thought of as probabilities then s can be considered to be a (probability) density matrix. For convenience we will now switch to the notation in [Etter, 1996]. Assuming $OP_i = T$ is constant and writing $s_0=S$, $s_1 = S'$, $s_2 = S''$, etc. then

$$\begin{array}{ccccc} & T & & T & \\ F_0 & \rightarrow & F_1 & \rightarrow & F_2 \ldots \\ S\uparrow & & S'\uparrow & & S''\uparrow \\ F_0 & \rightarrow & F_1 & \rightarrow & F_2 \ldots \\ & T & & T & \end{array}$$

commutes and the diagram represents a sequence of "anticausal" changes and we have

$$S' = TST^{-1} = S + [T,S]T^{-1}$$

etc. where, as mentioned above, S can be thought of as a density operator. Now assuming T=exp(iHt/h) where t is an "enfolding parameter" representing successive steps, h is a scaling constant and (remember T can be complex) H is an operator then for small t we can write

$$S' = (1 - iHt/h)\, S\, (1 + iHt/h)$$

so that

$$ih(S' - S)/t = HS - SH = [H, S].$$

So that in the limit as $t \rightarrow 0$ we have

$$ihdS/dt = [H, S]$$

which Hiley notes has the same form as the Heisenberg equation of motion. As S is a density matrix we can assume that it is factorisable in the form $S = \Psi\Psi^*$ and then we have

$$ihd\Psi\Psi^*/dt = ih(d\Psi/dt)\Psi^* + ih\Psi d\Psi^*/dt = (H\Psi)\Psi^* + \Psi(\Psi^* H)$$

which splits naturally into two equivalent formulae thus

$$ihd\Psi/dt = H\Psi \quad \text{and} \quad ihd\Psi^*/dt = -\Psi^* H$$

(postmultiply the first equation by Ψ^* and premultiply the second by Ψ and add. This ends Hiley's argument. Putting the Hamiltonian $H=V-\nabla^2h^2/2m$ gives the usual form of the Schrodinger equation, $ihd\Psi/dt = (V-\nabla^2h^2/2m)\Psi$.)

In detail, the commutative diagram for this Logical System is

$$\begin{array}{ccccc} & & e^{iHt/h} & & e^{iHt/h} \\ & F_1 & \rightarrow & F_2 & \rightarrow & F_3 \ldots \\ \Psi\Psi^* \uparrow & & \Psi'\Psi'^* \uparrow & & \Psi''\Psi''^* \uparrow & \\ & F_1 & \underset{e^{iHt/h}}{\rightarrow} & F_2 & \underset{e^{iHt/h}}{\rightarrow} & F_3 \ldots \end{array}$$

Finally we look briefly at Kauffman and Noyes derivation of Maxwell's equations. In quantum mechanics it is usual to define position and momentum operators R_i and P_j

such that for any location and momentum vectors **r** and **p** of a particle

$$\langle \mathbf{r} \mid R_i \mid \Psi \rangle = r_i \langle \mathbf{r} \mid \Psi \rangle \quad \text{and} \quad \langle \mathbf{p} \mid P_i \mid \Psi \rangle = p_i \langle \mathbf{p} \mid \Psi \rangle \quad i=1,2,3$$

where R_1, R_2, R_3 and P_1, P_2, P_3 designate respectively position and momentum operators X, Y, Z and P_x, P_y, P_z respectively. Then it is easy to show (eg., [Cohen-Tannoudji, 1977] p.150) the canonical commutation relations

$$[R_i, R_j] = 0,$$
$$[P_i, P_j] = 0,$$
$$\text{and } [R_i, P_j] = ih\delta_{ij}$$

for i,j = 1, 2, 3. In particular $R_iP_i - P_iR_i = ih$; all other variables commute. (For example in one dimension R and P are defined by the equations

$$R\Psi = x\Psi \quad \text{and} \quad P\Psi = (h/i)\partial\Psi/\partial t.$$

Thus

$$(RP-PR)\Psi = x(h/i)\partial\Psi/\partial x - (h/i)(\partial/\partial x)(x\Psi) = (ih)\Psi.$$

Hence RP - PR = ih.)

Following Dyson [1990], after an idea by Feynman, Kauffman and Noyes [1995] consider a point in a discrete space whose position and velocity are governed by the analogous equations

$$[X_i, X_j] = 0,$$
$$\text{and } [X_i, X_j^t] = \kappa\delta_{ij} \text{ for } i,j = 1, 2, 3$$

where $X^t = dX/dt$ is defined as J(X'-X) where the operator J, whose sole purpose is to keep track of the time shifting, is defined thus

$$J' = J,$$
$$AJ = JA' \text{ for all } A$$

and XdX/dt becomes X'(X'-X) as we have to take a time step to make the differential. The commutator $[X, X^t]$ is clearly nonzero. They then develop the Discrete Ordered Calculus (DOC) and show that

$$d(XY)/dt = (dX/dt)Y + XdY/dt$$

as is appropriate. They define

$$E_j = d^2X_j/dt^2 - \varepsilon_{ijl}(dX_k/dt)H_l$$

(by analogy with $E = F - v \times H$) where

$$H_l = (1/2\kappa)\,\varepsilon_{ijl}[dX_j/dt,\, dX_k/dt]$$

and prove after a lengthy calculation that

$$\text{div}\, H = 0 \quad \text{and} \quad \partial H/\partial t + \nabla \times E = 0.$$

Following Dyson they define

$$\text{div}\, E = \rho \quad \text{and} \quad \partial E/\partial t - \nabla \times H = j$$

to complete Maxwell's equations. They note that from (1) and $\partial X_i/\partial X_j = \delta_{ij}$ that we can write

$$\partial X_i/\partial X_j = [X_i, X_j^t]/\kappa$$

which allows us to define

$$\partial G/\partial X_j = [G, X_j^t]/\kappa$$

To simplify and clarify their derivation, introduce div, grad and curl with noncommutative coefficients

$$\nabla G = [\ldots \partial G/\partial X_i \ldots] = [\ \ldots\ [G, X_i^t]/\kappa \ \ldots\ ] = (GX^t - X^tG)/\kappa \qquad (G \text{ scalar})$$

$$\nabla.G = \partial G_i/\partial X_i = [G_i, X_i^t]/\kappa = (G.X^t - X^t.G)/\kappa$$

$$\nabla \times G = [\ \ldots\ \varepsilon_{ijk}\partial G_i/\partial X_j \ \ldots\ ] = [\ \ldots\ \varepsilon_{ijk}[G_i, X_j^t]/\kappa \ \ldots\ ] = -(G \times X^t + X^t \times G)/\kappa$$

with $A \times B = \varepsilon_{ijk}A_iB_j$. Note of course that $A \times B \neq -B \times A$ and thus that $A \times A \neq 0$! Thus we can write Kauffman and Noyes' definitions E and H,

$$E = X^{tt} + X^t \times H \text{ and } H = \nabla \times X^t/2\kappa$$

and $\partial H/\partial t = dH/dt + (\partial X_j/\partial t)\,\partial H/\partial X_j$

$$= d(X^t \times X^t)/\kappa dt - (X^t.\nabla)H$$

$$= (X^{tt} \times X^{t} + X^{t} \times X^{tt})/\kappa - (X^{t}.\nabla)H$$

$$= -\nabla \times X^{tt} - (X^{t}.\nabla)H$$

$$= -\nabla \times (E + X^{t} \times H) - (X^{t}.\nabla)H$$

$$= -\nabla \times E - X^{t}(\nabla.H) + H(\nabla.X^{t}) - (H.\nabla)X^{t} + (X^{t}.\nabla)H - (X^{t}.\nabla)H$$

$$= -\nabla \times E$$

where we have omitted the detailed calculations. These will be discussed in a forthcoming paper. Mike Peskin has pointed out that if the J operator is considered as a time step operator of the form exp(iHt) then from AJ=JA' we recover ELS like structures as shown above.

The structure of Maxwell's equations can be shown on an extended Roth diagram [Roth, 1955 or Bowden, 1990] or Tonti diagram [Tonti, 1976 or Bossavit, 1992] of the form

$$\begin{array}{ccccccc} & \text{grad} & & \text{div} & & \text{curl} & \\ \psi & \rightarrow & M+H & \rightarrow & J+sD & \rightarrow & . \\ \downarrow & & \mu s \downarrow\uparrow & & \downarrow\uparrow \sigma+\varepsilon s & & \uparrow \\ . & \leftarrow & s(\mu M+B) & \leftarrow & E+sA & \leftarrow & -\phi \\ & \text{div} & & \text{curl} & & \text{grad} & \end{array}$$

on which the internal structure of the twisted isomorphism can be shown as in [Bowden, 1990]. This is a special case of the general form

$$\begin{array}{ccccccccccc} & & \partial & & \partial & & \partial & & \partial & & \\ \ldots & 4 & \rightarrow & 3 & \rightarrow & 2 & \rightarrow & 1 & \rightarrow & 0 & \text{(dimension)} \\ & \text{OP}\downarrow\uparrow & & \text{OP}\downarrow\uparrow & & \text{OP}\downarrow\uparrow & & \text{OP}\downarrow\uparrow & & \text{OP}\downarrow\uparrow & \\ \ldots & 4 & \leftarrow & 3 & \leftarrow & 2 & \leftarrow & 1 & \leftarrow & 0 & \text{(dimension)} \\ & & \delta & & \delta & & \delta & & \delta & & \end{array}$$

showing the dimensional structure of the Logical Systems described above. This structure is becoming more important in practical electrical field analysis [Bossavit, 1992 and 1996]. Indeed Bossavit asks "When can we discretise a Tonti diagram?"

[Bossavit 1992].

4. COMPUTERS AND CONCLUSIONS

The title of this paper is that given by Jessel to the book he and I were to write together, before his untimely death in 1992. It was hoped to include some application of our general theory to computing in this paper, but this is still proving notoriously elusive. It is hoped that we will be forgiven for the title out of respect for Jessel. The problem essentially involves applying the theory over the group $Z_2=\{0,1\}$, rather than the reals or the integers. We have given some examples in the text, in particular the one relating to exclusive OR, showing that independence does not just depend on pairwise application, but in general we have found it hard to get useful results. It is believed that this new theory should help. Dubois [1996] describes the problem in his paper in this issue. It may be that Pawel Siwak's [1996] work also points the way. Recent work with Mike Manthey and Clive Kilmister on Manthey's Topsy, which is naturally defined over $Z_3=\{-1,0,1\}$ is looking more promising. It is hoped to publish some results in the near future.

The frequency of new publications on derivations of physical laws is remarkable. One physicist remarked that "Derivations of Schrodinger's equation are ten a penny!" This in no way detracts from the value of Etter's work which can be thought of as having an underlying (statistical) ontological picture contributing to the physical understanding of our world (such as why we add amplitudes instead of probabilities in Quantum Theory). The degree to which Etter's picture corresponds to that which we have developed in the past, relating the work of Jessel and Kron is remarkable. This has allowed the beginnings of the creation of a "covering theory" which is general enough to allow description of both the classical and Quantum worlds. This has been one of the major problems with the orthodox theory and is known as the Measurement Paradox. The value of these ideas for General Systems Theory itself should also not be underestimated. A general algorithm for recovering topological structure from unstructured data is implied. We hope to develop all these ideas much further in future publications.

ACKNOWLEDGEMENTS

Thanks are due to John Amson, Ted Bastin, Tom Etter, Basil Hiley, Clive Kilmister and, of course, to the late Maurice Jessel.

REFERENCES

Bastin, T. and C. W. Kilmister [1995], *Combinatorial Physics*. World Scientific, London.

Bohm, D. and B. J. Hiley [1993], *The Undivided Universe*. Routledge, London.

Bossavit A. [1992], "A New Viewpoint on Mixed Elements", *Meccanica*, 27, pp. 3-11.

Bossavit A. [1996], "Edge Elements for Magnetostatics", *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, **9**, pp. 19-34.

Bowden K. [1990], "On General Physical Systems Theories", *International Journal of General Systems*, **18**, pp. 61-79.

Bowden K. [1994], "Hierarchical Tearing: An Efficient Holographic Algorithm for System Decomposition", *International Journal of General Systems*, **23**(1), pp. 23-37.

Bowden K. [1996], "What is to be Done", *Proceedings of the 17th International ANPA Conference*, University of Cambridge, September 1995.

Cohen-Tannoudji C. et al [1977], *Quantum Mechanics Vol. 1*. Wiley, International and Hermann, Paris.

Dweck E., Editorials, *Matrix and Tensor Quarterly*.

Etter T. [1996], "Process, System, Causality and Quantum Mechanics", *International Journal of General Systems, this issue.*

Hiley B. J. [1996], "The Algebra of Process", *Conciousness at the Crossroads, Proceedings of the Tempus Project 'Phenomenology and Cognitive Science', Maribor, Slovenia, 23-7 August 1994*, pp. 52-67.

Jessel M. [1962], *Contribution aux Theories du Principe Huygens et de la Diffraction*. Thesis for Doctorate of Physical Sciences, University of Paris.

Kauffman L. H. and H. Pierre Noyes [1996], "Discrete Physics and the Derivation of Electromagnetism from the Formalism of Quantum Mechanics", *Proceedings of the Royal Society of London*, 452, pp 81-95.

Kron G. [1963], *Diakoptics: The Piecewise Solution of Large Scale Systems*. MacDonald, London.

Mesarovic M. D. and Yasuhiko Takahara [1975], *General Systems Theory:*

Mathematical Foundations. Academic Press, New York.

Parker-Rhodes F. [198?], *The Inevitable Universe*, unpublished.

Roth J. P. [1955], "The Validity of Kron's Method of Tearing", *Proceedings of the National Academy of Sciences*, **41**.

Russel B. [1919], *Introduction to Mathematical Philosophy*. Allen and Unwin, London.

Sudkamp T. [1992], "The Semantics of Plausibility and Possibility." *International Journal of General Systems*, **21**, pp. 273-289.

Tonti E. [1976], "The Reason for Analogies between Physical Theories", *Appl. Math. Modelling*, **1**, pp. 37-50.

HOW TO DO TRIGONOMETRY ON THE COMBINATORIAL HIERARCHY

JOHN AMSON*

Prepared for ANPA, 1999 March 12

1. Introduction.

Of the many ways in which the ideas of a Cosine and Sine of an Angle can be introduced, at a fairly elementary stage, the one which appears to be most intuitive for the present article starts out from the familiar vector calculus definitions : given two vectors $\mathbf{x}$ and $\mathbf{y}$ inclined to each other at the angle θ, we have

$$\mathbf{x} \bullet \mathbf{y} = |\mathbf{x}|\,|\mathbf{y}|\,\mathtt{COS}\,\theta$$
$$\mathbf{x} \wedge \mathbf{y} = |\mathbf{x}|\,|\mathbf{y}|\,\mathtt{SIN}\,\theta\,\mathbf{u}$$

where $\bullet$ resp. $\wedge$ denote the scalar (or *inner*) product (a number) and the vector (or *outer*) product (a vector) respectively, $\mathbf{u}$ is a unit vector in the same direction as $\mathbf{x} \wedge \mathbf{y}$, and $|\ |$ denotes the length (or *norm*) of a vector. Recall that $|\mathbf{x}|^2 = \mathbf{x} \bullet \mathbf{x}$ identically, and that $\mathtt{COS}\,\theta = 1$ and $\mathtt{SIN}\,\theta = 0$ if and only if the vectors $\mathbf{x}$ and $\mathbf{y}$ are *coincident*. Recall also that $(\mathtt{COS}\,\theta)^2 + (\mathtt{SIN}\,\theta)^2 = 1$ identically so that if neither $\mathbf{x}$ nor $\mathbf{y}$ are $\mathbf{o}$, then

$$(\mathtt{SIN}\,\theta)^2 = 1 - \frac{(\mathbf{x} \bullet \mathbf{y})^2}{(|\mathbf{x}|\,|\mathbf{y}|)^2} = \frac{(|\mathbf{x}|\,|\mathbf{y}|)^2 - (\mathbf{x} \bullet \mathbf{y})^2}{(|\mathbf{x}|\,|\mathbf{y}|)^2}$$

I shall now show, by exploiting related concepts in the context of the Combinatorial Hierarchy, how we may introduce two functions, COSB and SINB (for "Boolean-Cosine" and "Boolean-Sine"). Though not the same as Cosine and Sine they have many identical or similar properties, and may be used to simulate the trigonometric/geometric properties of triangles, parallelograms, and so forth.

The definitions and results are novel and general and have wider implications for any finite Vector Space over the two-element field $(0, 1)$ in which the notions of a Boolean Ring are present, and may readily be extended to arbitrary infinite systems.

* The Old Manse, 5 Shore, Anstruther, Fife KY10 3DY Scotland
Tel & Fax : 01-333-310-087 e-mail john.amson@which.net

2. Inner-Products.

The Combinatorial Hierarchy (**CH**) is usually described as an organised collection of binary vectors. Its organisation is usually presented as an inter-related set of four vector spaces (*levels*) of increasing dimension — 2, $4 = 2^2$, $16 = 4^2$, $256 = 16^2$ — the elements of each of the 2nd, 3rd and 4th levels playing the rôle of linear operators or mappings acting on the previous level. The vector spaces are over the field of two elements $(0, 1)$. The 'vectors' are ordered strings of elements from $(0, 1)$, and are usually referred to in **CH**-speak as 'bit-strings'. Vectors are added component-wise according to the logical XOR rule denoted by $\oplus$ which acts in the field $(0, 1)$ by the usual 'discrimination' law $0 \oplus 0 = 1 \oplus 1 = 0$, $0 \oplus 1 = 1 \oplus 0 = 1$. Linear operators at one level act as matrices multiplying vectors at the previous level. They do this by multiplying a vector at the lower level by each of the column-vectors of the matrix in turn, to produce the individual components of the product. This vector-vector multiplication is in fact component-wise multiplication according to the logical AND rule denoted throughout this paper by $\otimes$ which acts in the field $(0, 1)$ by the usual binary law $0 \otimes 0 = 0 \otimes 1 = 1 \otimes 0 = 0$, $1 \otimes 1 = 1$ — followed immediately by discriminately adding up the products of each pair of components, for instance:-

$$\begin{aligned} a_1 \otimes \mathbf{x} &= (a_{11}, a_{12}, a_{13}, a_{14}) \otimes (x_1, x_2, x_3, x_4) \\ &= (a_{11} \otimes x_1) \oplus (a_{12} \otimes x_2) \oplus (a_{13} \otimes x_3) \oplus (a_{14} \otimes x_4) \end{aligned}$$

$$\begin{aligned} \textit{e.g.} \qquad (1,0,1,1) \otimes (0,1,1,1) &= (1 \otimes 0) \oplus (0 \otimes 1) \oplus (1 \otimes 1) \oplus (1 \otimes 1) \\ &= 0 \oplus 0 \oplus 1 \oplus 1 = 0 \end{aligned}$$

The two special vectors which have all their components 0 or 1 respectively are known as the *null* and *anti-null* vectors, respectively, and denoted here by $\mathbf{o}$, $\mathbf{1}$. Use is also made of the 'dualising' operation which turns one vector (bit-string) into its 'dual' by discriminating the first vector with the anti-null vector, *e.g.* $(1,0,1,1) \oplus (1,1,1,1) = (0,1,0,0)$, an operation which at component level is simply the rule $x' = \mathbf{x} \oplus \mathbf{1}$.

It is also useful in calculations to remember (i) that multiplication of elements from $\{0, 1\}$ is 'idempotent', *i.e.* $x^2 = x$, and (ii) that the $\oplus$ action can be replaced by integer arithmetic : $x \oplus y = (x - y)^2$ for all $x, y \in \{0, 1\}$, and that $(x - y)^2 = x^2 + y^2 - 2.x.y = x + y - 2.x.y$ in $\mathbf{Z}$ by idempotency.

The rule for vector-vector multiplication illustrated above is in fact a rule for forming an "inner-product" of two bit-strings. However, there happens to be (at least) two crucially different ways in which we can form an "inner-product", all depending on how we 'add-up' the component-products. Based on the previous example, consider:-

$$\begin{aligned} [\mathbf{x}, \mathbf{y}] &= [(1,0,1,1), (0,1,1,1)] = 0 \oplus 0 \oplus 1 \oplus 1 = 0 \\ \langle \mathbf{x}, \mathbf{y} \rangle &= \langle (1,0,1,1), (0,1,1,1) \rangle = 0 + 0 + 1 + 1 = 2 \end{aligned}$$

. . . in the [] case we are using discrimination to add the components, but in the $\langle\rangle$ case we are using 'ordinary' integer addition.

In the latter case we can also write

$$\boxed{\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i \cdot y_i}$$

where the symbols '$\sum$' and '$\cdot$' means addition and multiplication in the ring of integers $\mathbf{Z}$ and the components x_i, y_i are restricted to 0, 1.

It turns out that neither of these two rules create a proper "inner-product". But the second one comes closer than the first, and so we give it preference. It is :-

symmetric	$: \langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
totally positive	$: \langle \mathbf{x}, \mathbf{y} \rangle \geq 0$ (as a function)
positive	$: \langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ (as an operator)
definite	$: \langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{o}$
sub-bilinear	$: \langle \mathbf{x} \oplus \mathbf{y}, \mathbf{z} \rangle \leq \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$, and
	$\langle \mathbf{x}, \mathbf{y} \oplus \mathbf{z} \rangle \leq \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$

(for all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in question).

By positivity and idempotency we have $0 \leq \langle \mathbf{x}, \mathbf{x} \rangle = \sum_i x_i^2 = \sum_i x_i$ so let us write $\|\mathbf{x}\|$ for the last summation. Then $\|\mathbf{x}\| \geq 0$ and is 0 iff $\mathbf{x} = \mathbf{o}$, and we also have $\|\lambda \mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|$ for all $\mathbf{x}$ and either of the two scalars λ from the two-field $\{0, 1\}$. We shall soon see that the Triangle Inequality $\|\mathbf{x} \oplus \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ holds, so that $\|\mathbf{x}\|$ is a genuine *norm* in the **CH**. In fact, the value of $\|\mathbf{x}\|$ is just the integer sum of the number of elements x_i in $\mathbf{x}$ which have the value 1 — in other words it is (just) the well-known *Hamming Norm*. Anyone familiar with inner-product space theory will know that the norm which is derived from an inner products has to involve the square-root of the inner-product (" $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}} = \left(\sum_i x_i^2 \right)^{\frac{1}{2}}$ ") but the peculiar rôle of idempotency renders the square-root unnecessary here!

Now a proper inner-product would be bilinear rather than sub-bilinear, *i.e.* would have equality signs instead of the inequality signs in the 'sub-linear' statement above. Here is half of the proof of the sub-bilinearity (the other half is similar) :-

$$\begin{aligned} \langle \mathbf{x} \oplus \mathbf{y}, \mathbf{z} \rangle &= \sum_i (x_i - y_i)^2 \cdot z_i \\ &= \sum_i x_i \cdot z_i + y_i \cdot z_i - 2 \cdot x_i \cdot y_i \cdot z_i \\ &= \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle - 2 \cdot \|\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}\| \\ &\leq \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle \end{aligned}$$

To show that strict inequailty can hold we only need to find three bit-strings $\mathbf{x}, \mathbf{y}, \mathbf{z}$ with $\|\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}\| > 0$; for example, with bit-strings of length 2, $\mathbf{x} = (0,1) = \mathbf{y}$, $\mathbf{z} = (1,1)$ have $\|\mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}\| = \|(0,1)\| = 1$.

Because our $\langle \mathbf{x}, \mathbf{y} \rangle$ is almost an inner-product we shall call it a

PSEUDO INNER-PRODUCT — "PIP".

It is easy (employing the classical Cauchy Inequality/Equality for Natural Numbers, and idempotency) to verify that PIP satisfies

the Pseudo Cauchy-Schwarz Inequality : $\langle \mathbf{x}, \mathbf{y} \rangle \leq \langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \|\mathbf{x}\| . \|\mathbf{y}\|$.

the Pseudo Cauchy Equality : $\langle \mathbf{x}, \mathbf{y} \rangle^2 = \|\mathbf{x}\| . \|\mathbf{y}\|$ iff $\mathbf{x} = \mathbf{y}$, or $\mathbf{x} = \mathbf{o}$, or $\mathbf{y} = \mathbf{o}$.

The PIP can also be given in terms of the norm of the product (logical AND) of two vectors :-

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum x_i . y_i = \|\mathbf{x} \otimes \mathbf{y}\| .$$

Making use of the same sort of arithmetic as used to prove sub-linearity we can not only show that the Triangle Inequality holds but can express it in the even stronger form of the

Triangle Equation : $\|\mathbf{x} \oplus \mathbf{y}\| + 2.\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| + \|\mathbf{y}\|$,

or, equivalently, $\|\mathbf{x} \oplus \mathbf{y}\| + 2.\|\mathbf{x} \otimes \mathbf{y}\| = \|\mathbf{x}\| + \|\mathbf{y}\|$.

or, in other words, $\|\mathbf{x}\,\mathtt{XOR}\,\mathbf{y}\| + 2.\|\mathbf{x}\,\mathtt{AND}\,\mathbf{y}\| = \|\mathbf{x}\| + \|\mathbf{y}\|$.

(which reveals by just how much (namely, $2.\langle \mathbf{x}, \mathbf{y} \rangle$) the norm of the sum falls below the sum of the norms) and hence we can immediately derive two further properties :-

(1) the Polarization Law : $\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2} (\|\mathbf{x}\| + \|\mathbf{y}\| - \|\mathbf{x} \oplus \mathbf{y}\|)$,

which shows us how we can reconstruct our PIP from the norm; and

(2) the (Hamming) norm is Archimedean.

To be Archimedean, a norm must fail to satisfy identically the 'Non-Archimedean' property

$$\|\mathbf{x} \oplus \mathbf{y}\| \leq \max (\|\mathbf{x}\| , \|\mathbf{y}\|) .$$

Using the Triangle Equation, that means it has to fail the condition

$$\|\mathbf{x}\| + \|\mathbf{y}\| - 2.\langle \mathbf{x}, \mathbf{y} \rangle \leq \max (\|\mathbf{x}\| , \|\mathbf{y}\|) ;$$

but this is easy to fault — take any two non-$\mathbf{o}$ vectors $\mathbf{x}$ and $\mathbf{y}$ with $\langle \mathbf{x}, \mathbf{y} \rangle = 0$; then $\|\mathbf{x}\| \neq 0$ and $\|\mathbf{y}\| \neq 0$ and we plainly have $\|\mathbf{x}\| + \|\mathbf{y}\| > \max (\|\mathbf{x}\| , \|\mathbf{y}\|)$.

The Triangle Equation will also soon be used to let the ghost of Pythagoras into the Combinatorial Hierarchy.

Just as in Euclidean space with its proper inner-product, with PIP we can introduce 'orthogonality' into the CH :- two bit-strings $\mathbf{x}, \mathbf{y}$ are *orthogonal* (write $\mathbf{x} \perp \mathbf{y}$) iff $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. Of course, this is the same as asking for $\mathbf{x} \otimes \mathbf{y} = \mathbf{o}$. There are immediate examples : $\mathbf{x} \perp \mathbf{x}'$, $\mathbf{x}' \perp \mathbf{x}$, $\mathbf{x} \perp \mathbf{o}$, $\mathbf{o} \perp \mathbf{x}$, $\mathbf{o} \perp \mathbf{o}$. Moreover we now have the valuable

Pseudo Pythagoras Law : $\|\mathbf{x} \oplus \mathbf{y}\| = \|\mathbf{x}\| + \|\mathbf{y}\|$ iff $\mathbf{x} \perp \mathbf{y}$.

Of course, the original Euclidean version of the Pythagoras Law involved the squares of lengths :-

$$\text{“ } \|\mathbf{x} \oplus \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \text{ iff } \mathbf{x} \perp \mathbf{y}\text{”.}$$

From the Pseudo Pythagoras Law we see at once that for equality to hold in the norm's Triangle Inequality $\|\mathbf{x} \oplus \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ in the CH it is not only sufficient but necessary that the bit-strings $\mathbf{x}$ and $\mathbf{y}$ be orthogonal. This goes directly against the intuitive idea drawn from our experience of ordinary Euclidean space. In the latter there is equality in the Triangle Inequality whenever two vectors and their sum-vector are all co-linear and have the same direction — the "triangle" formed by them having degenerated into a single straight line.

Before leaving this topic of orthogonality it is worth commenting that one can derive 'pseudo' versions of all the essential ideas about Orthonormal Systems, Fourier Expansions, Fourier Coefficients, Bessel's Equality, Bessel's Inequality, Bessel Residuals, and Bessel-Fourier Approximation, *etc.*. But this theory is significantly less deep than in the Euclidean or Hilbert Space situation, primarily because any Orthonormal System in the CH is necessarily a subset of the set of relevant unit bit-strings (*i.e.* bit-strings with only a single element '1' and all others '0').

3. Outer-Products.

We have seen that a (pseudo) inner-product is obtained by 'multiplying' two vectors (bit-strings) in the CH in a particular way to produce an non-negative integer; in effect we go from two vectors to an object of lower order, a scalar. But it is also possible to head in the opposite direction, and go from two vectors to an object of higher order, namely a matrix, to create an *outer product.* For this we use the standard 'Kronecker' product. For example :-

$$\begin{aligned} \text{with} \quad & \mathbf{x} = (x_1, x_2, x_3) = (0,1,1) \\ \text{and} \quad & \mathbf{y} = (y_1, y_2, y_3) = (1,1,0) \end{aligned}$$

$$\text{put} \quad \mathbf{x} \bigotimes \mathbf{y} = \left(x_i \cdot y_j \right)_{i,j \in \{1,2,3\}} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

$$\text{and} \quad \mathbf{y} \bigotimes \mathbf{x} = \left(y_i \cdot x_j \right)_{i,j \in \{1,2,3\}} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

We see at once that this Outer Product $\otimes$ is not symmetric (*i.e.* commutative). This is valuable in itself, but we also need a symmetrized version.

This is readily defined by creating our

SYMMETRIZED OUTER-PRODUCT — "SOP" : $\mathbf{x} \Diamond \mathbf{y} = (\mathbf{x} \otimes \mathbf{y}) \oplus (\mathbf{y} \otimes \mathbf{x})$

which, in our example, would be

$$\mathbf{x} \Diamond \mathbf{y} = \left((x_i \cdot y_j) \oplus (x_j \cdot y_i) \right)_{i,j \in \{1,2,3\}} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

There is a curious relationship between SOP and PIP and Norms :-

$$\boxed{\tfrac{1}{2} \| \mathbf{x} \Diamond \mathbf{y} \| + \langle \mathbf{x}, \mathbf{y} \rangle^2 = \|\mathbf{x}\| \cdot \|\mathbf{y}\|}$$

as the following calculation shows.

$$\begin{aligned} \| \mathbf{x} \Diamond \mathbf{y} \| &= \sum_{i,j} (x_i \cdot y_j - x_j \cdot y_i)^2 \\ &= \sum_{i,j} x_i^2 \cdot y_j^2 + \sum_{i,j} x_j^2 \cdot y_i^2 - 2 \cdot \sum_{i,j} x_i \cdot y_j \cdot x_j \cdot y_i \\ &= \sum_i x_i \cdot \sum_j y_j + \sum_j x_j \cdot \sum_i y_i - 2 \cdot \sum_i x_i \cdot y_i \cdot \sum_j x_j \cdot y_j \end{aligned}$$

(by idempotency, and re-arrangement and separation of summations)

$$\begin{aligned} &= \|\mathbf{x}\| \cdot \|\mathbf{y}\| + \|\mathbf{x}\| \cdot \|\mathbf{y}\| - 2 \cdot \langle \mathbf{x}, \mathbf{y} \rangle \cdot \langle \mathbf{x}, \mathbf{y} \rangle \\ &= 2 \left(\|\mathbf{x}\| \cdot \|\mathbf{y}\| - \langle \mathbf{x}, \mathbf{y} \rangle^2 \right) \end{aligned}$$

— whence $\mathbf{x} \Diamond \mathbf{y} = \mathbf{O}$ iff $\mathbf{x} = \mathbf{y}$, or $\mathbf{x} = \mathbf{o}$, or $\mathbf{y} = \mathbf{o}$ (by Pseudo Cauchy Equality).

...continued overleaf...

4. Boolean Cosine, Boolean Sine and Boolean Tangent.

We now have enough material to be able to define the following three 'pseudo' trigonometrical functions on the CH :-

Boolean COSINE

$$\begin{aligned} \mathtt{COSB}(x,y) &= \langle x,y\rangle^2 \,/\, \|x\| \,.\, \|y\| && (x,y \neq o) \\ &= 1 && (x = o, \text{ or } y = o) \end{aligned}$$

Boolean SINE

$$\begin{aligned} \mathtt{SINB}(x,y) &= \tfrac{1}{2}\,\|x \Diamond y\| \,/\, \|x\| \,.\, \|y\| && (x,y \neq o) \\ &= 0 && (x = o, \text{ or } y = o) \end{aligned}$$

Boolean TANGENT

$$\begin{aligned} \mathtt{TANB}(x,y) &= \|x \Diamond y\| \,/\, \left(2 \,.\, \langle x,y\rangle^2\right) && (x,y \neq o, x \not\perp y) \\ &= 0 && (x = o, \text{ or } y = o, \text{ but } x \not\perp y) \\ &= \infty && (x \perp y) \end{aligned}$$

Reciprocal functions CSCB, SECB, COTB, can be defined by $1/\mathtt{SINB}(x,y)$, *etc.*. Note that COSB and SINB are rational numbers in the closed interval $[0,1]$ and TANB being the ratio of SINB to COSB (*cf. tangent = sine / cosine*) is a rational or infinite number in the compactified infinite interval $[0,\infty]$. The 'range' of values clearly depends on the dimension of the underlying vector space (Boolean Algebra) *i.e.* the length of the bit-strings in question. For example, in the case of COSB for length 4 bit-strings, whilst the numerators $\langle x,y\rangle^2$ can take the values $0,1,4,9,16$ and the denominators $\|x\| \,.\, \|y\|$ can take the values $0,1,2,3,4,6,8,9,12,16$, the range of values that actually occurs is just

$$\text{range(COSB)} \;=\; [\;\; 0,\;\; 1/6,\;\; 1/4,\;\; 1/3,\;\; 4/9,\;\; 1/2,\;\; 2/3,\;\; 3/4,\;\; 1 \;\;]$$

and hence

$$\text{range(SINB)} \;=\; [\;\; 1,\;\; 5/6,\;\; 3/4,\;\; 2/3,\;\; 5/9,\;\; 1/2,\;\; 1/3,\;\; 1/4,\;\; 0 \;\;]$$

$$\text{range(TANB)} \;=\; [\;\; \infty,\;\; 5/1,\;\; 3/1,\;\; 2/1,\;\; 5/4,\;\; 1/1,\;\; 1/2,\;\; 1/3,\;\; 0 \;\;]$$

For the next higher level with bit-strings of length 16 the range is much more numerous since the two factors $\|x\|$ and $\|y\|$ which form the denominator in COSB can each take the

non-zero values $1 \ldots 16$. An ordered list of a selection of the possible values of COSB(x, y) for 16-bit-strings is appended at the end of this paper. It is impracticable to give the complete Table. There are $2^{16} = 65336$ distinct 16-bit-strings x. A complete (symmetric) table of values of COSB(x, y) at this Third Level of the **CH** would therefore have $2^{16} \times 2^{16}/2 = 2^{31} = 2147483648$ separate entries, many of which would contain duplicated values. (It would be interesting to list how many duplicates there were of each distinct value.) Such a huge table cannot realistically be computed *in toto*. This size problem would be even worse at the top (Fourth) Level of the **CH** whose bit-strings have length 256: a complete table would have $2^{256} \times 2^{256}/2 = 2^{511} \approx 6.7 \times 10^{153}$ entries — far more than the number of particles in the universe.

Here are a just some of the many facts that compare and contrast the rôles played by COSB, SINB, and TANB in **CH** and those of Cosine, Sine, and Tangent in Euclidean spaces. The most important one is the first :-

(1) $\mathtt{COSB}(x, y) + \mathtt{SINB}(x, y) = 1$ identically (*cf.* "$\sin^2 + \cos^2 = 1$").

This states that SINB is the unitary-supplement of COSB. In effect, this means that we need never calculate SINB from the more complicated procedures using the Outer Products — we need only calculate COSB from the far simpler Pseudo Inner-Product and Norms and subtract the result from unity. Whence :-

(2) $\mathtt{TANB}(x, y) = (1 \dot{-} \mathtt{COSB}(x, y)) \,/\, \mathtt{COSB}(x, y) = \mathtt{SECB}(x, y) - 1.$

From orthogonality and the Pseudo Pythagoras law we have :-

(3) $\mathtt{COSB}(x, y) = 0$ and $\mathtt{SINB}(x, y) = 1$ iff x and y are orthogonal and both non-o.

(4) $\mathtt{COSB}(x, y) = 1$ and $\mathtt{SINB}(x, y) = 0$ iff x and y are identical or at least one is o.

(5) Two distinct non-o vectors (bit-strings) always have a Coincidence less than unity.

5. Right-Angled Triangles, general Triangles, Parallelograms.

To give ready access to the properties of triangles and parallelograms, I will use specific examples drawn from the **CH** level with bit-strings of order 3 or 4, *i.e.* with 3 or 4 component elements. here we shall still calculate the Outer Products in detail, even though we know that SINB is just the unitary-supplement of COSB.

(1) An Isoceles Triangle.

Let O, X, Y be points with coordinates $(0,0,0,0)$, $(0,1,1,1)$, $(1,1,1,0)$ respectively. Let x resp. y, be the vectors OX, OY, then $\|x\| = 3$, $\|y\| = 3$ and the triangle XOY is isoceles (two equal sides). The sum $x \oplus y$ is a vector z corresponding to a point Z with coordinates $(1,0,0,1)$ and $\|z\| = 2$. The new triangle OXZ is also isoceles, the side XZ having norm $\|((1,0,0,1) \oplus (0,1,1,1))\| = \|(1,1,1,0)\| = 3 = \|y\|$.

None of the sides of this isoceles triangle OXZ are orthogonal :

$$\begin{aligned}\langle x, y\rangle = \|x \otimes y\| &= \|(0,1,1,0)\| = 2,\\ \langle x, z\rangle = \|x \otimes z\| &= \|(0,0,0,1)\| = 1,\\ \langle y, z\rangle = \|y \otimes z\| &= \|(1,0,0,0)\| = 1.\end{aligned}$$

Now form the Outer Products :-

$$x \Diamond y = \begin{bmatrix}0&0&0&0\\1&1&1&0\\1&1&1&0\\1&1&1&0\end{bmatrix} + \begin{bmatrix}0&1&1&1\\0&1&1&1\\0&1&1&1\\0&0&0&0\end{bmatrix} = \begin{bmatrix}0&1&1&1\\1&0&0&1\\1&0&0&1\\1&1&1&0\end{bmatrix}$$

$$\|x \Diamond y\| = 10$$

Then some trigonometry :-

$$\begin{aligned}\mathtt{COSB}(x,y) &= \langle x, y\rangle^2 \;/\; \|x\|.\|y\| = 4/9\\ \mathtt{SINB}(x,y) &= \tfrac{1}{2}.\|x \Diamond y\| \;/\; \|x\|.\|y\| = 5/9\\ \mathtt{COSB}(x,y) + \mathtt{SINB}(x,y) &= 4/9 + 5/9 = 1\\ \mathtt{TANB}(x,y) &= (5/9)/(4/9) = 5/4\end{aligned}$$

(2) A Right Triangle.

Let O, X, Y be points with coordinates $(0,0,0)$, $(0,1,1)$, $(1,0,0)$ respectively. Let x resp. y, be the vectors OX, OY, then $\|x\| = 2$, $\|y\| = 1$. The sum $x \oplus y$ is a vector z corresponding to a point Z with coordinates $(1,1,1)$ and $\|z\| = 3$.

$$\begin{aligned}\langle x, y\rangle = \|x \otimes y\| &= \|(0,0,0,0)\| = 0,\\ \langle x, z\rangle = \|x \otimes z\| &= \|(0,0,1,1)\| = 2,\\ \langle y, z\rangle = \|y \otimes z\| &= \|(1,0,0,0)\| = 1.\end{aligned}$$

So $x \perp y$ and the triangle XOY has a right angle between the sides OX and OY. In effect, the side OX is the *Base*, the side OY is the *Upright*, and the side XY is the *Hypotenuse*.

$$x \Diamond y = \begin{bmatrix}0&1&1\\1&0&0\\1&0&0\end{bmatrix}$$

$$\|x \Diamond y\| = 4$$

and you can work out that

$$\begin{aligned}\mathtt{COSB}(x,y) &= 0\\ \mathtt{SINB}(x,y) &= 1\\ \mathtt{COSB}(x,y) + \mathtt{SINB}(x,y) &= 0 + 1 = 1\\ \mathtt{TANB}(x,y) &= 1/0 = \infty\end{aligned}$$

It also follows that in a Right Triangle such as this [1] we have :-

(A) $x \otimes z = x; \qquad \langle x, z \rangle = \|x\| = 2$

("projection of Hypotenuse onto Base").

(B) $y \otimes z = y; \qquad \langle y, z \rangle = \|y\| = 1$

("projection of Hypotenuse onto Upright").

(C) $\mathtt{COSB}(x,z) = \|x\| / \|z\| = \frac{2}{3} = \mathtt{SINB}(y,z)$

("cosine of angle between Base and Hypotenuse" = "Base/Hypotenuse"
= "sine of angle between Upright and Hypotenuse")

(D) $\mathtt{COSB}(y,z) = \|y\| / \|z\| = \frac{1}{3} = \mathtt{SINB}(x,z)$

("cosine of angle between Upright and Hypotenuse" = "Upright/Hypotenuse"
= "sine of angle between Base and Hypotenuse")

(E) $\mathtt{TANB}(x,z) = (1-0)/0 = \infty; \quad \mathtt{TANB}(y,z) = (1-\frac{2}{3})/\frac{2}{3} = 1/2$

$\mathtt{TANB}(y,z) = (1-\frac{1}{3})/\frac{1}{3} = 2/3$

Here we see that COSB and SINB are playing the rôles of "Cosine" and "Sine" whereas we have already seen that they also play the rôles of " Cosine^2 " and " Sine^2 ". This doubling of rôle appears to stem directly from the idempotency characteristic of all multiplications in the **CH**.

(3) General Triangles.

The Triangle Equation can be further elaborated :-

$$\begin{aligned} \|x \oplus y\| &= \|x\| + \|y\| - 2.\langle x, y \rangle \\ &= \|x\| + \|y\| - 2.\|x\|.\|y\|.\frac{\langle x, y \rangle}{\|x\|.\|y\|} \end{aligned}$$

i.e.

$$\|x \oplus y\| = \|x\| + \|y\| - 2.\|x\|.\|y\| \,.\, \mathtt{C}(x,y) \qquad (*)$$

where the new factor $\mathtt{C}(x,y)$ in $(*)$ differs from $\mathtt{COSB}(x,y)$ simply by not having the squared Pseudo Inner-Product, $\langle x, y \rangle^2$, in the numerator.

It is remarkable that this modified form $(*)$ of the Triangle Equation has the same form as the classical Euclidean Trigonometrical "Cosine Rule"

$$a^2 = b^2 + c^2 - 2.b\,.\,c\,.\,\cos(A)$$

for a general triangle with sides b and c and included angle A and opposite side a.

[1] Remember we are using '$\otimes$' for 'logical AND'

(4) Parallelograms.

Parallelograms in the CH differ from those in a proper inner product space :-

they have only a single diagonal.

This rather disconcerting feature arises from the fact that there is no arithmetic operation of 'subtraction' in the field of two elements and any of the subsequent vector operations derived therefrom. Each vector (bit-string) is its own 'negative' :-

Let $\mathbf{x}$ and $\mathbf{y}$ be two distinct non-$\mathbf{o}$ vectors, put $\mathbf{z} = \mathbf{x} \oplus \mathbf{y}$, and consider the four distinct points $\mathtt{O}$, $\mathtt{X},\mathtt{Y},\mathtt{Z}$ associated with $\mathbf{o},\mathbf{x},\mathbf{y},\mathbf{z}$. In a proper inner-product space over a field with characteristic 0 (*e.g.* the real numbers), these four points would form a Euclidean Parallelogram with sides $\mathtt{OX}$, $\mathtt{XZ}$, $\mathtt{ZY}$, $\mathtt{YO}$ and diagonals $\mathtt{oz} = \mathbf{x}+\mathbf{y}$ and, $\mathtt{xy} = \mathbf{x}-\mathbf{y}$. The lengths of these sides and diagonals satisfy the classical PARALLELOGRAM LAW :-

the sum of the squares of the lengths of all 4 sides
= the sum of the squares of the 2 diagonals

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \;=\; \|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2$$

But in a vector space over the two element field the diagonals $\mathtt{OX}$, $\mathtt{OY}$ are identified with the same vector $\mathbf{x} \oplus \mathbf{y}$ and so have the same lengths $\|\mathbf{x} \oplus \mathbf{y}\|$. Consequently the corresponding Parallelogram Law in CH is false except for the special "rectangle" case in which $\mathbf{x}$ and $\mathbf{y}$ are orthogonal and in which it therefore reduces to the Pseudo Pythagoras Law.

(5) Some 'Compound' expressions using COSB and SINB.

The classical theorems for the Cosine and Sine of a compound 'angle' $\alpha+\beta$ are :-

$$\cos(\alpha+\beta) \;=\; \cos(\alpha).\cos(\beta) \;-\; \sin(\alpha).\sin(\beta)$$
$$\sin(\alpha+\beta) \;=\; \sin(\alpha).\cos(\beta) \;+\; \cos(\alpha).\sin(\beta)$$

However, I have nowhere defined what might be meant by 'angle' in the context of the Combinatorial Hierarchy so we cannot attempt to derive completely analogous results for COSB and SINB. But what can be done is to examine compound expressions as suggested by the right hand sides of those classical results. The reductions of the formulæ follow from replacing SINB by its unitary-supplement COSB. The underlying arithmetic is of course in the ring of Rationals; there is no question of idempotency in these simple calculations.

Let $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, be three bit-strings in CH, then :

$$\begin{aligned}
\text{(a)}\quad & \mathtt{COSB}(\mathbf{x},\mathbf{y}).\mathtt{SINB}(\mathbf{z},\mathbf{y}) \;+\; \mathtt{SINB}(\mathbf{x},\mathbf{y}).\mathtt{COSB}(\mathbf{y},\mathbf{z}) \\
= \quad & \mathtt{COSB}(\mathbf{x},\mathbf{y}).\,(1-\mathtt{COSB}(\mathbf{y},\mathbf{z})) \;+\; (1-\mathtt{COSB}(\mathbf{x},\mathbf{y}))\,.\mathtt{COSB}(\mathbf{y},\mathbf{z}) \\
= \quad & \mathtt{COSB}(\mathbf{x},\mathbf{y}) + \mathtt{COSB}(\mathbf{y},\mathbf{z}) \;-\; 2.\mathtt{COSB}(\mathbf{x},\mathbf{y}).\mathtt{COSB}(\mathbf{y},\mathbf{z})
\end{aligned}$$

and

$$\text{(b)} \quad \texttt{COSB}(x,y).\texttt{SINB}(z,y) - \texttt{SINB}(x,y).\texttt{COSB}(y,z)$$
$$= \quad \texttt{COSB}(x,y) - \texttt{COSB}(y,z)$$

whilst

$$\text{(c)} \quad \texttt{COSB}(x,y).\texttt{COSB}(y,z) - \texttt{SINB}(x,y).\texttt{SINB}(z,y)$$
$$= \quad - \texttt{SINB}(x,y) + \texttt{COSB}(y,z)$$

$$\text{(d)} \quad \texttt{COSB}(x,y).\texttt{COSB}(y,z) + \texttt{SINB}(x,y).\texttt{SINB}(z,y)$$
$$= \quad \texttt{SINB}(x,y) - \texttt{COSB}(y,z) + 2.\texttt{COSB}(x,y).\texttt{COSB}(y,z)$$

(4) Conclusions.

In this short, introductory account of some work that I had begun back in 1984 I have shown that it is feasible to introduce into the Combinatorial Hierarchy (**CH**) the notions comparable to the fundamental ideas of Euclidean Trigonometry.

The ideas of the Cosine, Sine and Tangent of an angle re-appear in modified forms as the Boolean Cosine COSB, Boolean Sine SINB and Boolean Tangent TANB of the 'angle' between two vectors (bit-strings), based on the ideas of a Pseudo Inner-Product PIP and a Symmetrized Outer-Product SOP of two vectors.

A number of relationships are deduced between the PIP, SOP and Norms of vectors, including those that characterize a strong Triangle Equality and a Pseudo Pythagoras Law, and the Cauchy-Schwarz Inequality and Equality, *etc.*. These are utilised in some simple examples of low order to illustrate the kind of calculations that are to be encountered in this wholly novel extension to the classical theory.

The fundamental ideas are in no way restricted to the case of the (finite) Combinatorial Hierarchy which happens to be of special interest to ANPA members, and which, though it happens to be both a graded vector and a graded Boolean Ring, is only of interest here insofar as each grade ('Level') is a vector space over the two-element field and also a finite Boolean Ring.

The ideas have far reaching applications to any finite Boolean Ring (a ring with at least two elements $0, 1$, in which every element is idempotent ($x^2 = x$), since any such Boolean Ring can be similarly equipped with the necessary Pseudo Inner Product, Norm and Symmetrized Outer Product. It is expected that comparable ideas can be developed in the context of infinite Boolean Rings following the introduction of limiting processes in the correspondingly modified definitions of the Pseudo Inner Product, Norm and Symmetrized Outer Product.

oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo

— table appended —

A PARTIAL TABLE OF VALUES OF COSB FOR BIT-STRINGS OF LENGTH 16

This Table of 180 values in both decimal and fractional format was constructed *via* a Pascal Program by creating 4096 pairs x, y of quasi-random 16-bit-strings; forming the integer numerator $\langle x, y \rangle^2$ and the integer denominator $\|x\| . \|y\|$ for COSB(x, y), and the value of the latter as their real ratio; saving ASCII-string values of these three quantities to a text file; sorting this file into ascending order of COSB(x, y), filtering this sorted file to cast out duplicated values, and finally editing the resulting file to create a formatted 'include file' for use with a TEX processor.

Note the increasing sparseness of values towards the high end of the range (larger numerators (PIPs) — bit-strings further from orthogonality, closer to coincidence).

0.00000,	0/ 12;	0.01667,	1/ 60;	0.01818,	1/ 55;	0.01852,	1/ 54;
0.02041,	1/ 49;	0.02083,	1/ 48;	0.02222,	1/ 45;	0.02273,	1/ 44;
0.02381,	1/ 42;	0.02500,	1/ 40;	0.02778,	1/ 36;	0.02857,	1/ 35;
0.03030,	1/ 33;	0.03125,	1/ 32;	0.03333,	1/ 30;	0.03571,	1/ 28;
0.03704,	1/ 27;	0.04000,	1/ 25;	0.04167,	1/ 24;	0.04762,	1/ 21;
0.05000,	1/ 20;	0.05556,	1/ 18;	0.05714,	4/ 70;	0.06061,	4/ 66;
0.06250,	1/ 16;	0.06349,	4/ 63;	0.06667,	4/ 60;	0.07143,	1/ 14;
0.07273,	4/ 55;	0.07407,	4/ 54;	0.08000,	4/ 50;	0.08163,	4/ 49;
0.08333,	4/ 48;	0.08889,	4/ 45;	0.09091,	4/ 44;	0.09524,	4/ 42;
0.10000,	4/ 40;	0.10227,	9/ 88;	0.11111,	4/ 36;	0.11250,	9/ 80;
0.11429,	4/ 35;	0.11688,	9/ 77;	0.12121,	4/ 33;	0.12500,	4/ 32;
0.12857,	9/ 70;	0.13333,	4/ 30;	0.13636,	9/ 66;	0.13846,	9/ 65;
0.14063,	9/ 64;	0.14286,	4/ 28;	0.14815,	4/ 27;	0.15000,	9/ 60;
0.16000,	4/ 25;	0.16071,	9/ 56;	0.16162,	16/ 99;	0.16364,	9/ 55;
0.16667,	4/ 24;	0.17778,	16/ 90;	0.18000,	9/ 50;	0.18182,	16/ 88;
0.18367,	9/ 49;	0.18750,	9/ 48;	0.19048,	4/ 21;	0.19753,	16/ 81;
0.20000,	4/ 20;	0.20455,	9/ 44;	0.20513,	16/ 78;	0.20779,	16/ 77;
0.21429,	9/ 42;	0.22222,	4/ 18;	0.22500,	9/ 40;	0.22727,	25/110;
0.22857,	16/ 70;	0.23077,	9/ 39;	0.23148,	25/108;	0.24038,	25/104;
0.24242,	16/ 66;	0.25000,	9/ 36;	0.25253,	25/ 99;	0.25397,	16/ 63;
0.25714,	9/ 35;	0.26042,	25/ 96;	0.26667,	4/ 15;	0.27473,	25/ 91;
0.27778,	25/ 90;	0.28125,	9/ 32;	0.28409,	25/ 88;	0.28571,	16/ 56;
0.29091,	16/ 55;	0.29630,	16/ 54;	0.29762,	25/ 84;	0.30000,	9/ 30;
0.30769,	16/ 52;	0.30864,	25/ 81;	0.31250,	25/ 80;	0.32000,	16/ 50;
0.32143,	9/ 28;	0.32468,	25/ 77;	0.32653,	16/ 49;	0.32727,	36/110;
0.33333,	9/ 27;	0.34615,	36/104;	0.34722,	25/ 72;	0.35556,	16/ 45;
0.35714,	25/ 70;	0.36000,	36/100;	0.36364,	16/ 44;	0.36735,	36/ 98;
0.37121,	49/132;	0.37500,	9/ 24;	0.37692,	49/130;	0.37879,	25/ 66;
0.38095,	16/ 42;	0.38462,	25/ 65;	0.39063,	25/ 64;	0.39560,	36/ 91;
0.39683,	25/ 63;	0.40000,	16/ 40;	0.40496,	49/121;	0.40833,	49/120;
0.40909,	36/ 88;	0.41667,	25/ 60;	0.41880,	49/117;	0.42857,	9/ 21;
0.44444,	16/ 36;	0.44545,	49/110;	0.44643,	25/ 56;	0.45000,	36/ 80;
0.45370,	49/108;	0.45455,	25/ 55;	0.45714,	16/ 35;	0.46296,	25/ 54;
0.46753,	36/ 77;	0.47115,	49/104;	0.48485,	64/132;	0.49000,	49/100;
0.49231,	64/130;	0.49495,	49/ 99;	0.50000,	16/ 32;	0.50794,	64/126;
0.51020,	25/ 49;	0.51042,	49/ 96;	0.51429,	36/ 70;	0.52083,	25/ 48;
0.52597,	81/154;	0.52893,	64/121;	0.53333,	16/ 30;	0.54444,	49/ 90;
0.54545,	36/ 66;	0.55556,	25/ 45;	0.55682,	49/ 88;	0.56250,	36/ 64;
0.57143,	36/ 63;	0.57857,	81/140;	0.58182,	64/110;	0.58333,	49/ 84;
0.59259,	64/108;	0.60000,	36/ 60;	0.60494,	49/ 81;	0.61250,	49/ 80;
0.61364,	81/132;	0.63636,	49/ 77;	0.64000,	64/100;	0.64286,	36/ 56;
0.64646,	64/ 99;	0.66667,	64/ 96;	0.66942,	81/121;	0.67500,	81/120;
0.68056,	49/ 72;	0.69444,	25/ 36;	0.71111,	64/ 90;	0.72727,	64/ 88;
0.73636,	81/110;	0.76563,	49/ 64;	0.79012,	64/ 81;	0.82645,	100/121;
0.83333,	100/120;	0.84028,	121/144;	0.85714,	36/ 42;	0.90000,	81/ 90;

WHY SPACE HAS THREE DIMENSIONS: A QUANTUM MECHANICAL EXPLANATION.

PETER MARCER WALTER SCHEMP

53 Old Vicarage Green *Lehrstuhl fur Mathematik I*
Keynsham, BS31 2DH, UK *University of Siegen, D-5900, GERMANY*
petermarcer@aikido.freeserve.co.uk *schempp@mathematik.uni-siegen.de*

(draft, 24 March 2000)

The theoretical physics of a quantum mechanical model of space, relativistic quantum holography, is described. It specifies three dimensions, such as is validated by the nature of our spatial experience, but where additionally, quantum non-locality, which Feynman described as the only mystery of quantum theory, is made manifest by means of observable phase relationships. For example, synchronicity between events, and other phenomena such as are described by the geometric/Berry phase, etc, which are outside the bounds of classical explanation.
It can therefore be hypothesized :
a) that we live in a entirely quantum mechanical world/universe and not a classical mechanical one (where quantum phenomena are confined to the microscopic scale) as is the current generally held scientific view,
b) that three spatial dimensions are a fundamental consequence of quantum mechanics,
c) that quantum holography is a natural candidate to explain quantum gravity, such that mass/inertia concerns not the eigenvalues of some operator, but rather the observable gauge invariant phases of a state vector, postulated to be that of the universe itself, as a whole, and
d) that this model provides a natural explanation in terms of relativistic quantum signal processing of why each individual's perception and cognition will be of a three dimensional world, defined similarly in relation to each individual's quantum state vector, describing its mind/body and associated gauge invariant phases or mindset, which have observable consequences, such that mental processes and events can cause neural events and processes!
These testable hypotheses, if validated, will have profound implications for our understanding, radically changing our scientific perspective on the world, as we enter the new millennium.

1 INTRODUCTION

Quantum holography [Schempp, 1992; Marcer, Schempp, 1999] as defined in terms of the Heisenberg group, which provides, based on the transactional interpretation of quantum mechanics [Cramer, 1986], the mathematical foundations of the signal processing [Schempp, 1986] at work in functional magnetic resonance imaging machinery [Schempp, 1998] is a natural application of the concept of the Berry phase [Anandan, 1992].

This follows from the fact that the three dimensional Heisenberg group G with generators of the form

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \text{ written as (x,y,z) for convenience} \qquad (A)$$

qualifies the generic Hamiltonian with the parametric dependence (x,y,z), so enabling it to describe a part of a larger system as if it were isolated, with some observable effects, which are gauge invariant phases [Resta, 1997]. For a quantum system having such parametric dependence cannot be isolated and parameters schematize couplings with other variables not described by the given Hamiltonian, or more generally with the "rest of the universe" to quote Berry's original words. Thus, in quantum holography, phase difference, which is the essential quantity of physical significance, is potentially a gauge invariant quantity of the state vector and physically observable, even though it cannot be expressed as the eigenvalue of some operator, (see section 3.

2 THE HEISENBERG GROUP G

G [Schempp, 1992] with unit (0,0,0) is a simply connected nilpotent Lie group, diffeomorphic to the differential manifold $(R\oplus R)xR$, with one dimensional centre $C_G = \{(0,0,z) \mid z\epsilon R\}$, where the Haar measure of G is Lebesgue measure $dx\otimes dy\otimes dz$ of the underlying differential manifold R^3, and the canonical basis {P,Q,Z} of g, the Lie algebra of G, formed by the upper triangular matrices $\{(x,y,z)-(0,0,0) \mid x,y,z\epsilon R\}$, defines the Heisenberg commutation relations as follows,

$$[P,Q] = PQ - QP = Z$$
$$[P,Z] = 0$$
$$[Q,Z] = 0$$

as was known to Weyl in 1928. Furthermore :
a) x and y, forming a Fourier duality pair, are the parameters, which determine the quantum holographic encoding/decoding procedure in the hologram plane $(R\oplus R)$ of the particular 3D object illumination, for both the degenerate and non-degenerate** filtered back propagation formulae of phase conjugate coherent four wavelet mixing under the harmonic analysis of Fourier transform action, [Schempp, 1992],

---

** as many frequences as there are degrees of freedom.

b) G is a nilpotent Lie group, where the log diffeomorphism is the mapping from such Lie groups to their associated Lie algebra. This mapping from G to its Lie algebra g and its dual, is a key feature of quantum holography distinguishing it from classical holography [Binz, Schempp, 1999], seen, for example, to be important in sections 3 and 4, and
c) quantum phase, a potentially physically measurable quantity of quantum holography, is, in these circumstances, sufficient to determine up to some degree of resolution, the complete state of the quantum mechanical system, ie an object's wavepacket or whole distribution of position and momentum values, known as the Wigner function, [Schleich, 1999] as a physical observable. This is because quantum holography, a Lie group-theoretical construct, is derived from the quantum mechanical commutation relations

$$q \circ p - p \circ q = i\hbar$$

Thus in a brain/quantum information system working quantum holographically :
i) the perception of an object would be an actual observation of the object's wavepacket, such that the parametric dependence x,y,z, as already described, makes the separation of the object from the rest of the universe possible, presenting the object to the brain/observer within the three dimensional manifold R^3. This provides a quantum mechanical explanation of why we sense/visualize our universe/world in geometric terms through three dimensions of space, independent of sensory apparatus and its spectral modality (typified, for example by the eyes and the visible spectrum of electromagnetic radiation, respectively), and
ii) the gauge invariant phases of the quantum state vector (see section 3) of the whole human organism, can be postulated to define brain/mind interaction [Marcer, Schempp, 1998], so as to provide a solution to the mind-body problem. Furthermore, the phase conjugation of the quantum holographic encoding /decoding procedures ensures that any 3D object image coincides with that of the object itself, and since any massive object must occupy a unique position in 3D space at any one time, all object imagery will be canonically labelled and therefore consistently categorized [Marcer, 1999]. Such quantum holographic encoding/decoding procedures concern a Bargmann-Fock model of quantics [Schempp, 1992; 1993](see Appendix). It is based on quantum field theory annihilation and creation operators for bosons, and U_v the infinite dimensional irreducible unitary representations of the Schrodinger type of G unique up to a unitary isomorphism in the standard Hilbert space $H = L^2(R)$, where $\underline{U}_v$ is the feedback or back-projection representation of G contragredient to U_v. Sensory modalities are therefore without limit, and would be naturally selected by organisms with respect to the most prevalent forms of environmental illumination; these constituting the reference illumination of the quantum

holography for the sensory apparatus/modality in question.

3 THE GEOMETRIC/BERRY PHASE; AN OUTLINE MATHEMATICAL ILLUSTRATION

Following Resta [1997], geometric phase begins with a quantum Hamiltonian H having a parametric dependence,

$$H(\xi)|\psi(\xi)> = E(\xi)|\psi(\xi)>$$

where ξ is defined in a suitable domain. A two dimensional ξ is chosen to illustrate the following facts, assuming that $|\psi(\xi)>$ is a non degenerate ground state for any ξ, so that phase difference $\Delta\vartheta_{12}$ between the ground eigenstates is defined by

$$\exp(-i\Delta\vartheta_{12}) = <\psi(\xi_1)|\psi(\xi_2)>/|<\psi(\xi_1)|\psi(\xi_2)>|$$

$$-i\Delta\vartheta_{12} = \ln(<\psi(\xi_1)|\psi(\xi_2)>/|<\psi(\xi_1)|\psi(\xi_2)>|)$$

hence

$$\Delta\vartheta_{12} = -\,\mathrm{Im}\log<\psi(\xi_1)|\psi(\xi_2)>$$

where Im is the imaginary part, and noting that $\log(|<\psi(\xi_1)|\psi(\xi_2)>|)$ has none.

Since any quantum mechanical state vector is arbitrary by a constant phase factor, this phase cannot have any physical significance, yet when one considers the total phase difference y along an arbitrary closed path C, all the arbitrary phases cancel out in pairs ie when N=4 for four points on C,

$$y = \Delta\vartheta_{12} + \Delta\vartheta_{23} + \Delta\vartheta_{34} + \Delta\vartheta_{41}.$$

$$= -\mathrm{Im}\log(<\psi(\xi_1)|\psi(\xi_2)><\psi(\xi_2)|\psi(\xi_3)><\psi(\xi_3)|\psi(\xi_4)><\psi(\xi_4)|\psi(\xi_1)>)$$

and it is realized that the phase difference y must be gauge invariant. This illustrates the fact that the geometric phase is a very general concept [Anandan, 1992].

The Berry phase y is most commonly defined in the continuum limit, such that there is a smooth curve C in the parameter domain, and $|\psi(\xi)>$ is single valued and varies in a differentiable way along the path, conditions which because of its Heisenberg Lie group formulation, quantum holography satisfies, so that when one discretizes the path C with any set of N points, and proceeds to the continuum limit, it can be shown more generally that the total phase difference y converges to the circuit of a real linear form called the Berry connection,

$$y = \sum_{1}^{N} \Delta\vartheta_{n,n+1} \quad \text{----->} \quad i\phi_C \langle \psi(\xi) \mid \nabla_\xi \psi(\xi) \rangle . \delta\xi$$

4 QUANTUM HOLOGRAPHY AS A CANDIDATE FOR QUANTUM GRAVITY

As Schempp [1993] has shown, quantum holography defines a creation /annihilation model of quantum mechanics, working by phase conjugate adaptive resonance, based on the three dimensional nilpotent Heisenberg Lie group G = (x,y,z). If therefore it is assumed that each phase conjugate adaptive resonant step, or creation/annihilation, corresponds to an increment in time dt, then the Lie derivative L_X appropriate to G

$$L_X = X_1 \partial/\partial x + X_2 \partial/\partial y + X_3 \partial/\partial z + X_4 \partial/\partial t$$

where X, the vector field

$$X = (X_1(x,y,z,t), X_2(x,y,z,t), X_3(x,y,z,t), X_4(x,y,z,t))$$

defines the invariant pathcurves, which are those found by solving the associated Pfaffian system,

$$dx/X_1 = dy/X_2 = dz/X_3 = dt/X_4.$$

But these are the four dimensional straightlines or geodesics, which are exactly the same geodesics as defined by general relativity. Moreover Feynman [1995] as described in lecture 6 of his 1962/63 lectures on gravitation, derives Einstein's equation for the stress energy tensor, as also following naturally from the construction of invariants with respect to infinitesimal (and therefore Lie) transformations, in general correspondance with the Lagrangian of a theory correct to all orders, as exists under an infinitesimal transformation of co-ordinates

$$x^\lambda = x'^\lambda + \varepsilon^\lambda \quad \text{where } x^\lambda = (x,y,z,t)$$

so as to define a reciprocal of the tensor field $g_{\nu\sigma}(x)$ according to the equation

$$g^{u\nu} g_{\nu\sigma} = \delta^u{}_\sigma \qquad (B)$$

where δ^{u}_{σ} is a Kronecker delta.This is exactly as the natural diffeomorphism of Lie transformation theory, or differentiable mapping with a differentiable inverse, says must exist for all orders, so that it can be hypothesised such theory will avoid the renormalization problems, inherently associated with theories of quantum gravity. That is to say, Lie tranformational theory, using Lie's second fundamental theorem will allow the theory of quantum gravity, based on quantum holography, to be linearized and expressed in terms of dimensionless coupling constants, so that it becomes renormalizable to all orders. Similarly the quantum mechanical Kronecker delta, a Heaviside operator equivalent to the Greens function of an integral formula over 3 spatial dimensions, tells us that the creation /annihilation operation of quantum holography is able to impose (at each step) for the metric tensor $g_{v\sigma}$, the condition of general covariance. A condition Einstein was able to intuit and then prove, leads to his formula for the stress-energy tensor, and one, which Feynman's equation (B) also predicts, leads away from this general covariance (at each step) to the general contragredience necessary for g^{uv}. This last prediction, associated with adaptive resonance, can therefore be intuited to relate g^{uv} to $g_{v\sigma}$, as defining the electro-weak theory that is essential to the phase conjugate adaptively resonant creation/annihilation evolutionary process [Marcer, 1999; Marcer, Dubois, 1992; Marcer, 1992; Amari, 1991] which quantum holography describes as taking place. Such an incremental evolutionary process described in terms of quantum mechanical creation /annihilation operators, would be quite different from that of a dynamic changing classical model based simply on general relativity. The possibility of this new model finds further confirmation in the fact that in standard quantum gravity, quantum theory is taken as implying the existence of exchange particles between matter ie the graviton, which must have two opposing helicity states. Whereas, these exchange particles, in quantum holography, are replaced by phase conjugate adaptive resonance, which consists of two opposing helicity processes, as Schempp [1992] has shown and has the Lorentz invariance necessary to be a description of quantum gravity. Here the active quantum mechanical medium of exchange is now the three dimensional continuum (x,y,z) defined by means of the Heisenberg Lie group G, where, it is to be noted, these spatial measures x,y,z can be distinguished relative to one another. This model would explain why mass is not defined like other quantum numbers, ie as the eigenvalues of some operator, but rather will be detemined in terms of continuum values of observable gauge invariant phases of the state vector Ψ of the universe, itself, as a whole. That is, will be defined, by means of the Universe's Berry or geometric phase; relative to the Universe as a whole ie with respect to the quantum vacuum state? Further Berry [1986] has shown that there exists a self-adjoint Hamiltonian operator obtained by quantizing some still unknown dynamic system without time reversal symmetry, whose phase space trajectories are chaotic and whose eigenvalues are the imaginary parts E_m, m =

1,2,3..., which can be expressed as phases, of the non trivial zeros of the Riemann Zeta function

$$\zeta(z) = \Pi_p (1-p^{-z})^{-1}, \text{ where } z = x+y\sqrt{-1} \text{ and p is prime.}$$

Here again x, y as a complex number, constitute a Fourier duality pairing. It is therefore of interest to speculate that this Hamiltion of a still unknown dynamical system, is that of the universe, where an irreversible evolution takes place by means of the phase conjugate adaptive resonance of quantum holography, and where Mach's principle of equivalence relates all masses to one another as a consequence of their quantum non-locality/entanglement, in accordance with Mach's global precept that the existence of any individual mass in the universe is consequence of the existence of all other masses. In this respect, it is of interest to note, that while the vacuum effects corresponding to quantum mechanical creation/annihilation models of electromagnetism, such as the Lamb Shift and the Casimir effect, constitute small and therefore second order corrections to such electromagnetic phenomena, that in the case of this proposed quantum holographic creation annihilation model of quantum gravity, they concern first order effects, illustrating the primacy of matter, inertia and fundamental energy, in line with the fact that these concern the universe as a whole. This model is therefore in line, with the concept that the universe and all matter/energy, ie Everything evolve out of the quantum vacuum ie from Nothing, [Puthoff, 1990; Haisch et al., 1998] as a sequence of "corrections/perturbations".

Other confirmations of quantum holography's candidature follows from that the fact that:

i) gauge invariance as a feature of quantum holography naturally incorporates the principle of general covariance necessary for general relativity [Feynman, 1996],

ii) it has a natural Lie commutator structure or bosonic formulation, for which there always exists a corresponding fermionic structure [Schempp, 1993]. Thus, each bosonic field has a fermionic partner, explaining how it can be expected that this formulation of quantum gravity, is renormalisable. That is, the ultraviolet behaviour of quantum theory is improved, because as the Lie formulation of quantum holography proves, the ordinary divergent bosonic (fermionic) contributions from Feynman diagram loops are cancelled by their partner fermionic (bosonic) contributions. Such a cancellation is the basis for the proposed approach to quantum gravity via the theory of supersymmetry,

iii) as Chapline [1999] has shown, his Heisenberg(-Weyl) group approach, has indeed much in common with both Schempp's quantum holography, supersymmetry, and string theory. It maybe therefore as shown above that quantum holography defines the hidden additional symmetry or condition, that

supersymmetry still requires to make it finite, and imposes on string theory the particular solution it needs to make it unique. That is, the Heisenberg group G, defines the dimension of space that is not at present an intrinsic property of string theory itself, see section entitled quantum gravity by Hatfield [Feynman, 1996], and

iv) the new model resolves the conflict between the necessary requirements that

a) as a quantum model describing a wholly attractive force, it must have two opposing helicities, and

b) the fact that general relativity concerns a symmetric tensor with ten components, that will have more than two degrees of dynamic freedom.

A conflict that would mean that the quantum theory would not be Lorentz invariant, which quantum holography is. It is resolved because quantum holography incorporates the gauge principles which lead to covariance, such that different classical field configurations describe the same physical state, and as are also necessary to the principle of equivalence.

5 CONCLUSION

Thus, while an extensive mathematical programme remains to be carried out, the presented evidence (and that of our senses) is already sufficient to hypothesize, that we live in a entirely quantum mechanical world/universe, and not a classical one where quantum phenomena are confined to the microscopic scale. It is perhaps timely here to quote Carl Friedrich von Weizsaeker's words in his talk given in memory of Werner Heisenberg at the Max Planck Institute in Munich on May 12, 1976.

"Diejenige Auffassung der Quantentheorie, die ich mir in Verfolgung des Wunsches von Heisenberg gebildet habe, eine weitere abgeschlossene Theorie zu entwickeln, veranlasst mich zu sagen. Wenn man das leisten koennte, was man Erachtens wird leisten koennen, auch den Begriff des driedimensionalen Raums aus der abstraken Quantentheorie herzuleiten -und damit auch die Begriffe von Feld und Teilchen aus der abstrakten Quantentheorie- dann ist Quantentheorie primaer nicht eine Theorie ueber Materie, sondern ueber information, genauer "ueber Bits in der Zeit."

Reality is therefore no longer confined to traditional 4 dimensional space time, nor even the 11 dimensions, as envisaged by string theorists, but maybe entended (as in quantum holography) to encompass the infinite dimensionality of Hilbert space. This may seem to be a wild conjecture. Yet even now in laboratories around the world, scientists are working on the design of quantum qubit computers over which users will require complete experimental control over the machine's Hilbert space of N dimensions, via the computer's N qubit machinery, necessary to calculate each quantum computational algorithm ; a feat which can only be achieved in classical computer by employing at least a 2^N bit

machinery, so that for large N, the calculation in the classical machine becomes impractical. In such quantum computers, the actions of the machine concerns the propagator

$$U = \exp(iH\Delta t/\hbar) = \exp(-i\Delta\vartheta)$$

where i is the $\sqrt{-1}$, $\Delta\vartheta$ is the change of phase, H is the quantum Hamiltonian, and Δt is a finite time interval -'the clock period'- at the end of which H is turned off, so that the propagator U acts as a gate by analogy with a classical computer. Such controls or 'gates', which are now thought feasible, can, it is hypothesized [Pribram, 1991; Marcer, Schempp, 1998, Marcer, Schempp, 1997], be effected in a quantum brain system working quantum holographically, utilizing the brain's geometric/Berry phase ie the total system phase difference effects, so that mental events cause neural events as well as vice versa, as postulated by Eccles [1986]. Such a Berry phase in this case, would provide a mathematical and quantum physically realizable model for the mind, which would be a global quantum non-local property of brains working quantum mechanically! That is, these total phase difference effects, would then be observable simultaneously at locations throughout the interior of the brain, so as to produce mind -->brain interactions. Similarly, the Berry phase as the brain's mind would be able to maintain a complete historical record of :
i) the time intervals (ie variable clock periods Δt) of the brain's/person's actions,
ii) where in 3 dimensional space the brain had been (ie the geometric component of its actions), and
iii) what quantum states these actions had passed through.
That is, a complete process mind record of the person's operational experience and actions, many of which could be labelled so as to enable their rapid recall for repeated future use, employing the brain as a transducer and the mind as the brain's operating system or control. And this Berry phase quantum model of the mind consisting of a complete process mental record of the history of the human organism, overcomes the old argument that postulating an homunculus implies infinite regression!

APPENDIX

Introduction of the complex mode co-ordinates [Schempp, 1992; 1993]

$$T = 1/2(P + iQ) \quad \text{and} \quad T^* = 1/2(P - iQ)$$

permits the different alternatives at say, the photon level that can exist in quantum complex linear superposition to be expressed in terms of creation /annihilation operators of an emitter/absorber model

$$a = U(T);\ a^* = U(T^*)$$

through the linear Schrodinger representation U of G, such that in terms of the number states $|n_k>$ which are the quantum states with n_k quanta occupying the mode k. These number states are the eigenstates of the number operator

$$N_k = aa^* \quad \text{and} \quad [a,a^*] = \pi$$

is the bosonic commutation relation. Thus if $H_v(\psi,\phi;\ x,y)$ expresses the probability of detecting a quantum, say a photon, of energy hv within a unit area attached to (x,y) in the hologram plane $R\oplus R$, where the wavelet mixing $\psi\otimes\phi$ necessary to the quantum holography, takes place, then $H_v(\psi,\phi;\ x,y)$ and $\underline{H}_v(\psi,\phi;\ x,y)$ respectively can be considered as the wavelet transform of the retarded signal $\psi\in L^2(R,dt)$ with respect to the advanced reference response wavelet $\phi\in L^2(R,dt)$ and vice versa, so that the time averaging performed by integation along the real line R by superposition of the net wavelets, which is expressed in the modular scalar product $< >$ of the complex Hilbert space $L^2(R,dt)$ effectively freezes the time t of the advanced signal wavelet packets and the retarded response wavelet packets into the spatial synchronization variables (x,y) of the symplectic hologram plane $(R\oplus R, \Omega_v)$, where $v \neq 0$ is the frequency. Thus the spatial encoding of the relative phase avoids the loss of phase information under wave packet reduction and the knowledge of the coordinates (x,y) with respect to the symplectic plane Ω_v allows the identification of the split photon channels in accordance with the non-local quantum property of individual photons passing by different pathways. This says that locally recording (x,y) makes the multiplexing coherent wavelet packet densities $\psi(t')dt'$ and $\phi(t)dt$ indistiguishable by relative time and phase corrections to the respective pathways, and that $H(\psi,\psi;\ ..)$ the holographic trace transform, which models the classical beam splitter quantum self interference experiment, is the natural transform to describe the Universe's state vector Ψ, so that $\Psi(t)dt \dashrightarrow H(\Psi,\Psi;\ x,y).dx\wedge dy$ would naturally extend a mapping of $L^2(R)$ into $L^2(R\oplus R)$ and the identity

$$H(\psi,\psi;\ x,y).dx\wedge dy\ = H(\phi,\phi;\ x,y).dx\wedge dy$$

implies $\psi(t)dt = c\phi(t)dt$, where c denotes a constant phase factor. This is the essential condition reflecting the fact that only the phase difference is important so that Ψ concerns the geometric/Berry phase.

REFERENCES.

Amari S. (1991) Dualistic Geometry of the Manifold of Higher-order Neurons, *Neural Networks*, 4, 443-451.

Anandan J. (1992) The Geometric Phase, *Nature*, **360**, 26, 307-313.

Berry M.V. (1986) Quantum Chaos and Statistical Nuclear Physics, eds. Seligam T.H. and Nishioka H. *Springer Lecture Notes in Physics 263* Springer, Berlin, Riemann's Zeta Function: a Model for Quantum Chaos? 1-17.

Binz E., Schempp W. (1999) Quantum Teleportation and Spin Echo. A Unitary Symplectic Spinor Approach. In: Aspects of Complex Analysis, Differential Geometry, Mathematical Physics and Applications, Dimiev S., Sekigawa K. editors, World Scientific, Singapore, 314-365.

Chapline G. (1999) Is theoretical Physics the Same Thing as Mathematics? *Elsevier Physics Reports*, **315**, 95 -105.

Cramer J. G. (1986) The Transactional Interpretation of Quantum Mechanics, *Review of Modern Physics*, 58 (3), 647-687.

Eccles J.C. (1986) Do mental states cause neural states analogously to the probability fields of quantum mechanics? *Proceedings of the Royal Society of London*, **B240**, 433-451.

Feynman R. P. (1999) Lectures on Gravitation, Penguin, London.

Haisch B., Rueda A., Puthoff H. E. (1998) Advances in the Proposed Zero-Point Field Theory of Inertia, *Proceedings 34th AIAA Joint Propulsion Conference*, paper AIAA 98 3143.

Marcer P. (1992) A Specification of the Unified Field, *Proceedings of the 13th International Congress of Cybernetics*, Namur, Belgium, 24-28th August, Symposium VII, Dubois/Mertens, 161-164.

Marcer P. (1999) Hypercomputation, *International Journal of Computing Anticipatory Systems*, Proceedings of CASYS'99. ed. by Dubois D. published by CHAOS (in press).

Marcer P., Dubois D. (1992) An Outline Model of Cosmological Evolution or Cosmogenesis, *Proceedings of the 13th International Congress of Cybernetics*, Namur, Belgium, 24-28th August, Symposium VII, Dubois/Mertens, 157-160.

Marcer P., Schempp W. (1997) Model of the Neuron Working by Quantum Holography, *Informatica*, **21**, 519-534.

Marcer P., Schempp W. (1998) The Brain as a Conscious System, International *Journal of General Systems*, 27 (1-3), 231-248.

Marcer P., Schempp W. (1999) Quantum Holography - The Paradigm of Quantum Entanglement, CP465, Computing Anticipatory Systems - Second International Conference edited by Dubois D. *The American Institute of Physics*, 461-467.

Pribram K.H. (1991) *Brain and Perception, Holonomy and Structure in Figural Processing*, Lawrence Erbaum, Publishers, Hillsdale, NJ.

Puthoff H. (1990) Everything from Nothing, *New Scientist*, **28th July**, 52-55.
Resta R. (1997) The Berry Phase, *Europhysics News*, 28, 19.
Schempp W. (1986) Harmonic Analysis on the Heisenberg Group with applications in signal theory, *Pitman Research Notes in Mathematics Series 147*, Longman Scientific and Technical, London.
Schempp W. (1992) Quantum Holography and Neurocomputer Architectures, *Journal of Mathematical Imaging and Vision*, 2, 279-326.
Schempp W. (1993) Bohr's Indeterminacy Principle in Quantum Holography, Self-adaptive Neural Network Architectures, Cortical Self-organization, Molecular Computers, Magnetic Resonance Imaging and Solitonic Nanotechnology, *Nanobiology*, 2, 109-164.
Schempp W. (1998) *Magnetic Resonance Imaging : Mathematical Foundations and Applications*, John Wiley & Sons, New York.
Schleich W.P. (1999) *Sculpting a Wavepacket*, Nature, **397**, 21st January, 207-208.

The Dirac Algebra and Charge Accommodation

Peter Rowlands* and J. P. Cullerne†

*IQ Group and Science Communication Unit, Department of Physics, University of Liverpool, Oliver Lodge Laboratory, Oxford Street, P.O. Box 147, Liverpool, L69 3BX, UK. E-mail prowl@hep.ph.liv.ac.uk and prowl@csc.liv.uk

†IQ Group, Department of Computer Science, University of Liverpool, Chadwick Laboratory, Peach Street, Liverpool, L69 72F, UK. E-mail john@jpcullerne.demon.co.uk

The broken symmetry responsible for the spectrum of fundamental particles and the characteristic properties of the strong, electromagnetic and weak interactions is derived from a representation of the three source terms or charges, *s*, *e*, *w*, by the quaternions operators, *i*, *j*, *k*. This symmetry is shown to be identical to that required by the Dirac equation in expressing the quantized version of the relativistic relation between energy, mass and momentum, and is used to derive the structures of baryons and mesons and their wavefunctions.

THE DIRAC ALGEBRA

Two symmetrical algebras are the basis of the work in this paper. These are the algebras of 4-vectors, with real vector units **i**, **j**, **k** and imaginary scalar *i*; and quaternions, with imaginary vector units *i*, *j*, *k* and real scalar 1. The imaginary quaternions follow the usual multiplication rules:

$$\begin{aligned} i^2 = j^2 = k^2 = ijk = -1 \\ ij = -ji = k \\ jk = -kj = i \\ ki = -ik = j \,, \end{aligned}$$

while the real vector units will be assumed to follow multivariate multiplication rules, identical to those for Pauli matrices, and parallel to those for quaternion algebra:

$$\begin{aligned} \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = 1 \\ \mathbf{ij} = -\mathbf{ji} = i\mathbf{k} \\ \mathbf{jk} = -\mathbf{kj} = i\mathbf{i} \\ \mathbf{ki} = -\mathbf{ik} = i\mathbf{j} \,. \end{aligned}$$

In effect, this means defining a 'full product' for two vectors **a** and **b** of the form

$$\mathbf{ab} = \mathbf{a}.\mathbf{b} + \mathrm{i}\,\mathbf{a} \times \mathbf{b} \,.$$

It can be readily shown that the algebra of the gamma matrices used in the Dirac equation

$$(\gamma^\mu \partial_\mu + im)\,\psi = 0 \,,$$

is completely isomorphic to a combination of the quaternion and 4-vector algebras, with assignments of the form:

$$\gamma^0 = -ii \qquad \text{or} \qquad \gamma^0 = ik$$
$$\gamma^1 = \mathbf{i}k \qquad\qquad \gamma^1 = \mathbf{i}i$$
$$\gamma^2 = \mathbf{j}k \qquad\qquad \gamma^2 = \mathbf{j}i$$
$$\gamma^3 = \mathbf{k}k \qquad\qquad \gamma^3 = \mathbf{k}i$$
$$\gamma^5 = ij \qquad\qquad \gamma^5 = ij\,.$$

The complete algebra has 32 parts: 1 real scalar, 1 imaginary scalar, 3 real vectors, 3 imaginary vectors, 3 quaternions, 3 imaginary quaternions, 9 real vector quaternions and 9 imaginary vector quaternions [Rowlands, 1996a, b, 1998]. The existence of 32 parts suggests that it can be generated from a binomial combination of five 'primitive' (though composite) components, of which the γ matrices are a characteristic set, with terms of the opposite sign obtained by reversing the order of multiplication:

$1, \gamma^0 = ik, \gamma^1 = \mathbf{i}i, \gamma^2 = \mathbf{i}j, \gamma^3 = \mathbf{i}k, \gamma^5 = ij, \gamma^0\gamma^1 = ij\mathbf{i}, \gamma^0\gamma^2 = ij\mathbf{j}, \gamma^0\gamma^3 = ij\mathbf{k}, \gamma^0\gamma^5 = i, \gamma^1\gamma^2 = -i\mathbf{k},$
$\gamma^1\gamma^3 = i\mathbf{j}, \gamma^1\gamma^5 = ik\mathbf{i}, \gamma^2\gamma^3 = -i\mathbf{i}, \gamma^2\gamma^5 = ik\mathbf{j}, \gamma^3\gamma^5 = ik\mathbf{k}, \gamma^0\gamma^1\gamma^2 = k\mathbf{k}, \gamma^0\gamma^1\gamma^3 = -k\mathbf{j}, \gamma^0\gamma^1\gamma^5 = \mathbf{i},$
$\gamma^0\gamma^2\gamma^3 = k\mathbf{i}, \gamma^0\gamma^2\gamma^5 = \mathbf{j}, \gamma^0\gamma^3\gamma^5 = \mathbf{k}, \gamma^1\gamma^2\gamma^3 = -ii, \gamma^1\gamma^2\gamma^5 = j\mathbf{k}, \gamma^1\gamma^3\gamma^5 = -j\mathbf{j}, \gamma^2\gamma^3\gamma^5 = j\mathbf{i}, \gamma^0\gamma^1\gamma^2\gamma^3 =$
$j, \gamma^0\gamma^1\gamma^2\gamma^5 = -ii\mathbf{k}, \gamma^0\gamma^1\gamma^3\gamma^5 = ii\mathbf{j}, \gamma^0\gamma^2\gamma^3\gamma^3 = -iii, \gamma^1\gamma^2\gamma^3\gamma^5 = k, \gamma^0\gamma^1\gamma^2\gamma^3\gamma^5 = -i.$

Since there is reason to suppose that space, time, mass and charge are the most fundamental set of parameters available to physics [Rowlands, 1990, 1991] we might expect that the 32-part algebra containing their units is of fundamental physical significance, and that significant physics will also be contained in the 5-fold structure represented by the binomial 'primitives'.

It is certainly possible, using this algebra, to derive the Dirac equation from the relativistic momentum-energy conservation equation

$$E^2 - p^2 - m^2 = 0\,,$$

by, first factorizing and attaching the exponential term $e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})}$, so that

$$(kE + ii\,\mathbf{p} + ij\,m)\,(kE + ii\,\mathbf{p} + ij\,m)\,e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})} = 0\,,$$

and then replacing E and $\mathbf{p}$ in the first bracket with the quantum operators, $i\partial\,/\,\partial t$ and $-i\nabla$, to give

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm\right)(kE + ii\,\mathbf{p} + ij\,m)\,e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})} = 0\,.$$

This can be written in the form

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm\right)\psi = 0\,,$$

where the wavefunction

$$\psi = (kE + ii\,\mathbf{p} + ij\,m)\,e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})}\,,$$

and the vector elements associated with the second term disappear when we take $\mathbf{p}$ in a preferred direction, as will often be convenient. ($p = \mathbf{1}.\mathbf{p}$ can be replaced by $\mathbf{p}$ if we

assume, as is conventional, that only one direction of the vector is well-defined.) With the m term fixed as positive, the equation allows for four solutions

$$\psi_1 = (kE + ii\,\mathbf{p} + ij\,m)\,e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})}$$
$$\psi_2 = (kE - ii\,\mathbf{p} + ij\,m)\,e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})}$$
$$\psi_3 = (-kE + ii\,\mathbf{p} + ij\,m)\,e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})}$$
$$\psi_4 = (-kE - ii\,\mathbf{p} + ij\,m)\,e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})}\ .$$

It can be shown that these solutions are identical to the four produced by the conventional Dirac spinor:

$$\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$$

and that the new version of the Dirac equation can be derived from the conventional matrix representation and vice versa (see the following section and Appendix I). The four solutions represent the four combinations of particle and antiparticle, and spin up and spin down states. Reversal of the sign of kE produces the wavefunction for an antiparticle, while reversal of the sign of $ii\mathbf{p}$ changes the direction of spin. As in conventional theory, to obtain a scalar probability density $\psi\psi^*$, we need to use a combination of all four solutions.

C-LINEAR MAPS AND LIFTS

In our expressions for the Dirac equation and the Dirac wavefunction, $\mathbf{p}$ is understood to be a multivariate momentum vector with the usual three components. The most usual representation for multivariate vectors is the set of Pauli matrices, σ_0, σ_x, σ_y, σ_z, which means that we may write the Dirac equation as follows:

$$\left(ik\frac{\partial}{\partial t} + i\,\vec{\sigma}.\,\vec{\nabla} + ijm\right)\psi = 0\,,$$

where $\vec{\sigma}.\,\vec{\nabla}$ is understood now to mean the scalar product, $\sigma_x\partial_x + \sigma_y\partial_y + \sigma_z\partial_z$. This has an equivalent 2×2 matrix form:

$$\begin{pmatrix} kE + ii\,p_z + ij\,m & ii\,(p_x - ip_y) \\ ii\,(p_x + ip_y) & kE - ii\,p_z + ij\,m \end{pmatrix}\begin{pmatrix} \phi \\ \chi \end{pmatrix} = 0\,,$$

which leads to the coupled linear differential equations

$$(kE + ii\,p_z + ij\,m)\phi + ii\,(p_x - ip_y)\chi = 0\,,$$

$$ii\,(p_x + ip_y)\phi + (kE - ii\,p_z + ij\,m)\chi = 0\,.$$

Choosing the momentum to be along the z-axis ($p = p_z$), these differential equations decouple to leave

$$(kE + ii\,p + ij\,m)\phi = 0\,,$$

$$(kE - ii\,p + ij\,m)\chi = 0\,,$$

with

$$\phi = \psi_1 = (kE + ii\,p + ij\,m)\begin{pmatrix}1\\0\end{pmatrix}e^{-i(Et\,-\,pz)}\,,$$

$$\chi = \psi_2 = (kE - ii\,p + ij\,m)\begin{pmatrix}0\\1\end{pmatrix}e^{-i(Et\,+\,pz)}\,.$$

These are the positive energy solutions; the quaternion representation allows a simple deduction of the negative energy solutions:

$$\psi_3 = (-kE + ii\,p + ij\,m)\begin{pmatrix}1\\0\end{pmatrix}e^{i(Et\,+\,pz)}\,,$$

$$\psi_4 = (-kE - ii\,p + ij\,m)\begin{pmatrix}0\\1\end{pmatrix}e^{i(Et\,-\,pz)}\,.$$

To carry out quantum mechanical calculations, we need to define a scalar product for the wavefunctions. However, the wavefunctions in the quaternion / spinor form are effectively a tensor product of two representations.

It is convenient to use a C-linear map $\mathbf{F}$: $H^C \rightarrow C^2$ is a mapping from complex quaternions, H^C ($q_\mu \mathbf{i}_\mu$, $q_\mu \in C$) to C^2.

$$\mathbf{F}: H^C \rightarrow C^2, \quad \mathbf{Q} \rightarrow \mathbf{Q}\begin{pmatrix}1\\0\end{pmatrix},$$

where $\mathbf{Q}$ is a complex quaternion, and

$$\mathbf{i}_0 = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix},\quad \mathbf{i}_1 = i = \begin{pmatrix}0 & -i\\-i & 0\end{pmatrix},\quad \mathbf{i}_2 = j = \begin{pmatrix}0 & -1\\1 & 0\end{pmatrix},\quad \mathbf{i}_3 = k = \begin{pmatrix}-i & 0\\0 & i\end{pmatrix}.$$

Applying this to ψ_1, we obtain

$$\psi_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}-i\\ \kappa + i\varepsilon\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix},$$

where $\kappa = p\ /\ E$, $\varepsilon = m\ /\ E$, and $\otimes$ is the tensor product between the two representations. Hence all four solutions may take the form

$$\psi_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ 0 \\ \kappa + i\varepsilon \\ 0 \end{pmatrix} \quad \psi_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ -i \\ 0 \\ -\kappa + i\varepsilon \end{pmatrix} \quad \psi_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 0 \\ \kappa + i\varepsilon \\ 0 \end{pmatrix} \quad \psi_4 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ i \\ 0 \\ -\kappa + i\varepsilon \end{pmatrix}.$$

This mapping leads to the conventional Dirac equation for these states, which in the case of ψ_1 is

$$\begin{pmatrix} -iE & 0 & -im+p & 0 \\ 0 & -iE & 0 & -im-p \\ im+p & 0 & iE & 0 \\ 0 & im-p & 0 & iE \end{pmatrix}\begin{pmatrix} -i \\ 0 \\ \kappa + i\varepsilon \\ 0 \end{pmatrix} = 0 .$$

Taking the four solutions, ψ_1, ψ_2, ψ_3, ψ_4, as defined above, we combine their values of

$$\psi^\dagger \psi = \psi^* \psi ,$$

to obtain the probability density

$$4\,(-E^2 - p^2 - m^2) = -8E^2 ,$$

which becomes 1 on application of the normalising factor $i/\sqrt{2}E$ for each solution.

For a Dirac wavefunction of the form

$$\psi = (kE + ii\,\mathbf{p} + ij\,m)\, e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})}$$

the Hermitian conjugate,

$$\psi^\dagger = \gamma^0 \psi \gamma^0 = \psi^* = (kE - ii\,\mathbf{p} - ij\,m)\, e^{i(Et\,-\,\mathbf{p}.\mathbf{r})}$$

and the adjoint wavefunction,

$$\bar{\psi} = \psi^\dagger \gamma^0 = \psi^\dagger\, ik = (-iE - j\mathbf{p} + ii\,m)\, e^{i(Et\,-\,\mathbf{p}.\mathbf{r})} .$$

It is, however, more convenient to replace this with the same function multiplied from the left by $-ik$, so that

$$\bar{\psi} = (kE + ii\,\mathbf{p} + ij\,m)\, e^{i(Et\,-\,\mathbf{p}.\mathbf{r})} ,$$

as this function equally satisfies the adjoint equation

$$\left(\frac{\partial \psi}{\partial t}\, ik + \nabla \psi i - \psi ijm\right) = 0 ,$$

and is only ever used as a multiplier from the left.

From this, we may derive the bilinear covariants, and hence the current density terms (using the four solution sum and the normalising factor):

$$\bar{\psi}\gamma^0\psi = \bar{\psi}ik\psi = ik\,E^2/E^2 = ik$$
$$\bar{\psi}\gamma^1\psi = \bar{\psi}ii\psi = -\mathbf{i}\,p^2/E^2$$
$$\bar{\psi}\gamma^2\psi = \bar{\psi}ij\psi = -\mathbf{j}\,p^2/E^2$$
$$\bar{\psi}\gamma^3\psi = \bar{\psi}ik\psi = -\mathbf{k}\,p^2/E^2$$
$$\bar{\psi}\gamma^5\psi = \bar{\psi}ij\psi = -\,ij\,m^2/E^2\,.$$

The first four quantities are the components of the current-probability density 4-vector $\bar{\psi}\gamma^\mu\psi$.

FERMION BOSONS AND WAVEFUNCTIONS

The power of the current version of the Dirac algebra lies in the fact that it uses nilpotents (or square roots of zero) rather than more conventional forms of the Dirac wavefunction. The two may, however, be easily related. Consider the equation:

$$(iE - ik\,\mathbf{p} - im)\,(iE - ik\,\mathbf{p} + im) = 0\,, \qquad (1)$$

which is just one version of the usual energy-mass-momentum equation. The two bracketed terms are clearly different, and so are not square roots of zero or nilpotents. However, if we multiply from the left by $-j$ and from the right by j, we get:

$$-j(iE - ik\,\mathbf{p} - im)\,(i\mathrm{E} - ik\,\mathbf{p} + im)j = 0\,.$$

This now becomes:

$$(kE + ii\,\mathbf{p} + ij\,m)\,(kE + ii\,\mathbf{p} + ij\,m) = 0\,. \qquad (2)$$

The two bracketed terms are now identical, and so each is a nilpotent or square root of zero. Equation (1) is, in effect, the usual way of writing the Dirac equation, with the left-hand bracket becoming the differential operator and the right-hand bracket the wavefunction, whereas (2) is just as valid, but much more powerful. In incorporating the m term directly, it effectively goes beyond the conventional Dirac wavefunction towards a quantum field interpretation, though it retains, as we have shown, all the physical interpretation available to the conventional term.

The vector-quaternion Dirac equation, especially in its nilpotent form, is, however, simpler and more powerful than the conventional representation, because it allows a more direct insight into the equation's physical meaning. In the first place, we have a direct expression involving E, $\mathbf{p}$ and m for the wavefunction, and, secondly, we can immediately see the explanation for such things as fermionic- and bosonic-type wavefunctions.

Since the exponential terms in the wavefunctions multiply as scalars, we will often find it convenient, here, to discuss the quaternion operators such as $(kE + ii\ \mathbf{p} + ij\ m)$ as the 'wavefunctions' when we are examining the properties of superposed states.

The immediately obvious aspect of fermionic wavefunctions is that they are composed of noncommutative operators, and hence are clearly antisymmetric. Also, multiplying two identical fermion wavefunctions produces zero, as in

$$(kE + ii\ \mathbf{p} + ij\ m)\ (kE + ii\ \mathbf{p} + ij\ m) = 0\ .$$

Pauli exclusion becomes an automatic property of the fact that $(kE + ii\ \mathbf{p} + ij\ m)$ is a square root of zero or nilpotent, always leading to the product $E^2 - p^2 - m^2$ when multiplied by itself. Reversing the sign of either kE or $ii\ \mathbf{p}$ or both in one of the factors, however, always produces a nonzero scalar product, such as a multiple of $E^2 + p^2 - m^2$, or $-E^2 + p^2 - m^2$, when summed up over the four solutions. Bosons are superposed states of fermion and antifermion, so

$$(kE + ii\ \mathbf{p} + ij\ m)\ (-kE + ii\ \mathbf{p} + ij\ m)$$

gives the wavefunction for a vector boson (spin 1), while

$$(kE + ii\ \mathbf{p} + ij\ m)\ (-kE - ii\ \mathbf{p} + ij\ m)$$

gives the wavefunction for a scalar boson (spin 0). The product

$$(kE + ii\ \mathbf{p} + ij\ m)\ (kE - ii\ \mathbf{p} + ij\ m)\ ,$$

though not a conventional boson state, will also be a scalar term, and can be taken as the wavefunction for a Bose-Einstein condensation.

It is notable that a spin 0 boson cannot be massless, as

$$(kE + ii\ \mathbf{p})\ (-kE - ii\ \mathbf{p}) = 0\ .$$

Hence, the Higgs scalar is necessarily massive, like the spin 0 mesons, while the massless gauge particles (gluons and photons) must all have spin 1. A boson with a scalar wavefunction will, in addition, be incapable of description via the Dirac equation, as the eigenvalue of a quaternionic quantum operator can only produce a zero product when multiplied by another quaternionic term.

The quaternionic version of the wavefunction also clarifies the meaning of CPT symmetry. Yet another significant aspect of this quaternion Dirac algebra is that it is already effectively second quantized. The full theory of C, P, T symmetry and of annihilation and creation operators are given in Rowlands and Cullerne [1999].

SPIN

In addition to these newer concepts, various standard results can be accommodated into the new formalism by replacing the gamma matrices with quaternion operators. Particularly important is the derivation of fermion spin. In the conventional treatment of spin, we write

$$[\hat{\sigma}, H] = [\hat{\sigma}, i\gamma_0\gamma.\mathbf{p} + \gamma_0 m] .$$

Also,

$$\hat{\sigma}_l = i\gamma_6\gamma_5\gamma_l \text{ , with } l = 1, 2, 3$$

and

$$i\gamma_6\gamma.\mathbf{p} = i\gamma_6\gamma_1 p_1 + i\gamma_6\gamma_2 p_2 + i\gamma_6\gamma_3 p_3$$

while

$$\gamma_6 = ik; \quad \gamma_1 = \mathbf{i}i; \quad \gamma_2 = \mathbf{j}i; \quad \gamma_3 = \mathbf{k}i; \quad \gamma_5 = ij .$$

So,

$$\hat{\sigma}_1 = -\mathbf{i} ; \quad \hat{\sigma}_2 = -\mathbf{j} ; \quad \hat{\sigma}_3 = -\mathbf{k}$$

or

$$\hat{\sigma} = -\mathbf{1} ,$$

and

$$\gamma = i\mathbf{1} ,$$

where $\mathbf{1}$ is the unit (spin) vector.

Since $\gamma_6 m = ikm$ has no vector term and $\hat{\sigma}$ no quaternion, they commute, and we may derive the conventional

$$[\hat{\sigma}, \gamma_6 m] = 0$$

and

$$[\hat{\sigma}, H] = [\hat{\sigma}, i\gamma_6\gamma.\mathbf{p}] .$$

Now,

$$i\gamma_6\gamma.\mathbf{p} = -j\,(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3) .$$

So,

$$\begin{aligned}[\hat{\sigma}, H] &= 2j\,(\mathbf{ij}p_2 + \mathbf{ik}p_3 + \mathbf{ji}p_1 + \mathbf{jk}p_3 + \mathbf{ki}p_1 + \mathbf{kj}p_2) \\ &= 2ij\,(\mathbf{k}(p_2 - p_1) + \mathbf{j}(p_1 - p_3) + \mathbf{i}(p_3 - p_2)) \\ &= 2ij\,\mathbf{1} \times \mathbf{p} .\end{aligned} \tag{3}$$

In more conventional terms,

$$\begin{aligned}[\hat{\sigma}, H] &= 2iki\,(\mathbf{k}(p_2 - p_1) + \mathbf{j}(p_1 - p_3) + \mathbf{i}(p_3 - p_2)) \\ &= 2ik\,\gamma \times \mathbf{p} \\ &= 2\gamma_6\,\gamma \times \mathbf{p} .\end{aligned} \tag{4}$$

Simultaneously, if $\mathbf{L}$ is the orbital angular momentum $\mathbf{r} \times \mathbf{p}$,

$$\begin{aligned}[\mathbf{L}, H] &= [\mathbf{r} \times \mathbf{p}, i\gamma_6\gamma.\mathbf{p} + \gamma_6 m] \\ &= [\mathbf{r} \times \mathbf{p}, i\gamma_6\gamma.\mathbf{p}] .\end{aligned}$$

Taking out common factors,

$$\begin{aligned}[\mathbf{L}, H] &= i\gamma_6\,[\mathbf{r}, \gamma.\mathbf{p}] \times \mathbf{p} \\ &= -ki\,[\mathbf{r}, \mathbf{1}.\mathbf{p}] \times \mathbf{p} \\ &= -j\,[\mathbf{r}, \mathbf{1}.\mathbf{p}] \times \mathbf{p}\,.\end{aligned}$$

Now,

$$[\mathbf{r}, \mathbf{1}.\mathbf{p}]\ \psi = -i\mathbf{i}\left(x\frac{\partial\psi}{\partial x} - \frac{\partial(x\psi)}{\partial x}\right) - i\mathbf{j}\left(y\frac{\partial\psi}{\partial y} - \frac{\partial(y\psi)}{\partial y}\right) - i\mathbf{k}\left(z\frac{\partial\psi}{\partial z} - \frac{\partial(z\psi)}{\partial z}\right)$$
$$= i\mathbf{1}\ \psi.$$

Hence,

$$[\mathbf{L}, H] = -ij\,\mathbf{1} \times \mathbf{p}\,. \tag{5}$$

This, again, can be converted into conventional terms:

$$\begin{aligned}[\mathbf{L}, H] &= iki\,\mathbf{1} \times \mathbf{p} \\ &= i\,\gamma_6\,\gamma \times \mathbf{p}\,.\end{aligned} \tag{6}$$

Using either (3) and (5) or (4) and (6), we may write

$$[\mathbf{L} - \mathbf{1}/2, H] = 0$$

or

$$[\mathbf{L} + \hat{\boldsymbol{\sigma}}/2, H] = 0\,.$$

Hence, $(\mathbf{L} - \mathbf{1}/2)$ or $(\mathbf{L} + \hat{\boldsymbol{\sigma}}/2)$ is a constant of the motion.

HELICITY

The term

$$\hat{\boldsymbol{\sigma}}.\mathbf{p} = -p_1 - p_2 - p_3 = -p$$

is defined as helicity, and, since it has no vector or quaternion terms, and has only terms of the form $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$ in common with

$$i\gamma_6\gamma.\mathbf{p} = -j\,(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3)$$

and also clearly commutes with $\gamma_6 m = ikm$, then

$$[\hat{\boldsymbol{\sigma}}.\mathbf{p}, H] = 0$$

and the helicity is a constant of the motion.

For a particle with zero mass, the term $k\,E + ii\,p + ij\,m$ reduces to $k\,E + ii\,p$, where p actually represents $\hat{\boldsymbol{\sigma}}.\mathbf{p}$. E also becomes equal to $\pm p$. For positive energy states,

$$E = \hat{\boldsymbol{\sigma}}.\mathbf{p}\,.$$

So the spin is aligned antiparallel to the momentum (has left-handed helicity). Then,

$$ij\,(k\,E + ii\,p) = ij\,(k - ii\,)\,E = (ii\,-k)\,E$$

and the spinor wavefunction follows the rule:

$$ij\,u_L = -\,u_L\,. \tag{7}$$

For negative energy states,

$$E = -\,\hat{\sigma}.\mathbf{p}\,,$$

In this case, the spin is aligned parallel to the momentum (has right-handed helicity). Then,

$$ij\,(k\,E + ii\,p) = ij\,(k + ii\,)\,E = (ii\,+k)\,E$$

and the spinor wavefunction follows the rule:

$$ij\,u_R = u_R\,. \tag{8}$$

From (7) and (8), we may derive the relations

$$\left(\frac{1-ij}{2}\right) u_L = \left(\frac{1-\gamma_5}{2}\right) u_R = u_L$$

and

$$\left(\frac{1-ij}{2}\right) u_R = \left(\frac{1-\gamma_5}{2}\right) u_R = 0\,.$$

If we define the right- and left-handed components of the wavefunction, ψ_R and ψ_L by the expressions

$$\psi_R = \left(\frac{1+\gamma_5}{2}\right)\psi = \left(\frac{1+ij}{2}\right)\psi$$

and

$$\psi_L = \left(\frac{1-\gamma_5}{2}\right)\psi = \left(\frac{1-ij}{2}\right)\psi\,,$$

we may derive the explicit expressions

$$\left(\frac{1+ij}{2}\right)(k\,E + ii\,p + ij\,m) = \left(\frac{1+ij}{2}\right)(k\,(E+p) + m) \tag{9}$$

and

$$\left(\frac{1-ij}{2}\right)(k\,E + ii\,p + ij\,m) = \left(\frac{1-ij}{2}\right)\left(\frac{-km}{E+p}\right)(k\,(E+p) + m)\,. \tag{10}$$

It is apparent from (9) that $\psi_R = 0$ when $m = 0$ and $E = -p = \hat{\sigma}\,.\mathbf{p}$.

SYMMETRIES OF MASS AND CHARGE

The Dirac algebra we have described has an interesting structure, which is very similar to the $SU(5)$, which has been proposed for the combined strong, weak and electromagnetic interactions [Georgi and Glashow, 1974]; and also has a clearly defined $SU(3)$ component. It would be interesting if some relationship could be found between the wavefunction algebra and the fundamental representations of particles in the Standard Model.

Here, we introduce some symmetries which we believe must exist between the fundamental parameters space, time, mass and charge. Space and time are conventionally described by a 4-vector algebra. It would be convenient if we could show that the symmetrical quaternion algebra was applicable to the only other parameters which appear to be as fundamental, the sources of the four known interactions. Here, we may refer to 'charge' as covering the sources of the three nongravitational interactions (essentially, the coupling constants) and refer to them individually as strong, weak and electromagnetic 'charges' (s, e, w), in the same sense as we refer to 'charge conjugation'. The source of the gravitational interaction we will refer to as 'mass' (M), although it is, in fact, energy or mass-energy, rather than the rest mass associated with individual particles (m).

One immediately noticeable aspect of the force laws representing these interactions is that the gravitational coupling is negative for identical particles (representing attractive force) where all the others are positive (representing repulsive force). To use the most convenient form of charge, the electrostatic repulsion between two identical e charges, in Coulomb's law, has the opposite sign to the gravitational attraction between two identical masses M in Newton's law:

$$F_e = \frac{e_1 e_2}{4\pi\varepsilon_0 r^2}$$

$$F_g = -\frac{GM_1 M_2}{r^2}.$$

Though this has been seen as a fundamental problem, it finds a convenient solution when we represent mass by real, and charge by imaginary, numbers. The corresponding force laws can then be written in an entirely symmetrical form.

$$F_e = -\frac{ie_1 ie_2}{4\pi\varepsilon_0 r^2}$$

$$F_g = -\frac{GM_1 M_2}{r^2}.$$

If we extend this by analogy to the other two forces, we require the existence of three independent systems of imaginary numbers, alongside one real one, to accommodate the existence of three types of 'charge' alongside the uniquely real quantity mass. It may then be that mass and charge are the corresponding components of a naturally existing quaternion, which might be considered as symmetrical to the 4-vector representation of space and time, and it will be convenient, on occasions, to represent the charge values on quaternion axes which are precisely isomorphic to the vector axes, and the mass values on a real number axis equally isomorphic to the imaginary axis which represents time in 4-vector geometry.

The use of imaginary numbers to represent charges also solves another problem in fundamental physics: the existence of antiparticles. Mathematically, any equation involving a positive imaginary number also has to have a negative solution. If we have

imaginary charges of any kind, then there also have to be charges of the opposite sign, which is exactly what we mean by antistates. Even electrically neutral particles such as the neutron and the neutrino, with no *e* charge, must have strong and / or weak charges, which means that they must also have antistates in which these charges are reversed. This mathematical condition does not, of course, apply to real numbers, and so negative states of mass are not required by the quaternion representation.

The four fundamental parameters, space, time, mass and charge, also have other contrasting properties which will be significant to us. The most important of these is the fact that the last two quantities are conserved, while the first two are not. The laws of physics are generally structured in various ways to make explicit this distinction, and we can imagine that, if an exact symmetry applies in the case of real and imaginary quantities, then an exact symmetry will also apply to the case of the conserved and the nonconserved quantities. In principle, nonconservation will be found to be as precise a concept as conservation, and every property which is attributable to the conservation properties of mass and charge will also apply to the nonconservation properties of space and time. Typical nonconservation properties are gauge invariance, and translation and rotation symmetries. Gauge invariance, in the Yang-Mills theories, is, in effect, a *local* manifestation of the absolute variability in space and time coordinates, while mass(-energy) and charge remain *locally* conserved. Space and time translation symmetries reflect the absolute indifference of the laws of physics to the actual position in space or moment in time for any conservative system, while the charge and mass values of that system remain absolutely untranslatable along their respective axes.

Of particular significance to us will be the contrast between the rotation symmetry of vector space (or absolute indifference of a vector to the composition of the components along its axes) and the rotation *a*symmetry of the three components of charge. If our symmetry assumes an exact contrast between conserved and nonconserved quantities, then the separate values of *e*, *s* and *w* in a system will remain fixed and not rotatable into each other. This, we will show, has a clear manifestation in the laws of lepton and baryon conservation, and it will also be significant to the question of the stability of the proton.

The rotation asymmetry of the charge components, *e*, *s*, *w*, however, presents us with a problem if we are to represent them by quaternion units, such as ***i***, ***j***, ***k***, for the quaternion units, like those of vectors in space, are clearly rotation symmetric, and we cannot simply write down an expression such as ***ie***, ***js***, ***kw***, and retain the full symmetry which the system seems to require. We can, however, use such an expression if we also include, in some other way, the required extra degrees of freedom. Here, we find that the answer is to provide the system of quaternions with an additional *vector* component, just as we combined vectors and quaternions in the Dirac algebra. The vector component turns out to manifest itself as the property of 'colour' in the quark system, and our difficulty of reconciling seeming opposites turns out to be an opportunity for understanding on a fundamental basis the otherwise mysterious properties of quark confinement and asymptotic freedom, for the added vector property must have no physically observable manifestation.

The combination of vectors and quaternions used in describing charges, and hence 'particles', is remarkably similar to the algebra used in fermionic wavefunctions, and we have found that it is possible to map one directly onto the other. We have also found that the system follows the same overall pattern of $SU(5)$ with a clearly defined $SU(3)$ component. Although, we have developed a precise algebraic description for the particle structures, we intend to first approach the subject in a more naïve or heuristic manner. For this, we have found it useful to return to the version of the quark system, with integral charges, originally proposed by Han and Nambu [1965], but seldom used in recent years.

HAN-NAMBU QUARKS AND CHARGE ACCOMMODATION

In the paper of 1965, which introduced the concept of quark 'colour', Han and Nambu proposed a model in which the differently coloured quarks carried either unit or zero values of electric charge, in preference to the fractional charges assumed in the original version of the quark theory. In the Han-Nambu theory, the charge assignments were of the form:

	Blue	Green	Red
up	e	e	0
down	0	0	$-e$

rather than the more familiar

	Blue	Green	Red
up	$2e/3$	$2e/3$	$2e/3$
down	$-e/3$	$-e/3$	$-e/3$

required by the original quark theory. We can imagine the Han-Nambu colour labels as representing three phases which are made unobservable by the requirement for colour composites imposed by the gauge-invariant strong interaction. The non-observability of these phases is actually a specific requirement of the Han-Nambu theory.

However, by the time that the coloured quark theory came into general use in the early 1970s, data on the ratio of hadron / muon production in electron-positron annihilation events had indicated that the values of charge entering into these processes were fractional, and the fractional charge representation for the three quark colour phases was subsequently almost universally adopted. Strictly speaking, however, as authors such as Close [1979] have occasionally pointed out, the evidence requires only the assumption that colour singlets have not been observed within a certain energy range, and does not rule out the possibility that integral charges actually exist at the quark level, though they will be made unobservable by the perfect gauge invariance of the strong interaction. In principle, phenomenological evidence from electromagnetic processes, especially in the low energy region, cannot distinguish between a theory in which the fractional charges are intrinsic to charge structure and one in which unobservable integral phases are necessarily averaged out to observed fractional values. It is perhaps significant that a parallel phenomenon has now emerged in condensed matter physics, where Laughlin's theory [1983] of the fractional quantum

Hall effect allows the phenomenological observation of fractionally charged states to emerge from a system possessing only integral charges.

Our theory requires both models. In our interpretation, the Han-Nambu theory does not apply to quarks as such, but to irreducible representations of fundamental charge states. Quarks are a physical consequence of the existence of unit fundamental charges; and fractional charges are evidence that quarks are bound states of these fundamental charges. Without observing fractional charges, we could not claim that quarks were bound. We do not thus expect to observe quarks with integral charge states; but physical evidence for fractional charge should not be used to assume that the underlying group structure is based on such charge units. The charges of quarks are truly fractional and will be observed as such, *at any energy*, but they are not fundamental, any more than those found for quasiparticles. Evidence of fundamental unit charge states will be obtained only from the group structure.

The Han-Nambu representation, in this sense, has several immediately attractive features. It incorporates a means of discriminating between the three phases of the strong interaction without having to invent the otherwise arbitrary property of 'colour'; it provides an easier route to the linking of quarks and leptons, without the difficulty of explaining why the charge units are different; and it also restores *single* units of charge to a level of matter which we suppose to be elementary. ($2e/3$ in the fractional representation clearly cannot be a *single* unit; and e for the electron cannot be a single unit if it is possible for another fundamental particle to have $e/3$.) Ultimately, we will show that the Han-Nambu model for fundamental charges gives better predictions than the more 'mechanistic' assumption of irreducible fractional charge for some of the most important numbers in particle physics. It will therefore be with this model of unit charge structure that we will attempt to unravel the origin of the group structure involving the three nongravitational interactions, though the direct correspondence of the combinations of the charge units with the quarks as physically observed will make it convenient to describe the components of this structure using the same term.

The great advantage of the Han-Nambu model for our analysis is that it reduces the problem of explaining 'particles' to that of defining the existence of units of charge. The theory is given in Rowlands and Cullerne [1999] as is that of the A, B, C, D and E representations, which are shown in a reduced form (restricted to u, d) in Appendix II. (The c, d, t, b states can be found by replacing w with $-w$.) It is significant that the E representation breaks the rules which we have devised for charge accommodation, and so, although it is needed to complete the five-fold symmetry, it also suggests a sense in which this overall symmetry might be broken.

THE ALGEBRA OF CHARGE ACCOMMODATION

We have previously shown [Rowlands and Cullerne, 1999] that the three expressions

$$(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3).\mathbf{i}$$
$$(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3).\mathbf{j}$$
$$(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3).\mathbf{k}$$

become the charge allocations for the **R**, **B** and **G** quarks in a ddd baryon, the total (before normalization) being given by $(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3).\mathbf{1}$. Whatever values of $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ are used, there are only five independent solutions for these scalar products, and these are the ones represented in tables A-E. (The ordering of **i**, **j**, **k** is, of course, completely arbitrary and merely represents the selection of names for the colour labels.)

The mathematical structure needed to model charges seems to be of the same kind as that which we have used in the Dirac equation, combining quaternions and vectors, though this time derived logically rather than mathematically. Can they be related in a more direct mathematical way?

Let us take the case for the ddd combination, which is represented by the scalar product of the terms $(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3)$ and $(\mathbf{i} + \mathbf{j} + \mathbf{k})$. For a baryon, the quarks are in the A, B or C representations, with $\mathbf{r}_1$ and $\mathbf{r}_3$ necessarily representing different unit vectors, and $\mathbf{r}_2$ cycling through all three possible values via the strong interaction. Let us, therefore, for illustration, take $\mathbf{r}_1$ to be **k** and $\mathbf{r}_3$ to be **j**. We can now write $(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3)$ in the form

$$(-j\mathbf{r}_1 + i\mathbf{1}\mathbf{r}_2 + k\mathbf{r}_3) = (i\mathbf{i}j + \mathbf{i}i\mathbf{r}_2 - i\mathbf{k}k) = -i\mathbf{i}\,(-\mathbf{j}j + i\mathbf{i}\mathbf{r}_2 + \mathbf{k}k)\,.$$

The terms $-\mathbf{j}j$, ii, $\mathbf{k}k$ form a closed cycle of the same form as the quaternion operators k, ii, ij in the Dirac algebra, although, this time, as the products of vectors and quaternions, they are commutative. With the second term again successively multiplied by the three vector operators **i**, **j**, **k**, we can see that the algebra of charge accommodation has at its core a structure of the same form as that applied to the Dirac equation – $SU(5)$ embedding $SU(3)$ – and we will find that we can use this principle to associate specific charge structures associated with particular particle states to specific quaternion wavefunctions.

CONSTRUCTING A BARYON

We have found, both heuristically and algebraically, that we may obscure specific quaternion-charge assignments by always combining unit charges in groups of three in such a way as to never know whether we have, for example, $1je$, $0ie$ or $0ke$. The different colour permutations give the same charge structure and at the same time hide the quaternion assignments, unit je, for example, being indistinguishable from zero ie or ke. The obscuring of quaternion assignments can also clearly be accomplished by combining unit charges with unit anticharges. The origin of both baryons and mesons thus becomes a natural result of this process, and their charge structures can be represented as in Rowlands and Cullerne [1999]. It is convenient, however, to outline here the process by which we may construct a baryon.

For the moment, let us remain within the A representation. What does a neutron (udd) look like within this representation? Well, to start with we need to consider the charge allocation tables for the quarks u and d. Within the A representation the table for u is:

u quark	B	G	R
$+e$	$1j$	$1j$	$0k$
$+s$	$1i$	$0k$	$0j$
$+w$	$1k$	$0i$	$0i$

If **B**, **G** and **R** correspond to **i**, **j** and **k** respectively in $(-j\mathbf{r}_1 + i\mathbf{r}_2 + k\mathbf{r}_3)$, then this table is the result if $\mathbf{r}_1 = \mathbf{k}$, $\mathbf{r}_2 = \mathbf{i}$, $\mathbf{r}_3 = \mathbf{i}$, with positive electric charge taking unit values in **B** and **G** and zero in **R**, and negative electric charge taking unit value in **R** and zero in **B** and **G**.

The table for d therefore is:

d quark	B	G	R
$-e$	$0i$	$0k$	$1j$
$+s$	$1i$	$0j$	$0k$
$+w$	$1k$	$0i$	$0i$

The construction of a neutron therefore goes as follows: First we need to combine three different columns from the tables corresponding to the three quarks u, d and d. Is there any preferred way of doing this?

There are three ways of constructing the total charge for each representation. For the A representation the neutron may be constructed by either,

$(i+j+k)$ in a colour, $-j$ in another colour, 0 in a third

or,

$(i+k)$ in a colour, $-j$ in another colour, j in a third

or,

$(i+k)$ in a colour, 0 in another colour, 0 in a third.

From this we can construct a matrix representation of the neutron where the columns and rows become the degrees of freedom for the combination of three unit charges; that is,

$$\begin{pmatrix} i+j+k & -j & 0 \\ -j & j & i+k \\ 0 & i+k & 0 \end{pmatrix} \tag{11}$$

Along every column and every row there is a valid charge structure in a given colour for the neutron. The columns and rows are therefore colour labels. How, then, is the matrix for the neutron in the A representation related to its equivalents in B, C, D and E representations?

The matrix in (11) may be called the charge structure matrix for the neutron in the A representation. However, the arbitrariness of the colour labels means that there are six ways of writing this matrix which correspond to the six permutations of the three colours **B**, **G** and **R**. QCD requires that the combination of three unit charges must

therefore admit all these combinations with a colour state given by the antisymmetric colour singlet of $SU(3)$:

$$\psi \sim (BGR - BRG + GRB - GBR + RBG - RGB). \tag{12}$$

The action of an arbitrary $SU(3)$ operation:

$$O(a_1, a_2, ..., a_8) = \exp\left(i \sum_{k=1}^{8} a_k \lambda_k \right) \tag{13}$$

where a_k are continuous parameters and λ_k are the generators of $SU(3)$, would leave ψ invariant. The equivalent transformations of the colour labels in the matrix (11), are performed by simply applying the six orthogonal permutations of S_3:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{14}$$

The first three are even permutations (Det = +1) and the second three are odd permutations (Det = −1).

Transforming (11) by the orthogonal transformations above, leads to the six other ways of writing (11) which correspond to the six linearly independent terms of ψ:

$$\begin{pmatrix} i+j+k & -j & 0 \\ -j & j & i+k \\ 0 & i+k & 0 \end{pmatrix} \begin{pmatrix} i+j+k & 0 & -j \\ 0 & 0 & i+k \\ -j & i+k & j \end{pmatrix} \begin{pmatrix} j & -j & i+k \\ -j & i+j+k & 0 \\ i+k & 0 & 0 \end{pmatrix}$$
$$\begin{pmatrix} j & i+k & -j \\ i+k & 0 & 0 \\ -j & 0 & i+j+k \end{pmatrix} \begin{pmatrix} 0 & i+k & 0 \\ i+k & j & -j \\ 0 & -j & i+j+k \end{pmatrix} \begin{pmatrix} 0 & 0 & i+k \\ 0 & i+j+k & -j \\ i+k & -j & j \end{pmatrix} \tag{15}$$

The symmetry that admits these possibilities into the A representation of the neutron, is only a special case of the full symmetry arising through the inclusion of all representations A, B, C, D and E. This symmetry would allow for the application of the permutations in (14) to any of the individual charge components of (11). Each charge type in a baryon has three orientations in colour space. Since there are three charge types, the total number of charge structures for a baryon is $3 \times 3 \times 3 = 27$. We could have come to this conclusion by realising that within each of the representations A, B, C and D, there are 6 orientations of charge structure. The E representation is the odd one out. It does not obey the rules for charge accommodation and therefore quaternion-charge assignments collapse to the trivial:

$$\pm is \pm je \pm kw. \tag{16}$$

The E representation cannot have distinct orientations corresponding to the six colour combinations in (12). Only three orientations are possible corresponding to the three colours that (16) can occur in. The 24 matrices from A to D and the 3 matrices from E, form the total of 27 charge structures.

THE STRONG INTERACTION

The overall $SU(5)$ structure represented by the tables and the charge-accommodation algebra has a clearly defined $SU(3)$ component. The *s* charge is the only one which takes the same value in all baryons or combinations of three unit charges and it is the only one which only ever occurs in one unit of charge at any given time. The 'colour' invariance requires that the single *s* component in the combination cannot be specified as belonging to one of the unit charges and a mechanism (which may be described as the strong interaction) must exist for a continual exchange of the *s* component between the three quarks in a baryon. Gauge invariance further requires that even the notion of an instantaneous location for *s* must be impossible. According to the tables above, transitions between systems A, B and C are equivalent to a transfer of the unit *s* component between the **B**, **R** and **G** quarks, without any change in the values of the *e* and *w* components, and the *s* component is the only one that can be transferred in this way. Consequently, it is the effective exchange of a single strong charge between the three bound quarks which prevents the identification of any one quark by its colour.

The strong interaction has been accommodated via quantum field theory into a colour $SU(3)$ group. An exact $SU(3)$ symmetry is a precise equivalent of a continuous gauge-invariant transformation between three colour states as required by the simultaneous and equally probable existence of the three quark systems A, B and C. The definition of a quark requires it to be necessarily in a strong-interacting state, but transitions from systems A, B, C to systems D and E cannot be involved in the pure strong interaction because they involve the additional exchange of a weak (or electric) component of charge. The weak interaction (or combined electroweak interaction) may thus be found to occur because the D and E representations of quarks, cannot be obtained in general from any of the others without an exchange in weak or electric, as well as strong, charges between the units. A pure weak or electroweak interaction, however, cannot take place where strong charges are present.

CONSTRUCTING A MESON

It is also possible to obscure quaternion charge assignments by always combining unit charges with unit anticharges in such a way as to make assignments like 1*je*, 0*ie* or 0*ke* indistinguishable. This becomes the origin of the meson.

Once again, let us start by considering the A representation alone. What does the charge structure of the K^0 meson (d with s antiquark) look like within this representation? The charge allocation table for the s antiquark is as follows:

s antiquark	Anti B	Anti G	Anti R
$+e$	$0i$	$0k$	$1j$
$-s$	$1i$	$0j$	$0k$
$+w$	$z_p k$	$0i$	$0i$

The K^o charge structure is constructed by taking a column from the d quark table and combining it with the corresponding column from the s antiquark. The resulting matrix has the structure:

$$\begin{pmatrix} 0i+(1+z_p)k & 0 & 0j \\ 0 & 0j & 0i+(1+z_p)k \\ 0j & 0i+(1+z_p)k & 0 \end{pmatrix} \tag{17}$$

which displays the equivalence of unit- and zero assignments by possessing six orientations corresponding to the six colour combinations in (12). Although mesons and baryons have very different structures, the charge accommodation is achieved in exactly the same way.

APPENDIX I: DERIVATION OF THE QUATERNION FORM OF THE DIRAC EQUATION FROM THE MATRIX REPRESENTATION

Here we interpret the matrices as dyadics formed from quaternion (or, alternatively, 4-vector) components, arranged by row and column. Starting with the equation

$$(\alpha.\mathbf{p} + \beta m - E)\ \psi = 0$$

and taking (for convenience, without loss of generality) $p = p_y$, we obtain [Eiserle, 1969]

$$\alpha_y = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix},$$

Using [Eiserle, 1969]

$$\beta = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & i \\ -i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix},$$

and applying the unit 4×4 matrix to E, the Dirac equation becomes

$$(\alpha \cdot p + \beta m - E)\psi = \begin{pmatrix} -E & 0 & im & -ip \\ 0 & -E & ip & im \\ -im & -ip & -E & 0 \\ ip & -im & 0 & -E \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$$

where the column vector is the usual 4-component spinor, and the terms E and $\mathbf{p}$ are the quantum operators which give the eigenvalues represented by these symbols when applied to the exponential term of the wavefunction. (It will be convenient here to refer to the components of the 4×4 matrix in terms of these eigenvalues rather than in terms of the operators which produce them.)

We can interpret the rows and columns of this matrix as having either 4-vector or quaternion coefficients. Let us choose the quaternion operators

$$j, i, k, 1$$

as the respective coefficients of the 4 rows. The 4×4 matrix now becomes a single row bra matrix with the columns:

$$\begin{array}{c} -jE - ikm + ip \\ -iE - ikp - im \\ ijm + iip - kE \\ -ijp + iim - E \end{array} .$$

If we multiply these terms from the left by the respective *column* coefficients

$$i, -j, 1, k,$$

we obtain in each case the expression

$$-kE + ii\,\mathbf{p} + ij\,m ,$$

which, when multiplied from the right by a wavefunction beginning with this term, gives a zero product. In order to show that this is equivalent to the Dirac equation in matrix form, we need to show that the four solutions ψ_1, ψ_2, ψ_3, ψ_4, multiplied by the appropriate quaternion row coefficients, each result in expressions beginning with this term.

Suppose, therefore, that we have the four solutions, as previously assumed:

$$\begin{aligned} \psi_1 &= (kE + ii\,\mathbf{p} + ij\,m)\, e^{-i(Et - \mathbf{p}.\mathbf{r})} \\ \psi_2 &= (kE - ii\,\mathbf{p} + ij\,m)\, e^{-i(Et - \mathbf{p}.\mathbf{r})} \\ \psi_3 &= (-kE + ii\,\mathbf{p} + ij\,m)\, e^{-i(Et - \mathbf{p}.\mathbf{r})} \\ \psi_4 &= (-kE - ii\,\mathbf{p} + ij\,m)\, e^{-i(Et - \mathbf{p}.\mathbf{r})} . \end{aligned}$$

We can show that the terms

$$\begin{matrix} k\ \psi_1\ k \\ -j\ \psi_2 j \\ 1\ \psi_3\ 1 \\ i\ \psi_4\ i \end{matrix}$$

each produce the expression

$$(-kE + ii\ \mathbf{p} + ij\ m)\ e^{-i(Et\ -\ \mathbf{p}.\mathbf{r})}\ .$$

So, multiplying each of the terms

$$\begin{matrix} k\ \psi_1 \\ -j\ \psi_2 \\ 1\ \psi_3 \\ i\ \psi_4 \end{matrix}$$

from the left by $(-kE + ii\ \mathbf{p} + ij\ m)$ results in a zero product.

For convenience, we may rearrange these to give a ket matrix of the form:

$$\begin{pmatrix} \psi_4 \\ \psi_2 \\ \psi_3 \\ \psi_1 \end{pmatrix} = i\ \psi_4 - j\ \psi_2 + 1\ \psi_3 + k\ \psi_1\ ,$$

where the row coefficients are identical to the column coefficients of the bra matrix (and even allow the elimination of the $-$ sign before $-j$). The resulting equation is identical to the quaternionic Dirac equation, which we have previously derived by direct means, with the four solutions representing the four possible combinations of $\pm E$ and $\pm p$ states.

The derivation demonstrates that the reason for the use of 4×4 matrices is, in fact, the fundamentally quaternionic nature of the Dirac wavefunction. Ultimately, this is because of the 4-vector space-time used in relativistic equations, which is symmetrical with the quaternion algebra used for mass and charge. In our understanding, the use of quaternionic operators to define the weak, strong and electromagnetic charges (w, s, e) maps directly on to the use of the same operators for the terms E, $\mathbf{p}$, m in the wavefunction, the existence of these terms as independent units stemming ultimately from the separate identities of the fundamental parameters time, space and mass (T, $\mathbf{S}$, M). Because quaternion operators define the meaning of the rows and columns in the Dirac matrix, the only way we can map charges w, s, e on to the E, $\mathbf{p}$, m terms via these operators is to have 4×4 matrices, and hence a 4-dimensional space-time signature in the equation.

It is significant that the application of quaternion operators to w, s, e and E, $\mathbf{p}$, m – and, by implication, to T, $\mathbf{S}$, M – is incomplete, each of these groups of three terms requiring a fourth to complete it. We can identify the respective fourth terms, without

difficulty, as mass (M), angular momentum (J) and charge (Q). The 4 × 4 Dirac matrix, in effect, *incorporates the fourth term as a zero quantity*. But, while the matrix requires four quantities to which the columns and rows apply, the 4-component spinor allows only four possible solutions from their combination. Interpreting the solutions in terms of the number of relative sign combinations of the component terms allows only *three* of the terms to be nonzero. Significantly, the excluded term in each case is an invariant, the system requiring only one invariant quantity (e.g. m or M) to demonstrate the variability of the others (E, **p**, or T, **S**).

APPENDIX II: REDUCED VERSION OF THE A-E REPRESENTATIONS

A

		B	G	R
u	$+e$	$1j$	$1j$	$0i$
	$+s$	$1i$	$0k$	$0j$
	$+w$	$1k$	$0i$	$0k$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$1i$	$0i$	$0k$
	$+w$	$1k$	$0j$	$0i$

B

		B	G	R
u	$+e$	$1j$	$1j$	$0k$
	$+s$	$0i$	$0k$	$1i$
	$+w$	$1k$	$0i$	$0j$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0k$	$1i$
	$+w$	$1k$	$0i$	$0k$

C

		B	G	R
u	$+e$	$1j$	$1j$	$0k$
	$+s$	$0i$	$1i$	$0j$
	$+w$	$1k$	$0k$	$0i$
d	$-e$	$0j$	$0k$	$1j$
	$+s$	$0i$	$1i$	$0k$
	$+w$	$1k$	$0j$	$0i$

D

		B	G	R
u	$+e$	$1j$	$1j$	$0i$
	$+s$	$0k$	$1i$	$0j$
	$+w$	$0i$	$0k$	$1k$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$1i$	$0i$
	$+w$	$0k$	$0j$	$1k$

E

		B	G	R
u	$+e$	$1j$	$1j$	$0i$
	$+s$	$0k$	$0j$	$1i$
	$+w$	$0i$	$0k$	$1k$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0i$	$1i$
	$+w$	$0k$	$0j$	$1k$

REFERENCES

Close, F. E. [1979]. *An Introduction to Quarks and Partons* (Academic Press), 167.

Eisele, J. A. [1969]. *Modern Quantum Mechanics with Applications to Elementary Particle Physics* (John Wiley), 194.

Georgi, H. and Glashow, S. L. 1974]. 'Unity of all elementary-particle forces'. *Phys. Rev. Lett.*, **32**, 438-41.

Han, M. Y. and Nambu, Y. [1965]. 'Three-Triplet Model with Double *SU*(3) Symmetry.' *Phys. Rev*, **139 B**, 1006-10.

Laughlin, R. B. [1983]. 'Anomalous quantum Hall effect: an incompressible field with fractionally charged excitations.' *Phys. Rev. Lett.* **50**, 1395-8.

Rowlands, P. [1990], 'A new formal structure for deriving a physical interpretation of relativity'. *Proceedings of the Conference on Physical Interpretations of Relativity Theory II*, British Society for Philosophy of Science, 264-8.

Rowlands, P. [1991], *The Fundamental Parameters of Physics* (PD Publications, Liverpool).

Rowlands, P. [1996a], 'Some interpretations of the Dirac algebra'. *Speculat. Sci. Tech.*, **19**, 243-51.

Rowlands, P. [1996b], 'A new algebra for relativistic quantum mechanics'. *Proceedings of the Conference on Physical Interpretations of Relativity Theory V*, British Society for Philosophy of Science, 381-7.

Rowlands, P. [1998], 'The physical consequences of a new version of the Dirac Equation', in G. Hunter, S. Jeffers and J.-P. Vigier (eds.), *Causality and Locality in Modern Physics* (Kluwer), 397-402.

Rowlands, P. and Cullerne, J. P. [1999], 'A derivation of particle structures and the Dirac equation from fundamental symmetires', in *Aspects II*, ed. K. G. Bowden, ANPA, 155-91.

Editor's notice

With this paper (which I asked them to split into two parts) Peter Rowlands and John Cullerne created both more interest and more controversy within ANPA than I have seen for a long time. With this in mind I asked Basil Hiley and Clive Kilmister to referee the paper and the formal reports are reproduced here. However, due to time constraints, Peter and John have unfortunately not been given the opportunity to reply to these reports in this issue of the Proceedings. They are invited to do so in the next Proceedings and/or the Newsletter. Our apologies to them for this omission.

Keith Bowden

Comments on "THE DIRAC ALGEBRA AND CHARGE ACCOMMODATION" by PETER ROWLANDS and J. P. CULLERNE

C. W. KILMISTER (ed KGB)

Red Tiles Cottage, High St, Barcombe, Lewes, E. Sussex

There are two things about the first paper which give me some misgivings. The first is this: PRJPC seem to have hit on a rather nice trick about the Dirac equation which, if it is new, ought to get a bit more publicity. To avoid any confusion, if I take $(\gamma^1)^2 = (\gamma^2)^2 = (\gamma^3)^2 = -1$, $(\gamma^4)^2 = 1$, so $(\gamma^4 E + \gamma^1 p_1 + \gamma^2 p_2 + \gamma^3 p_3)^2 = E^2 - \mathbf{p}^2 = m^2$, then if $\gamma^5 = \gamma^1\gamma^2\gamma^3\gamma^4$, so $(\gamma^5)^2 = -1$ we can take the Dirac equation

$$i\gamma^4 \partial\psi/\partial t - i\gamma^\alpha \partial\psi/\partial x^\alpha = m\psi$$

and multiply through by γ^5:

$$i\gamma^{45} \partial\psi/\partial t - i\gamma^{\alpha 5} \partial\psi/\partial x^\alpha - \gamma^5 m\psi = 0.$$

Now $\gamma^5 = \gamma^{15}\gamma^{25}\gamma^{35}\gamma^{45} = \gamma^1\gamma^2\gamma^3\gamma^4$ so we can take γ^{45} etc as a new set of γ's and have

$$i\gamma^4 \partial\psi/\partial t - i\gamma^\alpha \partial\psi/\partial x^\alpha - \gamma^5 m\psi = 0.$$

PRJPC seem to think that this new form makes various things more transparent. Well, if it does so, it does not depend on their idiosyncratic way of rewriting the equation and I wonder if tying it up with that will mean that few people will see it.

The second point is this: Rewriting the γ^α as products of a vector-set and a quaternion-set is more or less the same as writing them as a direct product of two quaternion sets – the only point of one set being a vector-set seems, to me, to be to "keep them apart". Now, of course, this seeing C_4 as $C_2 \times C_2 = Q \times Q$ is quite old – even Eddington had cottoned onto it by 1944. And another way of writing it is to see $C_2 \times C_2$ as isomorphic with $C_2(\ldots)C_2$, the linear quaternion functions of a quaternion. Indeed in my thesis I used the representation $\gamma^1 = i(\ldots)i$, $\gamma^2 = j(\ldots)i$, $\gamma^3 = k(\ldots)i$, $\gamma^4 = 1(\ldots)i$, (which is of the opposite sign convention to that used above). Now does this do anything for you? Only, I think, if in the later working you finish by using quaternions singly ie, not in the form of linear functions. This, I think, PRJPC do; indeed their ψ seems to be a quaternion. I used to work with a quaternion ψ in my thesis but Feza Gursey (who was not yet famous and was a friend, at IC) tried to persuade me that I should be looking at an ideal, not the whole algebra, although I cannot now remember why.

Comments on "THE DIRAC ALGEBRA AND CHARGE ACCOMMODATION" by PETER ROWLANDS and J. P. CULLERNE

B. J. HILEY (ed KGB)

TPRU, Birkbeck, Malet Street, London, WC1E 7HX

Keith Bowden has asked me to explain my worries about the paper presented by Peter Rowlands and John Cullerne (PRJPC) in these Proceedings. PRJPC are making the claim that the strong, electromagnetic and weak charges can be described by quaternions and that their methods give predictions that are closer to the experimental results than the standard model. If their claims are correct then this could provide a significant breakthrough in high energy physics.

Their approach depends on an algebraic treatment of the Dirac equation based on quaternions. Here it is unnecessary to first find a matrix representation to solve the Dirac equation. Solutions can be found using algebraic methods. I have a particular interest in these methods and in fact have written several papers using these techniques (Hiley and Frescura 1980, 1981, 1984, 1987 Bohm and Hiley 1983, 1984). I also used to give a series of lectures at MSc level on Clifford algebras and spinors so I have a keen interest in the techniques used in PRJPC. It was for this reason that I tried to reconstruct the results presented in the first part of their work. Unfortunately I was unable to derive their results. The purpose of these comments are to highlight my difficulties so that we can get a better understand the claims made by PRJPC.

In Part 1 I simply present my comments on their paper in note form. In Part 2 I present the details of a simple model, namely, the 2-dimensional space-time neutrino to illustrate the techniques involved. This model has the advantage that the Clifford algebra is only 4-dimensional and therefore the main points are not lost in the details. I hope this will clarify the algebraic approach, which has been around since the 1930s, but does not appear in the modern physics literature (see Sauter, 1930).

Part 1

1. PRJPC start by considering the two sub-algebras in the Dirac Clifford algebra. These sub-algebras are

   (i) The 'vector ' part based on **i**, **j**, **k**. This is R(2) in the notation of Porteous (1969)
   (ii) The quaternion part based on ***i***, ***j***, ***k***. This is H.

They are combined into the complex Dirac algebra H(2)⊗C. This algebra has 32 dimensions and is isomorphic to R(2)⊗H⊗C. In order to connect with the Dirac equation expressed in terms of γ's, it is necessary to chose which elements in H(2)⊗C will play the role of the γ^{μ}. You can do this in at least six ways once you stipulate, as do PRJPC, that γ^1, γ^2, γ^3 must contain **i**, **j**, **k** respectively. They then write down two sets of possible γ^{μ}, choosing the second set to work out the 32 elements of the Clifford algebra, which are shown in their table on page 2.

2.. Let us start with the Dirac equation as written in Bjorken and Drell.

$$\left(i\gamma^{\mu}\frac{\partial}{\partial x^{\mu}} - m\right)\psi = 0 \qquad (1)$$

PRJPC start with

$$\left(\gamma^{\mu}\frac{\partial}{\partial x^{\mu}} + im\right)\psi = 0 \qquad (2)$$

which follows from (1) by simply multiplying by $-i$. PRJPC do not explain which set of γ^{μ} they use. Since the second set are used in the table on page 2, I first assumed that this was the set to use in the Dirac equation (1). However if I do use this set I find

$$\left[ik\frac{\partial}{\partial t} + i\left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) + im\right]\psi = 0$$

If I write $\nabla = \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right)$, I obtain

$$\left[ik\frac{\partial}{\partial t} + i\nabla + im\right]\psi = 0 \qquad (3)$$

which is not the equation PRJPC start with. This is

$$\left[ik\frac{\partial}{\partial t} + i\nabla + ijm\right]\psi = 0 \qquad (4)$$

Note the extra j in the mass term.

Thus the equation that PRJPC are studying is NOT the DIRAC equation IF we use the set of γ^{μ} listed in the 32 element list.

If however we start from the set of γ^{μ} defined by the FIRST column at the top of page 2, we can get equation (4) from the Dirac equation. Thus by direct substitution for the γ^{μ} into (1) we find

$$\left(i\frac{\partial}{\partial t} + ik\nabla - m\right)\psi = 0$$

Then multiplying through by $-ij$ we obtain equation (4).

Thus the freedom to choose a set of γ^μ can lead to difficulties with the Dirac equation. Now the γ^μ define a Lorentz frame so it is important to fix on one set and stick to this set throughout the calculation as long as we remain in that Lorentz frame. However Lorentz invariance holds so that we can always Lorentz transform to a new frame and work in that frame. Thus we must expect sets of Lorentz equivalent γ^μ.

Unfortunately the two sets of γ^μ listed on page 2 of PRJPC are not Lorentz equivalent. This raises the interesting question as to the precise relation between the two sets of γ^μ and the meaning of this relationship.

3. Let us now solve (4) using the techniques outlined by Sauter (1930). Let us write

$$\psi = Ae^{-i(Et-p.r)}.$$

Here A will be any general element of the Clifford algebra. We will write the 32 elements of the algebra as Γ_r where r = 1, ... , 32, so that Γ_1=1, $\Gamma_2=\gamma^0$, $\Gamma_3=\gamma^1$... running through the elements generated by the first column of γ's on page 2. Then

$$A = \sum_{r=1}^{32} a_r \Gamma_r$$

We will assume the a_r are not functions of $\boldsymbol{r}$ and t. Then equation (4) becomes

$$[kE + iip + ijm]Ae^{-i(Et-p.r)} = 0 \qquad (5)$$

We can write

$$\psi = \sum_{r=1}^{32} a_r \Gamma_r e^{-i(Et-p.r)} = [kE + iip + ijm]Me^{-i(Et-p.r)}$$

Because $[kE + iip + ijm]$ is nilpotent, this is clearly a solution of equation (5). Here $M = \sum b_B \Lambda_B$ where the sum is over a suitable set B of elements of the algebra such that

$$[kE + iip + ijm]M = A.$$

Then a solution of equation (4) is

$$\psi = [kE + iip + ijm]Me^{-i(Et-p.r)} \qquad (6)$$

The presence of M, omitted by PRJPC, is important since it is this factor that shows that the solution is an element of a minimal RIGHT IDEAL. This fact follows from the general result that right ideals can be generated by a nilpotent element by multiplying

from the right. (See Crumeyrolle 1990). This RIGHT IDEAL is the (conjugate) SPINOR. These facts are detailed in Part 2.

[Note: This way of generating solutions of equations of the type (4) using RIGHT IDEALS was discussed by Sauter 1930. Using right ideals means that in a matrix representation we are discussing rows rather than columns.]

Thus (6) is equivalent to finding a 4-component Dirac spinor. It is tedious, but trivial to show that (6) spans a 4-dimensional complex vector space. This is left as an exercise for the interested reader.

This completes my discussion of ψ_1 in PRJPC.

4. Now let us move on to consider the other solutions. Unfortunately these are incorrect as written out. They should read

$$\psi_2 = [kE - iip + ijm]e^{-i(Et+p.r)}$$
$$\psi_3 = [-kE + iip + ijm]e^{i(Et+p.r)}$$
$$\psi_4 = [-kE - iip + ijm]e^{i(Et-p.r)}$$

Notice the exponentials must be written in these forms otherwise we would not get the correct expression for the quaternions. This follows directly from equation (4).

ψ_2 corresponds to a +ve energy particle travelling in the –ve p-direction.
ψ_3 corresponds to a –ve energy particle travelling in the +ve p-direction.
ψ_4 corresponds to a –ve energy particle travelling in the –ve p-direction.

These solutions do not correspond to "spin up and spin down states" as claimed in PRJCP. Since each solution can be multiplied from the right by an aggregate $\Sigma a\Gamma$ each will produce a 4-dimensional spinor which contains the spin states. Thus PRJPC have generated four different four-spinors. These spinors can be written in a column as on page (3), but when written in this form they are NOT IDENTICAL TO THE FOUR-COMPONENT SPINOR produced by solving the conventional Dirac equation.
[For details showing how spin actually arises see the discussion in part 2.]

5. Now all these results are supposed to be justified in Appendix 1. Here I am afraid I cannot follow the argument. PR has tried to explain this in a supplementary e-mail response to my problems but I was still do not follow the argument.

PRJPC start with

a. the Dirac equation considered in the Dirac (α, β)-representation and
b. consider only $p = p_y$ component of the momentum and
c. display the Dirac equation as a matrix using the Dirac representation.

It should be noted that the work in the section "The Dirac Algebra" does not assume any matrix representation.

Now on page 20 we have the statement "We can interpret the rows and columns of this matrix as having either 4-vector or quaternion coefficients". We can certainly do this but what does this mean in the context of the Dirac equation?

PRJPC go on to 'justify' the solutions ψ_1, ψ_2, ψ_3, ψ_4 written down on page 3 form a Dirac four-spinor by restricting the Dirac equation expressed in terms of α and β to the case where to $p = p_y$. But the subsequent equations they 'derive' are not restricted to $p = p_y$. They are supposed to hold for $p = ip_x + jp_y + kp_z$. This point is not discussed. Furthermore we have already shown ψ_1, ψ_2, ψ_3, ψ_4, that do NOT form a Dirac four-spinor. Clearly something is wrong here.

6. I now move on to the section "C-Linear Maps and Lifts." Here **i**, **j**, and **k** are replaced by σ_x, σ_y, σ_z and then the σ's are written in the 'standard' 2x2 Pauli matrix form without writing the corresponding i, j, k in matrix form. One cannot go to a situation where some algebraic elements are in a matrix representation and the others are not. One must find a full matrix representation of all the γ's and then proceed from there, otherwise problems will arise.

To see how problems occur, consider the second equation in this section. Recall that $ik = \gamma^0$ can be represented by a 4x4 matrix. Now in this equation kE is multiplied by ϕ but k is a 4x4 matrix and ϕ is only a 2 element column vector. How can you multiply a 4x4 matrix into a 2 dimensional column vector?

I am totally mystified by all that follows and have not been able to go further. What I will now do is show how the algebraic approach works in a simple model.

Part 2 Algebraic approach to a 2-dim space-time neutrino

In order to try to understand how the algebraic method of solving the Dirac equation actually works, we will consider a much simpler problem, namely, the two-dimensional space-time neutrino. We can examine the details of this problem because the Clifford algebra is only 4-dimensional and so it is much more manageable. We show how the algebraic method works and indicate how these solutions are related to the matrix solutions used in physics. More details of these methods can be found in Chevalley (1954), Crumyerolle (1990), Hestenes and Sobczyk (1984), Lounesto (1979) and Riesz (1958)

Once the method is understood it can then easily be generalised to the case that PRJPC are considering. I hope this will make it clear where my difficulties with their paper arise.

1. Dirac neutrino equation.

We start with the Dirac equation

$$\left(\gamma^0\frac{\partial}{\partial t}+\gamma^1\frac{\partial}{\partial x}\right)\psi=0 \qquad (7)$$

where $(\gamma^0)^2=1; \quad (\gamma^1)^2=-1; \quad \gamma^0\gamma^1+\gamma^1\gamma^0=0.$

This Clifford algebra is thus spanned by $1,\gamma^0,\gamma^1,\gamma^0\gamma^1$. Assume a solution of the form

$$\psi = Ae^{-i(Et-px)} \qquad (8)$$

where $A=\sum_{i=1}^{4}a_i\Gamma_i$ is a Clifford number with

$$\Gamma_1=1; \quad \Gamma_2=\gamma^0; \quad \Gamma_3=\gamma^1; \quad \Gamma_4=\gamma^0\gamma^1.$$

Then substituting (8) into equation (7) we find

$$\left(-iE\gamma^0+ip\gamma^1\right)Ae^{-i(Et-px)}=0$$

or equivalently

$$\left(E\gamma^0-p\gamma^1\right)Ae^{-i(Et-px)}=0. \qquad (9)$$

Then $\psi=\left(E\gamma^0-p\gamma^1\right)\left(a\gamma^1+b\right)e^{-i(Et-px)}$ will be a solution since

$$\left(E\gamma^0-p\gamma^1\right)\left(E\gamma^0-p\gamma^1\right)=E^2-p^2=0.$$

This is what we would expect of a zero rest mass particle. Thus $E=\pm p$ is the condition that equation (9) will have a non-trivial solution.

Using $E=\pm p$ we obtain two solutions:-

$$\psi_1=E\left(\gamma^0-\gamma^1\right)\left(a\gamma^1+b\right)e^{-i(Et-px)} \quad \text{and} \quad \psi_2=E\left(\gamma^0+\gamma^1\right)\left(a\gamma^1+b\right)e^{-i(Et-px)}.$$

From now on we will absorb E and the exponential term into a and b for convenience. Thus the two solutions are

$$\psi_1=a+b\gamma^0-b\gamma^1+a\gamma^0\gamma^1 \quad \text{and} \quad \psi_2=-a+b\gamma^0+b\gamma^1+a\gamma^0\gamma^1.$$

These are in fact minimal right ideals as we will now show.

2. **Conjugate spinors as minimal right ideals.**

Consider the two primitive idempotents $\varepsilon_1=\tfrac{1}{2}\left(1+\gamma^0\gamma^1\right)$ and $\varepsilon_2=\tfrac{1}{2}\left(1-\gamma^0\gamma^1\right)$.

Note that $\varepsilon_1+\varepsilon_2=1$ as required. The right ideal is generated by multiplication from the

right so that

$$\tfrac{1}{2}(1+\gamma^0\gamma^1)\gamma^0 = \tfrac{1}{2}(\gamma^0 - \gamma^1) \quad \text{and} \quad \tfrac{1}{2}(1+\gamma^0\gamma^1)\gamma^1 = -\tfrac{1}{2}(\gamma^0 - \gamma^1)$$

Thus the basis of the ideal is $\tfrac{1}{2}(1+\gamma^0\gamma^1) \quad \tfrac{1}{2}(\gamma^0 - \gamma^1)$ so that the spinor can be written in the form

$$c(1+\gamma^0\gamma^1)+d(\gamma^0 - \gamma^1) = c + d\gamma^0 - d\gamma^1 + c\gamma^0\gamma^1 .$$

Thus if we let $c = a$ and $d = b$, we see that this spinor is the solution ψ_1 of the Dirac equation given in section 1.

To see how this ties up with the matrix representation, we can choose a convenient matrix representation

$$\gamma^0 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \qquad \gamma^1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \qquad \gamma^0\gamma^1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then

$$\tfrac{1}{2}(1+\gamma^0\gamma^1) \to \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad (\gamma^0 - \gamma^1) \to \begin{pmatrix} 0 & 0 \\ -i & 0 \end{pmatrix},$$

so that our (right) spinor is of the form $\begin{pmatrix} 0 & 0 \\ b & a \end{pmatrix}$.

Remember the E and the exponents have been absorbed into the a and b. The reason for writing the spinor as a row follows from the fact that

$$\begin{pmatrix} 0 & 0 \\ . & . \end{pmatrix}\begin{pmatrix} . & . \\ . & . \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ . & . \end{pmatrix}$$

which is exactly what is meant by a right ideal.

The above solution is for $E = +p$. If we use the primitive idempotent $\tfrac{1}{2}(1-\gamma^0\gamma^1)$ we find the solution for which $E = -p$. This solution is spanned by the right ideal $\tfrac{1}{2}(1-\gamma^0\gamma^1) \quad \tfrac{1}{2}(\gamma^0 + \gamma^1)$, which gives the general spinor

$$c(1-\gamma^0\gamma^1)+d(\gamma^0 + \gamma^1) = c + d\gamma^0 + d\gamma^1 - c\gamma^0\gamma^1 .$$

This is just ψ_2 above if $c = -a$ and $d = b$.

The matrix representation of this spinor is of the form $\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix}$.

We can now collect together our results.

$$\psi_1 = \left[a\left(1+\gamma^0\gamma^1\right)+b\left(\gamma^0-\gamma^1\right)\right]e^{-i(Et-px)},$$

and

$$\psi_2 = \left[a\left(1-\gamma^0\gamma^1\right)+b\left(\gamma^0+\gamma^1\right)\right]e^{-i(Et-px)}.$$

If we recall that in this 'small' Clifford algebra, $\gamma^0\gamma^1$ corresponds to γ^5 in the usual Dirac theory. Here $(1\pm\gamma^5)$ plays the same role as the helicity operator in the two-component neutrino theory. Thus in our case we see that $(1-\gamma^5)$ denotes spin parallel to the momentum p, while $(1+\gamma^5)$ denotes the spin anti-parallel to the momentum p. These correspond to left handed and right handed neutrinos respectively. Because of non-conservation of parity recall that right handed neutrinos do not exist.

3. **Anti-neutrinos.**

To obtain the anti-particles we must choose the negative energy solutions. The solutions satisfying the equation

$$\left(\gamma^0\frac{\partial}{\partial t}+\gamma^1\frac{\partial}{\partial x}\right)Ae^{i(Et+px)}=0$$

describes an anti-neutrino moving in the x-positive direction. Here we again obtain two solutions, which can be written in the form

$$\psi_1 = \left(\gamma^0-\gamma^1\right)Ae^{i(Et+px)} \quad \text{and} \quad \psi_2 = \left(\gamma^0+\gamma^1\right)Ae^{i(Et+px)}$$

It is straight forward to show that these give both right handed and left handed anti-neutrinos.

4. **Right ideals or left ideals?**

The method that we have used to solve the Dirac equation produces the right ideal, which is essentially the conjugate row spinor as we have shown above. To obtain the column spinors we need the left ideals

$$\begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}\begin{pmatrix} \cdot & 0 \\ \cdot & 0 \end{pmatrix} = \begin{pmatrix} \cdot & 0 \\ \cdot & 0 \end{pmatrix}$$

To obtain left ideals using the operator formalism we need to write the Dirac equation as

$$\psi\left(\gamma\frac{\overleftarrow{\partial}}{\partial t}+\gamma\frac{\overleftarrow{\partial}}{\partial x}\right)=0$$

Then we go through the corresponding procedure by always multiplying from the left. We then find the two left ideals are respectively

$$\psi_1 = a\left(1-\gamma^0\gamma^1\right)+b\left(\gamma^0-\gamma^1\right) \quad \text{and} \quad \psi_2 = a\left(1+\gamma^0\gamma^1\right)+b\left(\gamma^0+\gamma^1\right)$$

6. **Two other pairs of solutions.**

They are

$$\psi_1 = \left(\gamma^0-\gamma^1\right)Ae^{-i(Et+px)} \quad \text{and} \quad \psi_2 = \left(\gamma^0+\gamma^1\right)Ae^{-i(Et+px)}$$

This corresponds to the neutrino travelling in the negative x-direction and

$$\psi_1 = \left(\gamma^0-\gamma^1\right)Ae^{i(Et-px)} \quad \text{and} \quad \psi_2 = \left(\gamma^0+\gamma^1\right)Ae^{i(Et-px)}$$

which corresponds to the anti-neutrino travelling in the negative x-direction.

This completes our discussion of the 2-dimensional Dirac neutrino equation.

7. References.

Bohm D. J. and Hiley B. J., *Relativistic Phase Space Arising out of the Dirac Algebra*, Old and New Questions in Physics, Cosmology, Philosophy and Theoretical Biology, ed. A. van der Merwe, 67-76, Plenum Press (New York, 1983).

Bohm D. J. and Hiley B. J., *Generalization of the Twistor to Clifford Algebras as a Basis for Geometry*, Rev. Brasil. Fis.,Vol. Especial Os 70 anos de Mario Schonberg,1-26, (1984).

Chevalley C., *The algebraic theory of spinors.* Columbia University Press, 1954.

Crumeyrolle A., *Orthogonal and symplectic Clifford algebras*, Kluwer, 1990.

Frescura F.A.M. and Hiley B. J., *The Implicate Order, Algebras and the Spinor*, Found. Phys. **10**, 7-31 (1980).

Frescura F.A.M. and Hiley B. J., *On the Geometric Interpretation of the Pauli Spinor*, Am. J. Phys., **49**, 152-157 (1981).

Frescura F.A.M. and Hiley B. J., *Algebras, Quantum Theory and Pre-Space*, Revista Brasileria de Fisica, Vol. Especial Os 70 anos de Mario Schonberg, 49-86 (1984).

Frescura F.A.M. and Hiley B. J., *Some Spinor implications unfolded*, in *Quantum Implications*, ed Hiley B.J. and Peat F.D., 278-288, (1987).

Hestenes D. and Sobczyk G., *Clifford algebras to geometric calculus*, Reidel, 1984.

Lounesto P., *Spinor valued regular functions in hypercomplex analysis*, Report-HTKK-MAT-A154, Helsinki University of Technology, 1979.

Porteous I. R., *Topological geometry*, van Nostrand Reinhold, 1969.

Riesz M., *Clifford numbers and spinors*, Lecture Series 38, Institute for Physical Sciences and Technology, University of Maryland, 1958.

Sauter F., *The solution of the Dirac equation without the need for a special representation of the Dirac algebra*, in German, Zeit. Fur Phys., **63**, 803 (1930).

SU(5) and Grand Unification

Peter Rowlands* and J. P. Cullerne†

*IQ Group and Science Communication Unit, Department of Physics, University of Liverpool, Oliver Lodge Laboratory, Oxford Street, P.O. Box 147, Liverpool, L69 3BX, UK. E-mail prowl@hep.ph.liv.ac.uk and prowl@csc.liv.uk

†IQ Group, Department of Computer Science, University of Liverpool, Chadwick Laboratory, Peach Street, Liverpool, L69 72F, UK. E-mail john@jpcullerne.demon.co.uk

On the basis of the representations outlined in the previous paper, the significant features of the Standard Model are reproduced from basic principles, together with their unification in an overall *SU*(5) structure. We also establish a direct correspondence between fermion and boson wavefunctions and particle charge states associated with the same quaternion operators. The model leads to a new value for the weak mixing angle which suggests Grand Unification at the Planck mass.

SYMMETRIES IN THE MATRIX REPRESENTATION

For a given baryon or meson, 27 orientations of charge structure exist spanning across 5 representations (A, B, C, D and E). These orientations are equivalent and derive from arbitrariness in colour labelling of individual charge components. A study of the transformations between orientations will reveal the symmetry that spans this 27 dimensional space. What, then, is the mechanism by which the full symmetry becomes broken into what is observed today?

The charge accommodation rules tell us that the E representation cannot be valid when the three types of charge (e, s, w) have separate conservation laws as they seem to at present. When E becomes invalid the full 27 dimensional space is reduced to 24 and therefore some transformations between orientations become invalid. An analysis of what remains valid under these circumstances will elucidate the observed broken symmetry.

To facilitate our discussion, let us consider the matrix below:

	$\overline{A}$	$\overline{B}$	$\overline{C}$	$\overline{D}$	$\overline{E}$
A	*	*	*		
B	*	*	*		
C	*	*	*		
D					
E					

(1)

The matrix in (1) represents all the transformations within and between the representations A, B, C, D and E. The region highlighted with stars represents just the transformations within and between A, B and C. An inspection of charge structure matrices from baryon or meson states within these three representations reveals symmetry involving the transformation of strong charge orientation:

A representation

$$\begin{matrix} 0 & 0 & -j \\ i & 0 & 0 \\ k & 0 & 0 \end{matrix}$$

B representation

$$\begin{matrix} 0 & 0 & -j \\ 0 & 0 & i \\ k & 0 & 0 \end{matrix}$$

C representation

$$\begin{matrix} 0 & 0 & -j \\ 0 & i & 0 \\ k & 0 & 0 \end{matrix}$$

(2)

(representations for a d quark)
The charts above show the relative positions of the three components of charge within the three representations. Transformations between A, B and C may be linearly combined into the following set of symmetry generators:

$$|1\rangle = (B\bar{A} + A\bar{B})/\sqrt{2}, |2\rangle = -i(B\bar{A} - A\bar{B})/\sqrt{2}, |3\rangle = (B\bar{B} - A\bar{A})/\sqrt{2},$$
$$|4\rangle = (B\bar{C} + C\bar{B})/\sqrt{2}, |5\rangle = -i(B\bar{C} - C\bar{B})/\sqrt{2}, |6\rangle = (A\bar{C} + C\bar{A})/\sqrt{2},$$
$$|7\rangle = -i(A\bar{C} - C\bar{A})/\sqrt{2}, |8\rangle = (A\bar{A} + B\bar{B} - 2C\bar{C})/\sqrt{6}, |9\rangle = (A\bar{A} + B\bar{B} + C\bar{C})/\sqrt{3}.$$

(3)

The representation with the line over it should be read as the representation leaving the interaction vertex. The reason for choosing this particular set of linear combinations becomes clear when we reintroduce the colour labels to the charts in (2). The system of operators in (3) becomes:

$$|1\rangle = (r\bar{b} + b\bar{r})/\sqrt{2}, |2\rangle = -i(r\bar{b} - b\bar{r})/\sqrt{2}, |3\rangle = (r\bar{r} - b\bar{b})/\sqrt{2},$$
$$|4\rangle = (r\bar{g} + g\bar{r})/\sqrt{2}, |5\rangle = -i(r\bar{g} - g\bar{r})/\sqrt{2}, |6\rangle = (b\bar{g} + g\bar{b})/\sqrt{2},$$
$$|7\rangle = -i(b\bar{g} - g\bar{b})/\sqrt{2}, |8\rangle = (r\bar{r} + b\bar{b} - 2g\bar{g})/\sqrt{6}, |9\rangle = (r\bar{r} + b\bar{b} + g\bar{g})/\sqrt{3}.$$

(4)

The labels *r*, *b* and *g* correspond respectively to red, blue and green. Lower case letters are to be used from now on to avoid confusion with letters labelling representations.

In terms of *SU*(3) symmetry (on which QCD is based), these nine states constitute a colour octet (transformations 1 to 8) and a colour singlet (transformation 9). The arbitrariness of labelling *r*, *b* and *g* that led to 6 orientations within a representation now appears as arbitrariness in the way we define colour in QCD.

If the singlet gluon existed, it would be as common and conspicuous as the photon. It cannot actually be the photon as it would couple to all baryons with the same strength, not (as the photon does) in proportion to their charge. Since the mass and baryon number are approximately proportional in bulk matter, such a force would, in fact, look very much like a contribution to gravity [Fischbach et al., 1986 a, b].

Charge accommodation rules require that all naturally occurring particles be colour singlets, and this explains why the octet combinations can never appear as free particles. The 8 transformations between A, B and C are therefore the eight gluon generators of the strong $SU(3)$ (the λ_k in (13) of the preceding paper).

THE ELECTROWEAK INTERACTION MECHANISM

Lepton structures are considered in Rowlands and Cullerne [1999]. An important additional point is that the colours other than red in D (and E) have structures that look like those of the left-handed antiparticles, including the photonic zero-charge structures occupying the places intended for the left-handed antineutrinos (while the A, B, C structures seem to correspond to the right-handed states).

We can see that the representations for the three generations of quarks collapse to three generations of leptons when there is no strong charge to accommodate. These generations still correspond to the three ways of accommodating the weak charge w; that is, $+w$ is indistinguishable from $-zw$, where z takes on the value of -1 for two reasons: 1) P is violated T is not and 2) T is violated and P is not. The representations for leptons in the tables are identical to those for quarks with the omission of strong charge.

The representation above, of course, concerns the D representation alone, because we have assumed that the lepton is the product of a weak transition between the A, B, C representations and the D representation. One should notice however, that in the absence of strong charge, A, B and C collapse to the same expression, which is different to D (and E).

Let us now analyse transformations from A, B or C into D or E. The two subspaces differ by the relative positions of the weak and electric charges.

A, B or C

$$\begin{matrix} 0 & 0 & -j \\ * & * & * \\ k & 0 & 0 \end{matrix}$$

D or E

$$\begin{matrix} 0 & 0 & -j \\ * & * & * \\ 0 & 0 & k \end{matrix} \qquad (5)$$

That is, either electric or weak charges are on different colours (A, B and C) or they are on the same colour (D and E).

Now let us apply the condition that the E representation is no longer valid. The only valid transitions are those within the $SU(3)$ subspace (A, B and C) or those that transform between that subspace and the D representation. Let the $SU(3)$ subspace be represented by the letter S. Then,

$$\begin{array}{ccc} & \overline{S} & \overline{D} \\ S & * & * \\ D & * & * \end{array} \tag{6}$$

is a matrix that represents all the transformations between the subspace S and the D representation. The symmetry transformations in (3) between S and D correspond to an interaction involving only the electric (j) and weak (k) charges. Having factored out the $SU(3)$ in the strong, we see that there are still two degrees of freedom which correspond to the two charge orientations in (5). However, whichever way we look at it, the two possible states may be taken up by one of the charges but the other charge can only take up one of the states. Below is a diagram that illustrates this point more clearly:

$$\begin{array}{cccc} \alpha & \beta & \alpha & \beta \\ -j & 0 & 0 & -j \\ 0 & k & 0 & k \end{array} \tag{7}$$

In (7) we see that the symmetry transformation between the two configurations may be written as though the k stays in the same phase (β) always and $-j$ moves between the phases ($\alpha \leftrightarrow \beta$). We could of course have done this with $-j$ staying in the same phase and k moving. However, if we try to analyse the origin of this problem we see that the whole business occurs because of the way in which we were forced to accommodate the weak charge (k). This we believe is the origin of the parity violation that occurs because of the way k is accommodated. For us, the weak charge only exists in one of the phases in (7); the two discrete phases are therefore states of helicity [Itzykson and Zuber, 1980].

We may write this system of generators from (6) as the following linear combinations:

$$\begin{aligned} |\sigma_x\rangle &= (S\overline{D} + D\overline{S})/\sqrt{2}, \\ |\sigma_y\rangle &= -i(S\overline{D} - D\overline{S})/\sqrt{2}, \\ |\sigma_z\rangle &= (S\overline{S} - D\overline{D})/\sqrt{2}, \\ |0\rangle &= (S\overline{S} + D\overline{D})/\sqrt{2}. \end{aligned} \tag{8}$$

Let us take a look at the meaning of these various generators in terms of the actual operations on the charge structure. The construction of the symmetry transformations

into (25) has the purpose of giving it a structure reminiscent of $SU(2)$. From (8) we see that:

$$D\overline{S} = \tfrac{1}{\sqrt{2}}\left(|\sigma_x\rangle - i|\sigma_y\rangle\right) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$
$$S\overline{D} = \tfrac{1}{\sqrt{2}}\left(|\sigma_x\rangle + i|\sigma_y\rangle\right) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \tag{9}$$

The matrices in (9) are in the basis

$$|S\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$|D\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{10}$$

The basis above is only a mathematical representation of the charge orientations S and D. Pictorially, (9) may be expressed as:

$$D\overline{S} = \begin{matrix} \alpha & \beta \\ 0 & -j \\ 0 & k \end{matrix} \mapsto \begin{matrix} \alpha & \beta \\ -j & 0 \\ 0 & k \end{matrix}$$

$$S\overline{D} = \begin{matrix} \alpha & \beta \\ -j & 0 \\ 0 & k \end{matrix} \mapsto \begin{matrix} \alpha & \beta \\ 0 & -j \\ 0 & k \end{matrix} \tag{11}$$

Inspecting the two helicities, we see that the β helicity states undergo $-j + k \leftrightarrow k$; that is, the symmetry transformation couples these two states of charge configuration. However, the α helicity states do not seem to have a charge structure to couple to ($-j \leftrightarrow 0$). The quaternion configurations are telling us that there is no coupling for the α helicity states between S and D; i.e., this transformation is helicity sensitive [Itzykson and Zuber, 1980].

The same may be done for

$$S\overline{S} = \begin{matrix} \alpha & \beta \\ -j & 0 \\ 0 & k \end{matrix} \mapsto \begin{matrix} \alpha & \beta \\ -j & 0 \\ 0 & k \end{matrix}$$

$$D\overline{D} = \begin{matrix} \alpha & \beta \\ 0 & -j \\ 0 & k \end{matrix} \mapsto \begin{matrix} \alpha & \beta \\ 0 & -j \\ 0 & k \end{matrix} \tag{12}$$

Here we see that in the top transformation a coupling in the α helicity state is allowed between $-j \leftrightarrow -j$. However, after that only β helicity states have structured couplings, $k \leftrightarrow k$ in the top, and $-j + k \leftrightarrow -j + k$ in the bottom transformation.

The above results are a statement of the standard model with the following correspondences between our model and that of GWS:

- Transitions between D and S:

$$D\overline{S} \leftrightarrow W^{+}\left(j_{\mu}^{+}\right), S\overline{D} \leftrightarrow W^{-}\left(j_{\mu}^{-}\right)$$

- Third component of weak isospin current coupling only left handed (β phase) charge structure:

$$\left|\sigma_{z}\right\rangle = (S\overline{S} - D\overline{D})/\sqrt{2}.$$

- Weak hypercharge current is the invariant construct coupling both left and right-handed (both α and β phases):

$$\left|0\right\rangle = (S\overline{S} + D\overline{D})/\sqrt{2}.$$

The four states in (8) constitute the generators of $SU(2)_L \otimes U(1)$. The L subscript reminds us that the $SU(2)$ involves only left-handed (β) states. Since these transformations correspond to relative positions of weak and electric charge in charge structure matrices, one cannot really make a distinction between changes in electric or weak orientations. Therefore changes in orientation due to generators (8) can only really be called electro-weak.

Globally then, the transformations between representations A, B, C and D, are generated by the group $SU(3) \otimes SU(2)_L \otimes U(1)$. This of course is the symmetry group of the standard model, with the $SU(3)$ corresponding to QCD and the $SU(2)_L \otimes U(1)$ corresponding to the GWS electroweak unification.

THE PRODUCTION OF LEPTONS

We have seen that the electroweak interaction is represented in our theory as the transformations between the subspace S and the representation D (*see* matrix (6)). The charts in (5) reduce these transformations to essentially (7); that is, the placing of the weak and electric charges in the same colour or different colours. Although the strong charge symmetry has effectively been factored out, we still have these two degrees of freedom (helicity).

Thus far we have considered only transformations between representations. Close inspection of the quark tables reveals that symmetry with exactly the same form as (4) is apparent between u $\leftrightarrow$ d, c $\leftrightarrow$ s, and t $\leftrightarrow$ b in the D representation. For example,

d quark in D

$$\begin{matrix} 0 & 0 & -j \\ * & * & * \\ 0 & 0 & k \end{matrix}$$

u antiquark in D

$$\begin{matrix} -j & -j & 0 \\ * & * & * \\ 0 & 0 & -k \end{matrix} \tag{13}$$

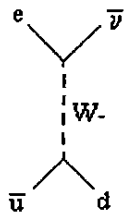
The charge accommodation rules required k to be unipolar in order to fit the tables. Therefore through an interaction identical to $S \leftrightarrow$ D, the charge structures of the d quark and u antiquark are interchangeable as a symmetry operation.

When the strong charge is present, these transitions between a d quark and a u antiquark can only occur within a π^- meson charge combination, since this would be the only way to accommodate the components of the two unit charges. However, in the decay of the π^- the strong charge and anticharge present presumably annihilate. The need for charge accommodation rules is relaxed and the two unit charges become 'free'. By inspecting the tables in the absence of the strong charge we can see the charge structures of the two free charges.

$$\begin{matrix} \alpha & \beta \\ 0 & -j \\ \underbrace{0 \quad k}_{d} \end{matrix} \qquad \begin{matrix} \alpha & \beta \\ -j & 0 \\ \underbrace{0 \quad -k}_{\bar{u}} \end{matrix} \tag{14}$$

The two columns represent the helicity phases mentioned above. The first column represents a free charge state made up of only one electric charge $-j$. The second column represents two free charges, one of structure $-j + k$ and the other of structure $-k$ alone.

When the d quark and u antiquark were in the π^- combination, their spins were paired to form a spin singlet state. When they separate as free particles, spin must remain conserved so the resulting particles must have opposite spin states but essentially the same helicity states. What the second column is telling us in (10), is that one possibility for the free charges is in states $-j + k$ and $-k$, which in principle have the same helicity. However, the first column essentially tells us that the other spin pairing possibility must be accounted for but that there is only one charge, $-j$, that can take this state up.



This is exactly what is observed in the π^- decay. The pion has a spin 0, so the electron and antineutrino must emerge with opposite spins, and hence equal helicities. The antineutrino ($-k$) is always right-handed, so the electron ($-j + k$) must be right-handed as well. However, the electron is not massless so must admit the other helicity state ($-j$), whereas the antineutrino cannot take up its opposite helicity state ($\therefore 0 \cdot k$).

According to GWS, if the electron were massless, then (like the neutrino) it would only exist as a left-handed particle. More precisely, the $1 - \gamma^5$ in the weak vertex factor would couple only to left-handed electrons, just as it couples only to left-handed neutrinos. If the electron were massless, the decay above would not occur at all. Since the electron is not massless but does have a very small mass, one would expect this decay to be heavily suppressed. This is indeed what is observed.

The violation of parity is evident. The requirement that the weak transition ($S \leftrightarrow$ D type transformations) treat $\pm k$ as if they were the same, leads to a violation of parity in the free charges. This violation of parity is not 'visible' in the π^- combination, but becomes conspicuous as soon as the electron and antineutrino free charges escape the bound state (there are *no* left-handed antineutrino states).

The fact that the weak interaction cannot tell the difference between $\pm k$ has other important consequences. Transitions u $\leftrightarrow$ s, c $\leftrightarrow$ d etc must be allowed to occur by exactly the same mechanism ($S \leftrightarrow$ D type transformations). The global structure of the quaternion model therefore requires a weak mixing of the Kobayashi – Maskawa type [1972], where the weak interaction may be thought of as a transformation between two subspaces of quark states (d, s and b) $\leftrightarrow$ (u, c and t). The standard model represents these transitions by offsetting the quark generations for the weak interaction:

$$\begin{pmatrix} u \\ d' \end{pmatrix} \begin{pmatrix} c \\ s' \end{pmatrix} \begin{pmatrix} t \\ b' \end{pmatrix} \tag{15}$$

where the quarks d', s' and b' are related to the physical quark states d, s and b by *the Kobayashi – Maskawa matrix.*

ELECTRO-WEAK MIXING

The quaternion description has made various assertions about the nature of electric and weak interactions. In this section we compare the GWS model with the quaternion model, in the hope that we can demonstrate the need for further work in this field.

The GWS model asserts that the three weak isospin currents couple, with strength g_w, to a weak isospin triplet of intermediate vector bosons, **W**, whereas the weak hypercharge current couples with strength $g'/2$ to an isosinglet intermediate vector boson, ***B***:

$$-i\left[g_w J_\mu \cdot W^\mu + \frac{g'}{2} j_\mu^Y B^\mu\right] \tag{16}$$

Within this structure is contained all of electrodynamics and all of the weak interactions.
The dot product can be written out explicitly:

$$J_\mu \cdot W^\mu = j_\mu^1 W^{\mu 1} + j_\mu^2 W^{\mu 2} + j_\mu^3 W^{\mu 3} \tag{17}$$

or, in terms of the charged currents, $j^{\pm}{}_\mu = j^1{}_\mu \pm i\, j^2{}_\mu$:

$$J_\mu \cdot W^\mu = \tfrac{1}{\sqrt{2}} j_\mu^+ W^{\mu +} + \tfrac{1}{\sqrt{2}} j_\mu^- W^{\mu -} + j_\mu^3 W^{\mu 3} \tag{18}$$

where

$$W_\mu^{\pm} \equiv (1/\sqrt{2})(W_\mu^1 \mp i W_\mu^2) \tag{19}$$

are the wave functions representing the $W^{\pm}$ particles.

The couplings are easily read off, from the coefficients of the charged W particles. For example, $e^- \rightarrow \nu_e + W^-$ is described in GWS as

$$j_\mu^- = \bar{\nu}_L \gamma_\mu e_L = \bar{\nu}\gamma_\mu \left[(1-\gamma^5)/2\right] e, \tag{20}$$

giving a term

$$-ig_w(1/\sqrt{2}) j_\mu^- W^{\mu -} = -\frac{ig_w}{2\sqrt{2}}\left[\bar{\nu}\gamma_\mu(1-\gamma^5)e\right] W^{\mu -} \tag{21}$$

The vertex factor is

$$-\frac{ig_w}{2\sqrt{2}}\left[\gamma_\mu(1-\gamma^5)\right] \tag{22}$$

In terms of global charge structure, the quaternion description exhibits exactly this symmetry transformation as

$$S\bar{D} = \tfrac{1}{\sqrt{2}}\left(|\sigma_x\rangle + i|\sigma_y\rangle\right) \tag{23}$$

in the *S*, D basis. Here we are dealing only with the symmetries within the quaternion description. Vertex probabilities within the quaternion description are left to later papers on this subject.

The underlying $SU(2)_L \otimes U(1)$ is mixed in GWS theory; that is, the two neutral *states*, W^3 and B, mix, producing one massless linear combination (the photon) and an orthogonal massive combination (the Z^o):

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W$$
$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \qquad (24)$$

Our quaternion description necessarily exhibits this kind of mixing as is evident in (12). Here we have evidence of a naturally occurring GWS structure, which arises from the quaternion description.

MASS AND SYMMETRY BREAKING

We have shown that the rules for the Standard Model follow immediately from those for charge accommodation, including the left-handed bias of the weak interaction for neutrinos and the implied zero mass of the neutrino. It is widely believed that oscillations observed between muon and tau neutrinos would necessarily imply a neutrino mass, in contradiction to the usual assumptions of the Standard Model [Fukuda et al., 1998]. However, in the present system, the muon and tau neutrinos are distinguished only by their respective violations of parity and time reversal symmetry. Such a distinction, however, does not seem to have a physically observable basis, and it would certainly be impossible without a mass scale; so masslessness of the neutrino (a particle which can only be detected by its interactions) would lead to a distinguishability determined only by the mass scales of the interacting partners.

It is also assumed in many contexts that parity violation occurs in preference to time-reversal symmetry, but nature, in fact, knows no such preference. It is impossible to tell, in principle, which violation has actually occurred in any given circumstance. It is simply a matter of convention and convenience that we take parity violation as occurring in the first instance, and time-reversal asymmetry in the second. The real distinction is between one symmetry violation and two. Systems that are assumed to violate time-reversal symmetry, such as the K^0 meson, have already been found to violate parity, and time-reversal asymmetry is only invoked because it can be no longer avoided.

We should not, therefore, expect a difference in mass-scale between *individual* P or T violations, but rather a difference between the breaking of one symmetry and the breaking of two. Some evidence of this may perhaps be found in the Cabibbo-Kobayashi-Maskawa mixing between the respective quark generations represented by d, s and b, where the mixing ratio between the first two generations (approximately tan θ_C) appears to be of the same order as that between the electric and symmetry-breaking weak interactions ($\sin^2\theta_W$), while that between second and third generations, with two symmetry-violations, takes this to a second order. The increasing mass-scale

for the quarks in the higher generations must be related in some way to this symmetry-breaking.

THE CHARGE-MASS MAPPING

Here we are using a 5-'dimensional' quantity to combine the effects of 3-dimensional conserved and nonconserved parameters (the term p having 3 dimensions, although only one is normally defined), and, just as the 4-vector representation requires complex algebra, its extension is only possible through a noncommutative (specifically, quaternion) algebra.

For a quantum system with stationary states, in which E and $\mathbf{p}$ are *fixed*, as well as m, we describe the fixity of E, $\mathbf{p}$ and m against the variation of the space and time coordinates in the exponential part of the Dirac wavefunction, $e^{-i(Et - \mathbf{p}.\mathbf{r})}$. Thus, in quantizing the energy-momentum conservation, the Dirac equation effectively structures mass (or energy-momentum-mass) as a 3-dimensional conserved parameter, like charge (with one of the 'dimensions' being itself dimensional).

Remarkably, the structures for particles generated by applying the quaternion operators i, j, k to the charges s, e, w seem to be essentially identical in form to those produced by the Dirac wavefunction in applying i, j, k to the 3-'dimensional' 'mass-operator' (iE, $\mathbf{p}$, m). This is because, with the theory of charge structures, we are effectively adding the same degree of 3-component variation to what is in principle a fixed quantity. The combination in each case requires a 5-dimensional term which automatically breaks the symmetry that would otherwise apply to the coefficients of the quaternion operators. The 5-dimensionality also suggests a possible link with the $SU(5)$ symmetry which is the simplest possible candidate for Grand Unification.

In principle, the Dirac equation introduces quaternion operators for E, p, m because of the need to remove cross-terms in the combination of squared quantities (in the same way as the individual conservation laws for e, s, w remove the cross-terms when these quantities are squared). In each case, the application of rotation symmetric quaternion operators to rotation asymmetric fixed quantities requires the application of the same symmetry-breaking (chiral) mathematical structure.

$SU(5)$ SYMMETRY

There are good reasons for believing that the grand unified gauge group is $SU(5)$ [Georgi and Glashow, 1974, Georgi, Quinn and Weinberg, 1974, Masiero, 1984, Rowlands and Cullerne, 1998]. This is a group of rank four and so has four invariant Casimir operators, which are nonlinear functions of the generators, and clearly correspond to the four fundamentally conserved quantities w, s, e, M. $SU(5)$ appears to be the underlying structure of the γ-matrix algebra (especially in the form of the 4-vector-quaternion combination), and its separate allocations to w, s, e and E, $\mathbf{p}$, m, in addition to the five possible quark representations A, B, C, D, E . We have good evidence for the following mappings of these structures onto each other:

E	p_x	p_y	p_z	m
w	s_G	s_R	s_B	e
D	A	B	C	E
ik	ii	ij	ik	j
γ^0	γ^1	γ^2	γ^3	γ^5

The mapping of the strong terms is always exact, but the electroweak terms are so closely linked physically that transposition to equivalent representations may be necessary to reflect the physical manifestations of these interactions.

THE *SU*(5) GENERATORS

The *SU*(5) structure is demonstrated by the existence of 24 generators, however these are grouped. Between representations we have the transitions A-B-C as pure strong interactions, with generators $A\bar{A}$, $A\bar{B}$, $A\bar{C}$, $B\bar{A}$, $B\bar{B}$, $B\bar{C}$, $C\bar{A}$, $C\bar{B}$, standing for the gluons, with $C\bar{C}$ arbitrarily selected as the excluded singlet state. The D-E transitions represent the pure electroweak forces, with $E\bar{D}$, $D\bar{E}$, $D\bar{D}$, $E\bar{E}$ representing the four electroweak bosons W^+, W^-, Z^0, γ via linear combinations. (In the strong interaction, we imagine e, w as fixed and s being transferred between quarks; in the electroweak interaction s is fixed and e, w transferred, the boson exchange producing the required current.) The twelve transitions A, B, C-E and A, B, C-D represent the X and Y bosons, carrying combinations of strong and electroweak forces. It is noticeable that the transitions form three groups of eight – in effect three *SU*(3) groupings, two of which are broken. The exact one is formed from transitions between A, B, C, and the broken ones from the respective transitions to and from D and E.

In terms of charges, we could represent the gluons by terms like $s_G \bar{s}_G$, $s_G \bar{s}_B$, $s_G \bar{s}_R$, $s_R \bar{s}_G$, $s_R \bar{s}_B$, $s_R \bar{s}_R$, $s_B \bar{s}_G$, $s_B \bar{s}_R$, excluding one combination (here, arbitrarily, $s_B \bar{s}_B$) as a singlet. W^+, W^-, Z^0, γ would be required to accommodate the four states $w\bar{e}$, $e\bar{w}$, $w\bar{w}$, $e\bar{e}$ by a linear combination, the actual representations being $\bar{e}\,w\,\bar{e}$, $w\,\bar{w}\,e$, $w\,\bar{w}$, $e\,\bar{e}$. The remaining twelve gauge bosons are then required to accommodate the states like $s_G\bar{w}$, $s_R\bar{w}$, $s_B\bar{w}$, $w\bar{s}_G$, $w\bar{s}_R$, $w\bar{s}_B$, $s_G\bar{e}$, $s_R\bar{e}$, $s_B\bar{e}$, $e\bar{s}_G$, $e\bar{s}_R$, $e\bar{s}_B$, though they are probably combinations of them.

With respect to the wavefunction terms, E, $\mathbf{p}$, m, the gluons become combinations of the form $p_x\bar{p}_x$, $p_x\bar{p}_y$, $p_x\bar{p}_z$, $p_y\bar{p}_x$, $p_y\bar{p}_y$, $p_y\bar{p}_z$, $p_z\bar{p}_x$, $p_z\bar{p}_y$, excluding the singlet, arbitrarily taken here as $p_z\bar{p}_z$. The physical significance of these representations is in the *actual* transfer of momentum state with transfer of strong charge, as in conventional QCD. The electroweak bosons W^+, W^-, Z^0, γ are represented by $E\bar{m}$, $m\bar{E}$, $m\bar{m}$, $E\bar{E}$, or combinations of them. It is apparent here why there must be three weak boson states involving mass transfer, while the gluons, involving only p transitions, remain massless. The remaining twelve boson states are of the form $p_x\bar{E}$, $p_y\bar{E}$, $p_z\bar{E}$, $E\bar{p}_x$, $E\bar{p}_y$, $E\bar{p}_z$, $p_x\bar{m}$, $p_y\bar{m}$, $p_z\bar{m}$, $m\bar{p}_x$, $m\bar{p}_y$, $m\bar{p}_z$, or combinations of them. It is the precise need for massless momentum transfer in the case of gluons and massive transfer in the case of the weak gauge bosons which demonstrates the exactness and physical meaning of

the $s \rightarrow$ p and $e \rightarrow m$ mappings; the charge structures generate wavefunctions (see section 28) because of the physical requirements that the interactions produce for their gauge bosons. The exactness of the $w \rightarrow E$ mapping will be demonstrated by its production of a filled vacuum for the weak interaction (see the section on the Higgs Mechanism).

In simplified form (leaving out any mixing and singlet states) the $SU(5)$ generators may be represented by the following tables:

	$\bar{A}$	$\bar{B}$	$\bar{C}$	$\bar{D}$	$\bar{E}$
A					
B		Gluons		Y	X
C					
D		Y		Z^0, γ	W^-
E		X		W^+	Z^0, γ

	$\bar{s}_G$	$\bar{s}_B$	$\bar{s}_R$	$\bar{w}$	$\bar{e}$
s_G					
s_B		Gluons		Y	X
s_R					
w		Y		Z^0, γ	W^-
e		X		W^+	Z^0, γ

	$\bar{p}_x$	$\bar{p}_y$	$\bar{p}_z$	$\bar{E}$	$\bar{m}$
p_x					
p_y		Gluons		Y	X
p_z					
E		Y		Z^0, γ	W^-
m		X		W^+	Z^0, γ

THE ELECTROWEAK FORCES

An $SU(5)$ structure requires 24 generators. The $SU(5)$ symmetry, however, is broken in practice, for quarks, because, below grand unification energies, the E representation is forbidden. Strictly speaking, also, the $E\bar{E}$ transition is forbidden. The electromagnetic component of the electroweak interaction is reduced to a phase term ($U(1)$). For quarks, also, as we have seen, there is no pure weak or electroweak transition across representations. The transitions between A / B / C and D are, in effect, strong-weak transitions, reflecting the fact that *the very existence of quarks* requires them to be always in the process of a strong interaction; and so, although we can have a weak interaction between quarks, as long as they remain quarks this interaction will always be manifested as a combination of strong and weak. Although the diagrammatic representation suggests that the weak transition must be from C to D, the representation of quarks with these particular charge structures as being in representations A, B or C must be as arbitrary an assignment as the colour labels Blue, Green or Red. They are always the result of superpositions. If, however, we remove the strong charge by converting the quarks into leptons, via quark-antiquark interactions, then A, B, C become identical, as do D and E, and the transition of the lepton derived from A / B / C to that derived from D becomes a pure weak (or, including the phase term from E, pure electroweak) interaction.

PROTON DECAY

According to our reasoning, such decays as that of the proton through an X-particle to neutral pion plus positron are forbidden by rotation asymmetry (or separate e, s, w conservation) [Rowlands and Cullerne, 1999], and they also neglect the fact that weak interactions are subject to the requirements of what may (for want of a better term) be called 'weak colour' – the necessary creation of fermion-antifermion states, or their equivalent. Also, creating or removing the vector state involved in strong-electroweak transitions necessarily always involves combining + and – parts of s and $\mathbf{p}$, so that charges can be added or subtracted and vectors in wavefunctions may be removed via, or created from, scalar products.

The unit charges in the D / E representations cannot lead to bound quark states like those in A, B, C, and are really only to be associated with lepton states. A strong-electroweak transition from A, B, C to D / E via X or Y can only take place through the removal of the s component, and such interactions can only be imagined as happening between quark and antiquark or baryon and antibaryon. The X-particles are thus assumed to be created at the vertex of a fermion-antifermion interaction and are therefore presumed to have the characteristics of a Bose-Einstein condensate (fermion-fermion or antifermion-antifermion combination), rather than those of a regular boson. CP violation in the weak decay of such states may be expected as a result of their creation of a boson-like particle with two fundamental units of weak charge, and it may, in fact, be expected as a general consequence of the weak decay of Bose-Einstein condensate states.

CHARGE STRUCTURES AND WAVEFUNCTIONS

The use of identical quaternion operators for the charges w, s, e and the wavefunction terms E, $\mathbf{p}$, m suggests that charge structures might map directly on to wavefunctions. A significant fact to take into consideration here is that, for composite particles, component wavefunctions are multiplied while charges are added. In constructing wavefunctions from charge structures, we also need to be aware that kE and kw are always positive for particles and negative for antiparticles, that is is positive for particles and negative for antiparticles, whereas the sign of $ii\mathbf{p}$ depends on the direction of spin, and that the sign of ijm is always positive irrespective of the sign of je.

If we assume that the E, $\mathbf{p}$, m components of a fermion wavefunction reflect the presence or absence of the charges s, e, w, then quarks, which have a charge structure with nonzero expectation values for all three charges, would have a wavefunction of the form

$$(kE + ii\,\mathbf{p} + ij\,m)\, e^{-i(Et - \mathbf{p}.\mathbf{r})},$$

reflecting a charge structure of the form

$$k\,w + i\,s + j\,e\,.$$

A meson state now combines this with an antistate reversing the signs of w and s. At the same time we reverse that of E, so that the operator part of the wavefunction becomes

$$(kE + ii\,\mathbf{p} + ij\,m)\,(-kE + ii\,\mathbf{p} + ij\,m),$$

leading to a scalar term on summation over all four possible solutions. We assume here, of course, that a meson is a single-particle state and that the values associated with E, $\mathbf{p}$ and m are averaged-out or composite values for the particle as a whole. We also assume that the momentum term $\mathbf{p}$ represents, in the first instance, the 'internal momentum' (spin) maintaining the structure of the particle through the colour force, though $\mathbf{p}$ and E are both modified when the system gains external momentum through the loss of mass or presence of an external field (in a manner parallel to the spin-orbit coupling of atomic theory).

If we now apply the rule of *addition* of components to the equivalent charge structure terms, we find that

$$(k\,w + i\,s + j\,e) + (-k\,w - i\,s + j\,e)$$

leads immediately to the global elimination of the w and s terms. The existence of a neutral boson with electric charge $(j\,e - j\,e)$, however, would not eliminate the mass term ijm, which is positive regardless of the sign of electric charge. (The e term is the only one whose sign in the charge structure does not determine whether a particle is a fermion or antifermion.)

It is particularly significant that all fermion wavefunctions have a term involving kE, corresponding to the weak charge kw, which is their defining characteristic. This, of course, merely reflects the physical fact that any particle, whose existence is defined

by mass or momentum, must necessarily also have energy. Conversely, a particle with a non-zero quaternion energy operator must also have momentum or rest mass, with a different quaternion operator, ensuring a quaternionic form for the total wavefunction. Bosons, however, defined by the absence of a global kw term, have a kE term combined with $-kE$ because they are defined as a fermion-antifermion combination.

Gluons are massless, and the energy E is entirely determined by the momentum term, corresponding to the strong charge s. The masslessness corresponds also to an absence of electric charge e. The gluon wavefunction operator, then, may be expected to have the form

$$(kE + ii\,\mathbf{p})\,(-kE + ii\,\mathbf{p})\,,$$

corresponding to the charge structure

$$(k\,w + i\,s) + (-k\,w - i\,s)\,.$$

The electroweak bosons W^+, W^-, Z^0, γ may be supposed to be combinations of leptons and antileptons, with no internal momentum, and so, if assumed stationary, become

$$(kE + ij\,m)\,(-kE + ij\,m)\,,$$

corresponding to

$$\begin{aligned}&(k\,w + j\,e) + (-k\,w + j\,e)\\&(k\,w - j\,e) + (-k\,w - j\,e)\\&(k\,w + j\,e) + (-k\,w - j\,e)\end{aligned}$$

for W^+, W^- and Z^0. External momentum would, of course, change the wavefunction to

$$(kE + ii\,\mathbf{p} + ij\,m)\,(-kE + ii\,\mathbf{p} + ij\,m)\,.$$

γ would have no existence without external momentum as E cannot be defined without m or $\mathbf{p}$, so γ becomes

$$(kE + ii\,\mathbf{p})\,(-kE + ii\,\mathbf{p})\,,$$

with $\mathbf{p}$ purely external.

Since the strong charge is absent, the corresponding wavefunction operator for an electron, without external momentum, is

$$(kE + ij\,m)\,,$$

corresponding to the charge structure

$$(k\,w + j\,e)\,.$$

With external momentum, this becomes

$$(kE + ii\,\mathbf{p} + ij\,m)\,.$$

The neutrino, if assumed massless, would again have no existence without external momentum, but, with this, it has the structure

$$(kE + ii\ \mathbf{p}),$$

corresponding to the charge structure

$$(k\ w).$$

One other wavefunction that is significant is that for the composite baryon state. This must incorporate three fermion (quark) wavefunctions, and itself be a fermion wavefunction. A wavefunction of the form

$$(kE + ij\ m)\ (kE + ij\ m)\ (kE + ii\ \mathbf{p} + ij\ m)$$

would satisfy these requirements, and also account for the fact that the baryon spin term ($ii\ \mathbf{p}$) is not derived from three component quark spins, in addition to establishing the parallelism of the mechanism of the exchange of a single unit of strong charge between three quark states and the exchange of the spin (momentum) term. To relate this to the complete mathematical structure of QCD, we need to interpret $\mathbf{p}$ in terms of the covariant derivative

$$D^{\mu} = \partial^{\mu} + is\lambda.\mathbf{A}^{\mu}\ ,$$

with λ representing the eight gauge bosons or gluons.

We can now represent the baryon wavefunction used to construct a baryon in the accompanying paper using a mapping such as:

BGR	$(kE + ij\ m)\ (kE + ij\ m)\ (kE + ii\ \mathbf{p} + ij\ m)$
$-BRG$	$(kE + ij\ m)\ (kE - ii\ \mathbf{p} + ij\ m)\ (kE + ij\ m)$
GRB	$(kE + ij\ m)\ (kE + ii\ \mathbf{p} + ij\ m)\ (kE + ij\ m)$
$-GBR$	$(kE + ij\ m)\ (kE + ij\ m)\ (kE - ii\ \mathbf{p} + ij\ m)$
RBG	$(kE + ii\ \mathbf{p} + ij\ m)\ (kE + ij\ m)\ (kE + ij\ m)$
$-RGB$	$(kE - ii\ \mathbf{p} + ij\ m)\ (kE + ij\ m)\ (kE + ij\ m)$,

with each term equivalent to $-p^2(kE + ii\ \mathbf{p} + ij\ m)$ or $-p^2(kE - ii\ \mathbf{p} + ij\ m)$.

With the spinor terms included, each if these is represented by a tensor product of three spinors, for example:

$$(kE + ij\ m)\ (kE + ij\ m)\ (kE + ii\ \mathbf{p} + ij\ m)\left(\frac{1}{2}\right)\otimes\left(\frac{1}{2}\right)\otimes\left(\frac{1}{2}\right)$$

where

$$\left(\frac{1}{2}\right)\otimes\left(\frac{1}{2}\right)\otimes\left(\frac{1}{2}\right)=\left(\frac{3}{2}\right)\oplus\left(\frac{1}{2}\right)\oplus\left(\frac{1}{2}\right)$$

So this representation encompasses both spin ½ and spin 3/2 baryon states.

THE ORIGINS OF THE HIGGS MECHANISM

In the Standard Model, the Higgs mechanism is used to supply masses to the fermions and weak gauge bosons. There it is brought in as an ad hoc procedure, but, though the exact values of mass allocation may remain unsolved without detailed mathematical

development, the principle that mass will result from a filled weak vacuum can be explained as a development of other fundamental principles.

In the Dirac theory, fermions are specified by positive energy states (kE) and antifermions by negative energy states ($-kE$). We create a new fermion state by applying the creation operator $\mathbf{a}^\dagger = (1 / 2E)\ (kE + ii\mathbf{p} + ijm)$ to the vacuum state. In principle, the same process should also apply to antifermions, with $\mathbf{a}^\dagger = (1 / 2E)\ (-kE + ii\mathbf{p} + ijm)$. In Dirac's original theory, however, the antifermion energy levels were completely filled and antifermions were created only by the annihilation of a fermion state. We have expressed this, in effect, in section 5, by defining the vacuum state as $(E - ij\mathbf{p} + iim)\ e^{-i(Et\ -\ \mathbf{p}.\mathbf{r})}$, which is the equivalent, up to a post-multiplication factor $-k$, of $(-kE + ii\mathbf{p} + ijm)\ e^{-i(Et\ -\ \mathbf{p}.\mathbf{r})}$. Applying $\mathbf{a}^\dagger$ for an antifermion to the vacuum state as defined here would produce a zero result. That is, a truly filled vacuum would discriminate physically between the fermion and antifermion states represented by the respective energy operators kE and $-kE$, and deny the equality of status required by charge conjugation symmetry.

In principle, the Dirac theory suggests that there is a preponderance of fermions over antifermions because it is possible to have a completely filled antifermion vacuum in the lowest possible energy state, while fermions exist as free particles, presumably reflecting the fact that positive energy is a real physical state, while negative energy is not. Thus, even the lowest energy state would have a positive value, while the existence of free antifermions, as in the real world, would indicate an energy level above the ground state. The expression $(E - ij\mathbf{p} + iim)\ e^{-i(Et\ -\ \mathbf{p}.\mathbf{r})}$ becomes the unique vacuum for this ground state.

Now, the energy operator k is also the weak charge operator, and the weak charge operator is the one that, according to the charge accommodation rules of section 10, violates charge conjugation symmetry, or the symmetry between fermions and antifermions for the weak interaction. The mechanism, according to the standard theory, is a filled vacuum for the weak charge, identical to the filled vacuum for antifermions postulated in the Dirac theory. The fact that charge conjugation is actually *allowed* for the weak interaction, as well as required, may thus be a reflection of the same physics that requires the predominance of fermions over antifermions.

Once we have charge conjugation violation in the weak interaction, but not in the others, we have, in principle, a mechanism for generating rest mass for the interacting particles. Essentially, if a particle has *only* weak charge, like the neutrino, then charge conjugation symmetry will be completely violated, and it will have only one direction of helicity (left-handed, according to the analysis in the accompanying paper). The wavefunction, therefore, has no right-handed component. If other charges are present, however, then the right-handed component cannot be totally suppressed because the symmetry-violation only occurs for weak charge. The introduction of a right-handed component means that the particle must have rest mass, so that its opposite handedness can be observed when it is 'overtaken'. The mass must in some way be a measure of the amount of right-handedness which it may display, and must therefore be related, in some way or other, to one or both of the other two charges. It is expected that a more detailed mathematical analysis of the processes involved will ultimately lead to the spectrum of boson and fermion masses.

PARTICLE MASS SPECTRUM

We are not yet able to generate rest masses for the fundamental fermions, the quarks and leptons, as these must depend on details of the Higgs mechanism which have not yet been worked out. Arguments can be put forward, however, for the origin of the mass-energies of the composite baryons and leptons, based on the numbers of zero units of charge [Rowlands and Cullerne, 1999].

That mass and charge are, in some sense, mutually exclusive components of the vacuum (effectively combining to form an invariant), is implied by standard treatments of the $U(1)$ component of the Weinberg-Salam theory. For example, Aitchison and Hey [1989], writing on the hypercharge value of the Higgs field, state that: 'we do not allow the particle physics vacuum to give an electrically charged field a non-zero value. Thus we require that the component of ϕ with non-zero vacuum value has zero charge.'

It is not really possible to provide a serious calculation of the mass of the Higgs boson until details of the mechanism have been fully worked out, and any attempt at such a calculation must be, at this stage, highly speculative. The value of 182 GeV, suggested in Rowlands and Cullerne [1999], is, however, almost identical to the sum of the masses of all fermion states (six quarks plus six leptons), suggesting that the vacuum energy might be distributed by the Higgs mechanism between these states according to some probability condition for the coupling. It is also virtually identical to the mass of two Z^0 bosons. If the vacuum energy is distributed between the possible fermion states, it is noticeable that the three quark *generations* are separated from each other by a factor of the order of α, effectively the separating factor between strong and electroweak couplings. No obvious relation yet presents itself between u, c, t and d, s, b states, or between the quarks and the leptons.

GRAND UNIFICATION

We have presented evidence that the symmetry between the three nongravitational forces is, indeed, based on an $SU(5)$ group, as Georgi and Glashow have proposed [1974], but this does not necessarily mean that this grouping will represent the final structure at grand unification. Our successful use of the Han-Nambu model suggests that we may have justification in recalculating some of the important constants of Grand Unification theory. One is the weak mixing angle in the GSW $SU(2)_L \otimes U(1)$ theory, calculated from

$$\sin^2\theta_W = \frac{\Sigma\, t_3^{\,2}}{\Sigma\, Q^2}.$$

For the weak component, with only left-handed contributions to weak isospin, from 3 colours of u, 3 colours of d, and the leptons e and ν,

$$\Sigma\, t_3^{\,2} = \frac{1}{4} \times 8 = 2 \ .$$

This should be unchanged for the Han-Nambu model. For the electromagnetic component, however, this is not the case. Intrinsic fractional charges, with both left- and right-handed contributions, would lead to

$$\Sigma Q^2 = 2 \times \left(\frac{4}{9} \times 3 + \frac{1}{9} \times 3 + 1 + 0 \right) = \frac{16}{3} ,$$

from which

$$\sin^2 \theta_W = 0.375 .$$

With intrinsic integral charges, we have

$$\Sigma Q^2 = 2 \times (1 + 1 + 0 + 0 + 0 + 1 + 1 + 0) = 8 ,$$

leading to

$$\sin^2 \theta_W = 0.25 ,$$

which is actually in much better agreement with the experimental value of 0.231.

An interesting thing happens when we apply this value to the renormalization equations for weak and strong couplings used in Grand Unified theories [Georgi and Glashow, 1974, Masiero, 1984, Weinberg, 1996]:

$$\frac{1}{\alpha_2(\mu)} = \frac{1}{\alpha_G} - \frac{5}{6\pi} \ln \frac{M_X^2}{\mu^2}$$

and

$$\frac{1}{\alpha_3(\mu)} = \frac{1}{\alpha_G} - \frac{7}{4\pi} \ln \frac{M_X^2}{\mu^2} ,$$

for, combining these equations with

$$\sin^2 \theta_W = \frac{\alpha(\mu)}{\alpha_2(\mu)} ,$$

gives

$$\sin^2 \theta_W (\mu) = \alpha(\mu) \left(\frac{1}{\alpha_3(\mu)} + \frac{11}{6\pi} \ln \frac{M_X}{\mu} \right).$$

Using typical values for $\mu = M_Z = 91.1867(21)$ GeV, $\alpha(M_Z^2) = 1 / 128$ (or 1/129), $\alpha_3(M_Z^2) = 0.118$ (or 0.12), and $\sin^2 \theta_W = 0.25$, we obtain 2.8×10^{19} GeV for the Grand Unified mass scale (M_X), which is of the order of the Planck mass (1.22×10^{19}). Current models of Grand Unification, based on $\sin^2 \theta_W = 0.375$, have assumed that this is a Grand Unified value which can be renormalized down to something like 0.19-0.21, but this has produced only moderate agreement with experiment [Georgi and Glashow, 1974, Masiero, 1984, Weinberg, 1996]. By our understanding, 0.25 is specifically the value for a *broken* symmetry, produced by asymmetric values of charge, whether or not it is contained within a larger group structure such as $SU(5)$. (It is also, in this theory, but not in minimal $SU(5)$, identical for both quarks and leptons separately considered.)

If we *assume* that M_X is the Planck mass, we obtain α_G (the Grand Unified value for all interactions) = 1 / 52.4, and $\alpha_2(M_Z^2) = 1 / 31.5$, which is, of course, well within the range expected from taking $\sin^2 \theta_W = 0.25$. Though even using the experimental value of $\sin^2 \theta_W = 0.231$ brings us fairly close to the Planck mass at about 5×10^{17} GeV, the fit to the experimental data is even more remarkable in that higher order corrections show that the presence of massive gauge bosons depresses the effective values of both

$1 / \alpha_2$ and $\sin^2\theta_W$ in the energy range M_W - M_z, where they are normally measured, possibly by the 0.02 required, while the application of the two-loop approximation reduces the value of M_X by a small factor (0.64 in the fractional charge model) [Kounnas, 1984, Novikov et al., 1999]. We also obtain unit strength for the strong interaction ($\alpha_3 = 1$) at the energy level of baryonic and mesonic structure (approximately, $\mu = m_\pi$), and an approximate 257 GeV (close to the total mass of the electroweak bosons) for the Higgs vacuum expectation value ($2M_W / g$).

The renormalization equations for weak and strong couplings require no modification as a result of changes in the quark model, but the $U(1)$ electromagnetic coupling requires alteration in the hypercharge numbers. In particular, $\begin{pmatrix} u \\ d \end{pmatrix}_L$ changes from 1 / 6 to 1 / 2, while $(u^c)_L$ goes from $-2 / 3$ to -1, -1 or 0, depending on the colour, and $(d^c)_L$ from 1 / 3 to 0, 0 or 1. The fermionic contribution to vacuum polarization is, conventionally (Masiero, 1984):

$$\frac{4}{3}\times\frac{1}{2}\times\left(\frac{1}{36}\times 3+\frac{1}{36}\times 3+\frac{1}{9}\times 3+\frac{4}{9}\times 3+\frac{1}{4}\times 1+\frac{1}{4}\times 1+1\right)\frac{n_g}{4\pi}=\frac{5}{3\pi},$$

where $n_g = 3$ is the number of fermion generations. However, when modified for integral charges, this becomes

$$\frac{4}{3}\times\frac{1}{2}\times\left(\frac{1}{4}\times 3+\frac{1}{4}\times 3+1+1+0+0+0+1+\frac{1}{4}\times 1+\frac{1}{4}\times 1+1\right)\frac{n_g}{4\pi}=\frac{3}{\pi},$$

corresponding to the change from $C^2 = 5 / 3$ to $C^2 = 3$, when $\sin^2\theta_W = 1 / (1 + C^2)$ changes from 0.375 to 0.25. Interestingly, the addition of the term $(3 / \pi) \ln (M_X^2 / \mu^2)$ to α_G leads *directly to the coupling α for the electromagnetic interaction*, and not to the modified coupling α_1, normalized to fit an overall gauge group, assumed in most Grand Unification schemes, for when $M_X = 1.22 \times 10^{19}$ GeV, $\mu = M_Z = 91.1867$ GeV, and $\alpha_G = 1 / 52.4$,

$$\frac{1}{\alpha_G}+\frac{3}{\pi}\ln\frac{M_X^2}{\mu^2}=128=\frac{1}{\alpha(\mu)}\ .$$

This is of particular interest, since it suggests, a little unexpectedly, that the 'unification' at M_X might involve a direct numerical equalization of the strengths of the three, or even four, physical force manifestations, without reference to the exact unification structure (unless, at grand unification, $C^2 = 0$ and $\sin^2\theta_W = 1$, in a group structure such as $U(5)$, involving gravitation). Such a unification would be far more exact than one dependent on the constants of a particular group structure, and would confirm the interpretation of $\sin^2\theta_W$ as the electroweak constant for a specifically broken symmetry, taking the value of 0.25 at the energy range (M_W - M_z) where the symmetry breaking occurs. It would also, in addition, overcome the perceived lack of coincidence of the three grand unified couplings (α_G) for which more complicated supersymmetric solutions have previously been invoked [Amaldi et al., 1991].

We could tentatively propose that the excluded singlet state in the $SU(5)$ scheme, with its universal coupling to all states, would become the generator responsible for gravity, suggesting a final $U(5)$ scheme, in which all the forces become infinite in range and the generators become phases, as in the electromagnetic interaction and the E representation.

The theory offers a good experimental test in that it predicts that the value of α at 14 TeV, the energy of the LHC now under construction at CERN, will be of the order of 1 / 118 rather than the 1 / 125 which would be predicted by minimal $SU(5)$ (with $\sin^2\theta_W = 0.375$).

CONCLUSION

Many of the most significant facts in particle physics have been shown to follow logically from a representation of the rotation asymmetric (conserved) charges *s, e, w,* by the rotation symmetric system of quaternions i, j, k. The $SU(3)$ and $SU(2)_L \otimes U(1)$ components are immediately apparent, as is the $SU(5)$ nature of the grand unified gauge group. Parity and CP violation, three generations, colour, bosons and fermions, baryons and mesons, quarks and leptons, and even the Higgs boson, can be seen as consequences of a logical structure that has proved susceptible to analysis. Preliminary numerical work provides a new determination of the weak mixing angle, which suggests Grand Unification at the Planck mass, and the incorporation of gravity within an overall $U(5)$ structure.

REFERENCES

Aitchison, I. J. R. and Hey, A. J. G. [1989]. *Gauge Theories in Particle Physics*, second edition (Adam Hilger), 434.

Amaldi, U., Boer, W. de and Fürstenau, H. [1991], 'Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP.' *Phys. Lett.* **260 B**, 447-455.

Fischbach, E. et al. [1986a]. *Phys. Rev. Lett.* **56,** 3.

Fischbach, E. et al. [1986a]. *Phys. Rev. Lett.* **56,** 2423.

Fukuda, Y. et al. [1998]. *Phys. Rev. Lett.* **81**, 1562.

Georgi, H. and Glashow, S. L. 1974]. 'Unity of all elementary-particle forces'. *Phys. Rev. Lett.*, **32**, 438-41.

Georgi, H., Quinn, H. R. and Weinberg, S. [1974], 'Hierarchy of interactions in unified gauge theories'. *Phys. Rev. Lett.*, **33**, 451-4.

Itzykson, C. and Zuber, J-B. [1980], *Quantum Field theory* (McGraw Hill Int.), 59, 87, 148, 243.

Kobayashi, M. and Maskawa, T. [1972]. *Prog. Theor. Phys.*, **49**, 282.

Kounnas, C. [1984], 'Calculational schemes in GUTs', in C. Kounnas, A. Masiero, D. V. Nanopoulos, & K. A. Olive. *Grand Unification with and without Supersymmetry and Cosmological Implication* (World Scientific), 145-281, 188-227.

Masiero, A. [1984], 'Introduction to Grand Unified theories', in C. Kounnas et al. [1984], 1-143, 20-25.

Novikov, V. A., Okun, L. B., Rozanov, A. Z. and Vysotsky, M. I. [1999], 'Theory of boson decays'. *Rep. Prog. Phys.* **62**, 1275-1332.

Rowlands, P. and Cullerne, J. P. [1998], 'A symmetry principle for deriving particle structures'. *Proceedings of the Conference on Physical Interpretations of Relativity Theory VI,* British Society for Philosophy of Science, 316-33.

Rowlands, P. and Cullerne, J. P. [1999], 'A derivation of particle structures and the Dirac equation from fundamental symmetires', in *Aspects II*, ed. K. G. Bowden, ANPA, 155-91.

Weinberg, S. [1996]. *The Quantum Theory of Fields*, 2 vols. (Cambridge University Press), Vol. II, 327-32.

EMERGENCE AND REDUCTION

Peter EISENHARDT; Dan KURTH

Figures and Diagrams by Vera HOHLSTEIN and Jens WALDECK

Institut für Geschichte der Naturwissenschaften
Johann Wolfgang Goethe Universität
D-60054 Frankfurt am Main
Germany
Tel.: (069) 798-28397/22337/22338
Fax: (069) 798-23275
E-Mail: Eisenhardt@em.uni-frankfurt.de

After a critical analysis of the relation of reduction and emergence as contradictory concepts we account for the thesis of their peaceful coexistence. In the light of our theory of emergence new entities come into being where this process is represented by a *discontinuous mapping* but in the final analysis nature consists only of basic entities and their hierarchically stratified interaction-structures.

Emergence and Reduction: A conceptual problem ...

In this paper we will discuss the problem of reduction and emergence hopefully coming to some ... point. In an abstract way the new emergent state of a system is not reducible to the old state. In Fig. 1 (non-reduct morphism) there is a mapping from S_{G+1} to S'_G representing this non-reducibility. In this paper we take the forming of levels for granted (for a detailed discussion of that subject see EISENHARDT, KURTH, WALDECK [1997]) and take for the old and the new states levels of respectively lower and higher complexity.

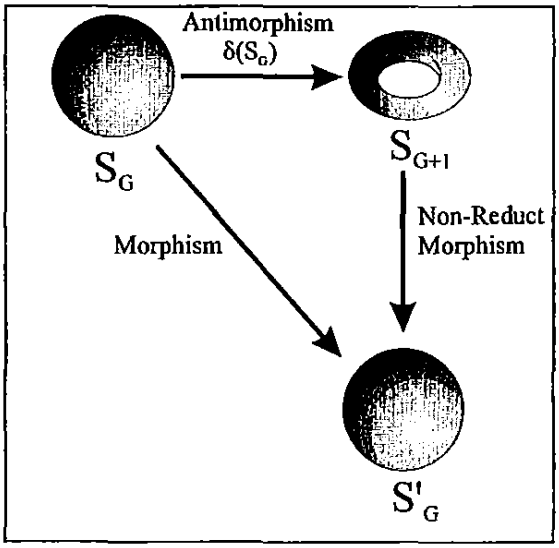


Fig. 1

Nature complexifies itself in the course of the evolution of the universe by getting more and more layered so that the respective lower levels will be preserved. The stratifying process of emergence is pictured in Fig. 2 where the meaning of the non-reduct morphism is shown in its context, namely that of emergence depicted as topological enrichment. The morphisms link the respective base levels at various stages of

complexification. So S_G could represent a class of basic entities at different times of stratification (S_G, S'_G , S''_G,...), whereas S_{G+1} symbolizes the entities which emerged from S_G and are preserved too (S_{G+1}, S'_{G+1}, S''_{G+1},...).

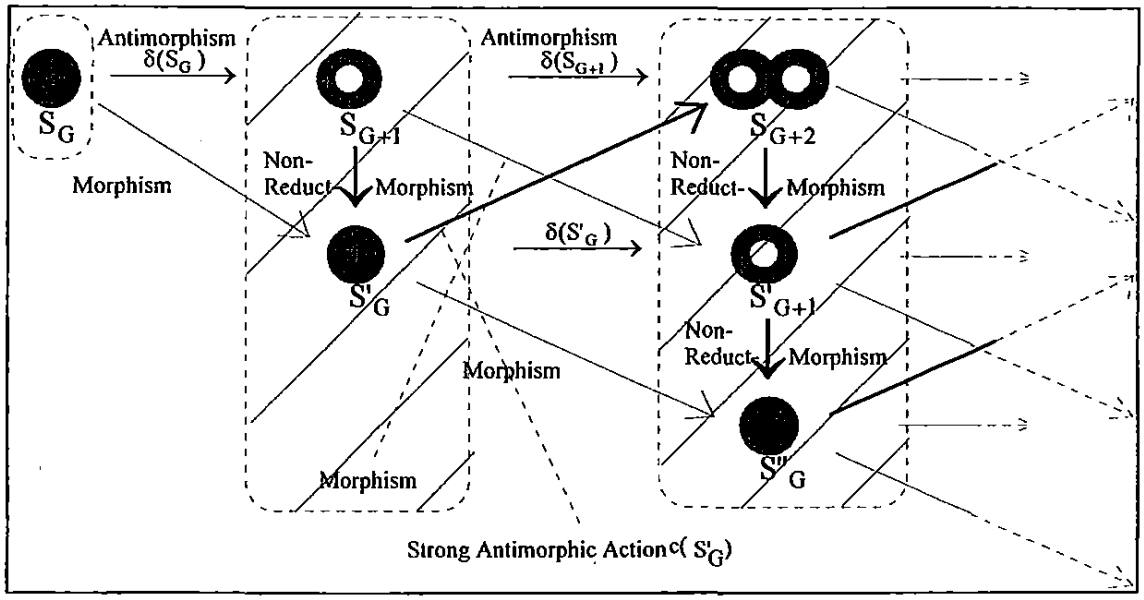


Fig. 2

But what does the phrase "B is not reducible to A" actually mean?

In a recent paper[1] an attempt has been made to reconstruct C.D BROAD's concept of emergence. The argument goes as follows:

1 (Definition): A macro-property P (or better: a $level_n$ -property) of a system S consisting of $C_n^k{}_{-1}$ constituents having the relation R to one another and always possessing this property in accordance with a natural law is *emergent* iff on the basis of the information about the fundamental properties of the C^k and about the laws of entities in general possessing such properties alone it is principally impossible to show that S has all the distinctive characteristics which are typical of P.[2]

For instance if on the basis of the information about the quantumchromo - and quantumelectrodynamical properties of ruby atoms and of gem atoms in general alone it is impossible to show that a ruby is solid (under 'normal' conditions) then solidity is an *emergent property* in accordance with this definition; or take 'transparent' instead of 'solid'; or if on the basis of the information about the physical and chemical properties of neurons and of particular neuronal switches in general alone it is impossible to show that a brain thinks[3] then thinking is an *emergent property*.

But what about the relations? Well, if you have informations about the bonding characteristics and the crystal structure of a ruby atom you have the solidity or transparency immediately. Furthermore your information depends on the meaning of the concept of *relation.* Is it only a (topological or geometrical) *arrangement* or does it involve physical (especially causal) *interactions*?[4] The former often gives poor information whereas the latter hands rich informations over to us so that we can perform an ontological reduction of P to S. (In the definition above 'relation' means 'arrangement'.)

And what about reduction? Is P reduced to S in the example? Is transparency nothing but a interactive relation of atoms? The latter claim sounds plausible. It appears as if the property P_n (of $level_n$) would in a sense be *realized* through a property Q_{n-1} ($level_{n-1}$) of S having all the distinctive characteristics which are typical of P_n. A solid is transparent when the light going through is scattered scarcely by (free) electrons - metals contain a dense gas of electrons that absorbs photons quickly whereas crystals are very transparent out of their absorption areas like the regions of lattice oscillations or these of electrons leaving ions. Let us call this scattering feature of solids Q. It is an arrangement causing interactions; it explains why light passes through some bodies and why it is absorbed (completely) by others. Let us call the property of transparency P. P is realized by Q in the same sense as temperature is realized by the mean kinetic energy of molecules.[5] (If P is realized by Q only and not by other properties R, S, ... then P is theoretically identical with Q. In that case we say that temperature is nothing but the mean kinetic energy of molecules, etc.)

Let us for a moment accept this terminology and two theses[6] will come up immediately because now we find ourselves confronted with a very interesting problem:

2: There is a basic stratum in nature.

(2': There are fundamental entities and relations in nature.)

3: All other strata are realized by the basic stratum.

(3': All other entities are realized by the basic entities.)

Thus we very well might find ourselves being compelled to draw the conclusion: *there is no emergence.*

Because (by presupposition) emergent properties cannot be reduced to the strata or entities they emerged from. In the case of emergence it is in principle impossible to show that the

basis system has all the distinctive features which are typical of the emergent property (see above).

The natural sciences suggest that our two theses are very plausible. Whether nature consists basically of quarks and electrons (still a standard model) or whether oscillating membranes are the fundamental entities (a very up to date model) - it seems as if there were no other entities because these latter ones consist ultimately of the former basic things[7]. So it is! The relations of the 'elements' get more complex and new properties arise but no new entities or substances come up. What seems to be a new entity (let's say: a stone) is nothing but a novel though reducible property of agglomerations of basic entities. A simple logical example is this. The alternative world is made of one (equivalence) class of basic entities C and various relations R^n.

4. What looks like an entity C_2 is 'really' an entity $(C_1\ R_2\ C_1)$ which again is in the end the entity $((C_0\ R_1\ C_0)\ R_2\ (C_0\ R_1\ C_0))$:

$$C_2$$
$$\downarrow$$
$$(C_1\ R_2\ C_1)$$
$$\downarrow$$
$$((C_0\ R_1\ C_0)\ R_2\ (C_0\ R_1\ C_0))$$

Ultimately this world (bearing a slight similiarity to our one) is nothing but a *linear chain of basic entities* $(...(C_0\ R_0'\ C_0)\ R_0''\ (C_0\ R_0'\ C_0\ )\ R_0'''\ (\ C_0\ ...)\ \ )$.[8] A stone or a planet are fictitious entities - they are not ultimately real. The stones and planets crumble into dust, dust disintegrates into atoms, atoms are left an imperceptible short time before they decay leaving nothing but the barren basics. Horrible?

The British emergentists accepted thesis 2 in a physicalistic interpretation (2" There are fundamental entities and relations in nature and these are physical) yet they negated thesis 3 (cf. the definition 1 of emergence above). But then they agreed to a particular interpretation of the 'agglomeration thesis' 4.[9]

The non-reducibility of the new sort of entities comes from the (mereological) agglomeration of parts becoming wholes being parts again of bigger and more complex wholes, and so on. This unexplainable factum (which one has to accept with 'natural piety', as S. ALEXANDER put it) furthermore allows the rise of new causal powers at some levels - the

new properties are causally effective not epiphenomenal. Should we assent? We should not if the recent interpreters of the emergentist are right stating that the emergence of new properties must be *explained* in accordance with the laws of nature and that such a property as

> "...something new...goes beyond the micro-structures. Hence...the emergent property must possess a reality independent of the physical bases." [10]

Differences of the microstructure (relative to a macrostructure which could itself be a microstructure in relation to a 'higher' macrostructure) must be responsible for differences of the macrostructure, or the latter is in a sense 'independent' of the former.

That seems to be a sort of mysterious property pluralism[11] because there exist non-realized properties. The stones, planets and minds emerge miraculously from the dust and its elements breathing creative powers into itself striving for the radical new one. Adorable?

Are we in the horns of a dilemma? On the one hand *reductionism,* on the other hand *dualism* or *pluralism*?[12] There must be a way out of this morass.[13] We want *explainable emergence* of *new strata* in a *monistic if possible physicalistic ontology.* Even though we do not have to argue for such a rather philosophical ideal in this paper because our draft here presented is neutral to ontological positions which one could prefer in a scientific framework, we still have to explain emergence at least in a structural way.

... and its structural solution: emergence as a discontinuous mapping

A structural explanation gives a *mathematical mapping* from one class of entities to another one this mapping being the *law of transformation.* In case of emergence the latter means essentially a separating transition not a realization. Precisely there has to be realization and emergence at the same time - realization representing the continuity and emergence referring to the discontinuity of the evolutionary process. So these concepts are not the horns of a dilemma but the branches of a necessary dichotomy.

Thus we hold the view that

1. There is a basic stratum in nature, and

2. Strata_{n+1} are realized by strata_n *and* strata_{n+1} emerge from strata_n.

We hurry to explain. Thesis 2 is trivially true if the realizing and emerging strata are different: 'transparency' may be realized and 'life' may have emerged. But that's not our

point, naturally. We do mean that the same $stratum_{n+1}$ is realized through $stratum_n$ and has emerged from it.

This might be possible because from the 'internal' point of view one has emergence and from the 'external' point of view one has realization in the sense of reduction - all these perspectives symbolized in Fig. 3 by the arrows of the *strong antimorphic action* and of the *non-reduct morphism*. The iterated relatedness of these arrows which can be seen in Fig. 2 implies that *emergence* is a concept referring to a perspective from below in a hierarchy of strata whereas the conceptual significance of *reduction* just comes up from above obviously.

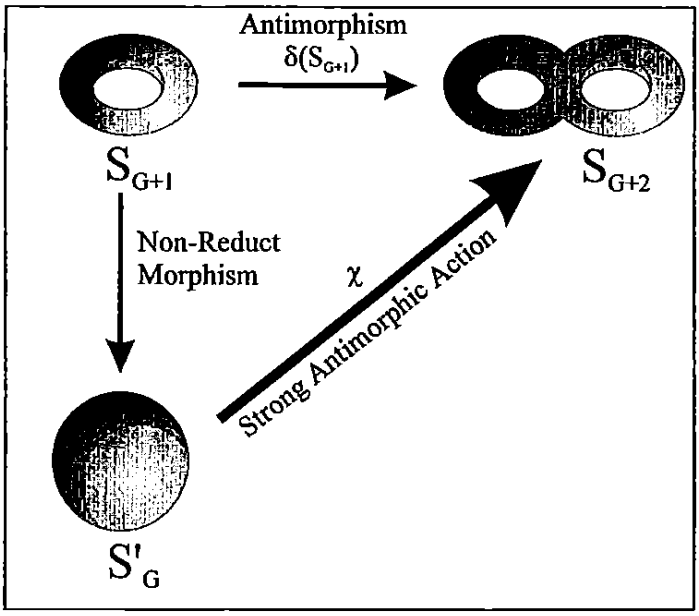


Fig. 3

A particular interpretation offers itself immediately[14]: Subscribe to the thesis of 'epistemological' reduction (involving ontological reduction) and say that the differences of the levels are conceptual differences of unknown boundary conditions respectively. These very boundary conditions are nothing but the distinctive characteristics mentioned above (in thesis 1). Emergent entities consisting ultimately of basic entities cannot be predicted because these typical features - possessing new specific laws requiring new concepts - are unknown. So we have emergence and reduction at the same time.

The emergent new specific features of nature are ontologically as *ultimately real* as the basic entities always had been and still are *originally real*. The *Emergent Complexity* is not compressible to respective lower levels and it is no epistemological quantity because (in a theoretical sense) we *are* into the mill of Leibniz. For that reason we rather would prefer the term 'perspective' instead of 'epistemological' in dealing with the level-relativity of the description of nature. The perspective relation is just as ontological as coarse graining is - not depending essentially on the epistemological description of objects but rather being relative to various 'ontological' viewpoints of 'real objects' to other 'real objects', as for example of human beings to cells or - to dig a bit deeper - superstrings, of organisms in general to molecules, of molecules to atoms, of atoms to quarks and of quarks to superstrings. And there

exist various more and other - less popular and less linearly ordered - level-relative descriptions or perspective relations in nature as well.

Yet if we do not want to argue in terms of 'perspective', let us come back to the characterization of the relations R linking the objects C. These relations consist not only in spatial arrangements but also in causal powers or specific interactions so that the agglomeration of objects forms in a sense new objects possessing new binding forces to agglomerate on a higher level. The regular behavior of an object $C_2 = (C_1\ R_2\ C_1)$ is explainable by the 'binding force' R_2 whereas the objects $C_1 = (C_0\ R_1\ C_0)$ are in a general sense 'determined' by R_1. The new 'causal powers' R_2 are not explainable by the C_1's and not reducible to them - yet under no circumstances does this mean that these interaction are 'independent of the basic objects' or have to be admired with 'natural piety'. Consider the difference of the states of two (quantum mechanical) particles, for instance photons, namely being in separate states, on the one hand, and being in entangled states, on the other hand. The latter form in a sense a new object consisting of two superposed particles. From this point of view quantum mechanics is a theory of natural forms or natural forming.

And there is more in this example than what becomes obvious at the first glance, namely an at least slightly more physicalistic underpinning of the previous rather philosophical discussion of the relation of emergence and reduction.

Fig. 2 shows the essentials of this thesis very clearly, because on the one hand the morphisms $S_G \rightarrow S'_G \rightarrow S''_G$ etc. represent the *invariance of a basic stratum* whereas the possible *reduction* (so to speak 'from above') of the ever *increasing strata* of nature S_{G+1} $[S'_G]$; S_{G+2} $[S'_{G+1}, S''_G]$; S_{G+3} $[S'_{G+2}, S''_{G+1}, S'''_G]$; S_{G+4} $[S'_{G+2}, S''_{G+1}, S'''_G, S''''_G]$[15] ... can be seen in any of these subsequent columns (with S_{G+1}, S_{G+2}, S_{G+3} and so on at the top) respectively - represented in Fig.2 by the (somehow ironically) so called *non-reduct morphisms*. On the other hand the antimorphisms $\delta\ (S_{G+n})$ mapping any S_{G+n-1} to the respective S_{G+n} or - in this respect even more forcefully - our *strong antimorphic action* as the action of the *emergent* events are radical *discontinuous* mappings (which could be viewed as the *inverse* of the *non-reduct morphisms* and - in that sense - as coming 'from below'), as for example from S'_G to S_{G+2}, or from S'_{G+1} to S_{G+3} and so on.

Such a *discontinuous mapping* is in a certain sense comparable to a kind of symmetrybreaking in the theory of dynamical systems which is characterized by the occurrence of a singularity in the otherwise continuous flow (cf. Fig. 4a and Fig. 5a) of the dynamics of the respective systems. Examples of such singularities can be seen in Fig. 4b and

Fig. 5b. Fig. 4b shows a two dimensional projection of a discontinuity similar to the original three dimensional cusp-catastrophe which had been described by R. THOM in his catastrophe-theory[16]. Fig. 5b shows a simple case of a bifurcation, which is a common feature of the theory of dynamical systems[17].

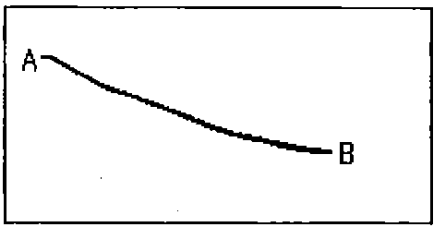


Fig. 4a

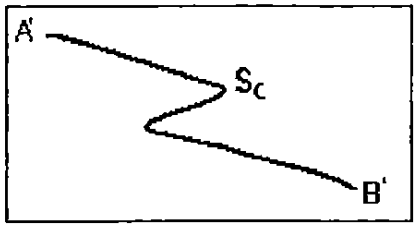


Fig. 4b

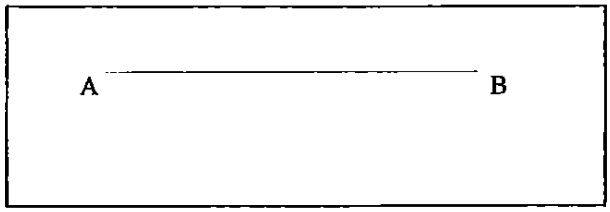


Fig. 5a

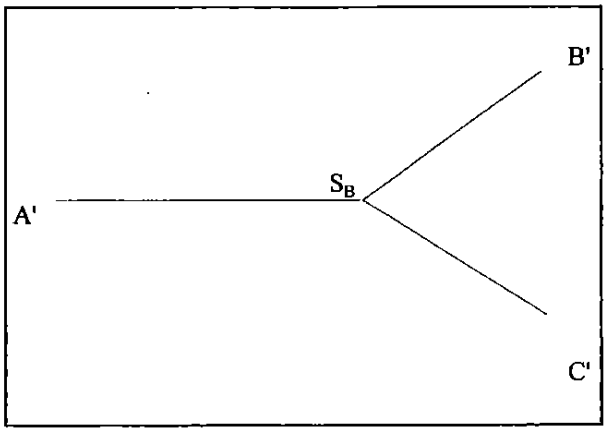


Fig. 5b

In these cases[18] the discontinuities are respectively made up of a point of nondifferentiability in an otherwise homogeneously differentiable curve resp. trajectory or field. Going beyond this established model of discontinuity H. HAKEN, the founder of Synergetics, regards the instability of a dynamical flow as a *qualitative change* exceeding

essentially a mere homeomorphic distortion like a grid transformation that had been already considered by D'ARCY THOMPSON in biology.

> "According to studies by D'Arcy Wentworth Thompson at the beginning of the twentieth century, the two kinds of fish [porcupine fish and sun fish, *the authors*] can be transformed into each other by a simple grid transformation. While from the biological point of view such a grid transformation is a highly interesting phenomenon, from the mathematical point of view we ar dealing here with an example of structural stability. In a mathematician's interpretation the two kinds of fish are the same. They are just deformed copies of each other. A fin is transformed in a fin, an eye into an eye, etc. In other words, *no new qualitative features* [italics by us, *the authors*] ... occur." [19]

In contrast to this homeomorphic biological transformation the development of a fertilized egg to an embryo or the genesis of a new species are examples of *structural* instabilities where new qualities emerge. Synergetics considers the forming of new qualitative features of the *trajectories* of dynamical systems. Seen in this light it lays claim to the title "General Emergence Theory"[20] but it fails because of the conventional concentration on *flows* of dynamical systems on just one level of qualitative complexity. Notwithstanding this defect HAKEN then pointed in the right direction by referring to the central problem of emergence, namely criticizing a simple homeomorphic mapping as possibly being a reasonable representation of an emergent tranformation. In contrary there has to be some sort of a *fundamental* discontinuous hiatus in the continuous manifold of representations. Apparently HAKEN had a topological strategy in mind to solve the problem of emergence when he gave a short example of a mathematical interpretation of structural change:

> "Let us draw one of the two fishes on a rubber sheet. Then by mere stretching or pushing together the rubber sheet, we may continuously proceed from one picture to the other. However, there is no way to proceed from [a one point flow diagram, *the authors*] to [a three point flow diagram, *the authors*] by merely stretching or deforming a rubber sheet continuously as it is possible [for two flow diagrams having the same number of points, *the authors*]. In other words, there is no longer any one-to-one mapping between the stream lines of one flow to the stream lines of the other[...]. In the mathematical sense we shall understand by "structural instability" or "structural changes" such situations in which a one-to-one mapping becomes impossible."[21]

This stressing of the failure of a possible one-to-one mapping in the case of real structural changes then had been an noteworthy insight in the required preliminaries of an appropriate concept of emergence. But one has to go ahead with the topological way[22] by further tightening the mathematical restrictions for an emergent transition which must be *non-homeomorphic,* i.e. not one-to-one *and* non-continuously 'back and forth' (technically this means that the mapping and *its reverse* shall be non-continuous).

Obviously a *discontinuous mapping* in a theory of emergence ought to be something very different from the rather familiar *discontinuity* in the theories of (the basically continuous) dynamical systems. And so it is. In the case of the *discontinuous mapping* the question is not

about trajectories or any other kinds of continuous propagation or flow of dynamical systems. The question here is about structure. Yet then there also ought to be at least a certain similarity as otherwise the explanatory significance of our comparison would be rather negligible.

That certain similarity consists in two respects. Namely in the first one, that the respective singularity - which is in the same extent instrumental for the discontinuous mapping in the antimorphic, i.e. structural transition (which the objects and systems in question underlie) as its counterpart is in the field of dynamical systems - is at least in minimal extension again of a pointlike kind.[23] And in the second one, that that singularity again also causes an obvious discontinuity, if not in the continuous flow of a dynamical system but therefore contingently in the sequence of morphisms between any two successive of the iterated automorphisms of the respective objects or systems. Which in fact doesn't mean anything else than that that singularity or discontinuity emerges in the continuance of these objects or systems.

Figs. 6b and 6c show the nature of that singularity in its most elementary design. I.e. an emerging hole in an otherwise homogenous space (Fig. 6a). Obviously such a hole can - at least in its infinitesimal extension - bear a pointlike feature. Yet then this aspect is only of importance for its emergent character, not for its significance as a fundamental and potentially universal characteristic of the discontinuous mapping which again is a defining feature of emergence.

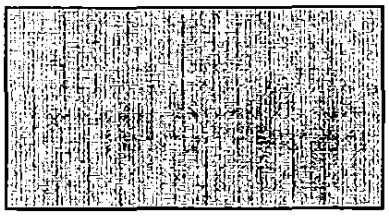


Fig. 6a

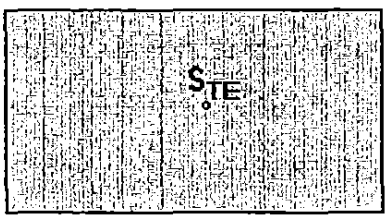


Fig. 6b

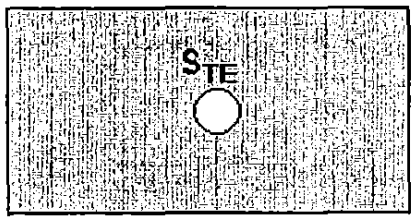


Fig. 6c

Such a singularity must not be seen as an entity in a physical space but rather as a property of a representation space. And that singularity is by far not confined to such a simple topological space as shown in Fig. 6c. It rather refers to the homotopy structure of arbitrary complex spaces, including representation spaces of complex entities of the physical world.

Thus it should have become clear by now that in that respect the notion of '*discontinuous mapping*' is plainly a comparably more general expression of what we already have introduced previously under the heading of '*antimorphic action*'[24]. Then the thrust of this generalization is obviously pointed in the direction of category theory. And exactly this makes the concept of discontinuous mapping a central aspect of the theory of emergent complexity which is based on the idea of antimorphic action.[25]

And then there is even another linkage between the old-fashioned discontinuities of dynamical systems and the emergent singularities in the evolution of structurally complex systems. A linkage which has intricately to do with the problem of - emergence and reduction, again. As it is known from quantum physics - in particular in its path integral formulation - a change of the homotopy structure of a representation space does - by changing its underlying symmetries - also yield a complete change in the possible dynamics which may take place in that space as it is shown in Figs. 7a-d.

Fig. 7a

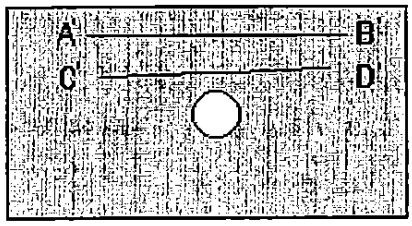


Fig. 7b

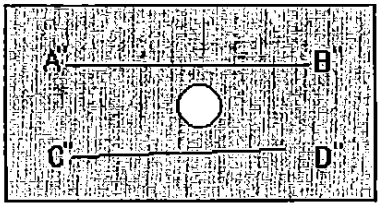


Fig. 7c

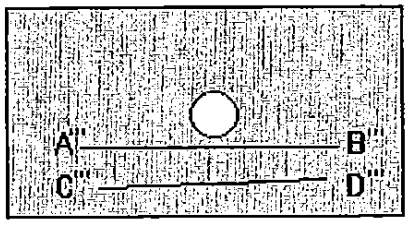


Fig. 7d

These Figures show - in a two dimensional projection onto respective intersecting planes - a change in the possible dynamics in that sense that there is only one relative position (see Fig. 7a) of two non intersecting lines in a homogenous space (with no relevant boundary faces) whereas after the emergence of a hole the symmetry of that space is broken and therefore three distinct relative positions of that lines can be found (see Figs. 7b, 7c, 7d). If such lines are seen as potential trajectories possible changes in the dynamics (e.g. relating to the principle of minimal action) follow immediately.

We hold that a similar change in the internal 'dynamics' of a rather complex entity - as for example in the metabolism of an organism - also takes place if - e.g. such an organism (or rather a kind or family of organism in the course of evolution) underlies a change in its internal or homotopy structure. The emergence of the mentioned singularities or 'holes' in the course of the cosmic and biological evolution then would - nearly literally - mark *points* of no return or better: *holes of irreversibility*.

This exposes a meaning of *discontinuous mapping* in our context which radically diminishes the seeming riddle of coexisting emergence and reduction to a mere question of aspect. I.e. there doesn't exist any possible elementary dynamical *representable description* of any particles trajectory across the discontinuous mappings between the integrative levels of the hierarchy of nature, or: if someone just gives you a fundamental hamiltonian you won't build up an universe, at least none which has any feature of complexity. But if someone gives you a complex universe you won't find anything in it which didn't come up along a hamiltonian from its beginning.

Thus the process of successively iterated emergence or the evolution of complexity might be illustrated by the picture of a weir: the most elementary as well as the increasingly more complex particles and entities being the fishes partly changing their composition by passing the nested stages and being poised to find their way - if at all - only in one direction: forward.

Thus apparently reduction is the one side of a coin of which the other is emergence. In the course of nature's complexification new specific features become evident coming into existence by 'punctuating' or 'punching holes into' the representation space and eventually forming new levels by the various and many layered possible interlockings of the theoretical entities of this space.

Literature

ALEXANDER, Samuel [1920], *Space, Time, and Deity — The Gifford Lectures at Glasgow 1916 - 1918* Vol.I, II, repr. London 1966

BARTELS, Andreas [1996], *Grundprobleme der modernen Naturphilosophie.* Paderborn 1996

BECKERMANN, Ansgar [1997], »Property Physicalism, Reduction and Realization«., in: CARRIER, Martin, MACHAMER, Peter [1997] (eds.), 303-321

CARRIER, Martin MACHAMER, Peter [1997], (eds.), *Mindscapes: Philosophy, Science and the Mind.* Konstanz/Pittsburgh 1997

EHRESMANN, A.C.; VANBREMEERSCH, J.-P.[1987], »Hierarchical Evolutive Systems: A Mathematical Model for Complex Systems«. *Bulletin of Mathematical Biology* 49 (1987) 13-50

EISENHARDT, Peter, KURTH, Dan, STIEHL, Horst [1988], *Du steigst nie zweimal in denselben Fluß.* Die Grenzen der wissenschaftlichen Erkenntnis, Reinbek 1988

EISENHARDT, Peter, KURTH, Dan [1990], »Aufriß einer Theorie der Emergenz«, in: SALTZER, Walter G., *Zur Einheit der Naturwissenschaft in Geschichte und Gegenwart*, Darmstadt 1990, 129-149

EISENHARDT, Peter, KURTH, Dan, WALDECK, Jens [1996], »Emergence as Antimorphic Action«, in: BOWDEN, Keith (ed.), Philosophies, Proceedings of ANPA 17, Cambridge (U.K.) 1996, 125-138

EISENHARDT, Peter, KURTH, Dan, WALDECK, Jens [1999a], »Emergence, Complexity, and Integrative Levels«, in: BOWDEN, Keith (ed.), *Aspects* I, Proceedings of ANPA 19, Cambridge (U.K.) 1999, 67-87

EISENHARDT, Peter, KURTH, Dan, WALDECK, Jens [1999b], »The Categorial Approach to the Part-Whole Relation: Mereological Extra-Level Emergence as the Emergence of new Limits«, in: BOWDEN, Keith (ed.), *Aspects* II, Proceedings of ANPA 20, Cambridge (U.K.) 1999, 46-54

GLENDINNING, Paul [1994], *Stability, Instability, and Chaos*, Cambridge (U.K.) 1994

HAKEN, Hermann [1983], *Advanced Synergetics*, Heidelberg etc. 1983

HAKEN, Hermann [1989], »Synergetik-Eine interdisziplinäre Theorie der Selbstorganisation«, in: WEINGARTNER, Paul, SCHURZ, Gerhard (eds.), *Philosophie der Naturwissenschaften.* Akten des 11. Internationalen Wittgenstein Symposiums, Wien 1989, 231-241

HAKEN, Hermann, WUNDERLIN, Arne [1991], *Die Selbstorganisation der Materie,* Braunschweig 1991

HOYNINGEN-HUEHNE, Paul [1992], »On the Way to a Theory of Antireductionist Arguments«, in: BECKERMANN et al. [1992], 289-301

KIM, Jaegwon [1992], »"Downward Causation" in Emergentism and Non-reductive Physicalism«, in: BECKERMANN, FLOHR, KIM [1992], 119-138

KIM, Jaegwon [1997], »Supervenience, Emergence, and Realization in the Philosophiy of Mind«, in: CARRIER, MACHAMER [1997], 271-293

SEARLE, John R. [1992], *The Rediscovery of the Mind.* Cambridge (Mass.), London 1992

THOM, René [1975], *Structural Stability and Morphogenesis.* London etc 1975

---

[1] Cf. BECKERMANN [1997]

[2] BECKERMANN [1997] p 308. The formulation is ours.

[3] Some sort of brains sometimes.

[4] Cf. SEARLE [1992] p 111.

[5] The latter example (temperature) is from Beckermann [1997]. The essentials of the definition run as follows: If as a consequence of natural laws a system when it has Q possesses all the distinctive characteristics which are typical of P then P is *realized* by Q.

[6] Cf KIM [1992] for a discussion of the following problem.

[7] Not things like stones or drops of water, of course. We do not need to discuss here the supposed superposed event feature of the basic entities being rather fields than particles. (Another problem neglected in this paper is the relation of body to mind.) Remind that speaking about 'entities' and their 'relations' implies speaking about 'properties' because these are (logically) monadic relations ('predicates').

[8] In one sense EHRESMANN , VANBREMEERSCH [1987] p 32 are right in that case saying "*In a hierarchical system which is based on the lowest level objects and their links, all the information is enclosed in these data: the study of an object of any level is reduced to that of ist 0-level organization, consisting of the 0-level components and their links.*" But in another sense they are wrong because the *morpological* information is *not enclosed* in the basic data. We discuss the theory of EHRESMANN, VANBREMERSCH in paragraph 6.2.2.

[9] "Emergent properties are had by aggregates of basic entities standing in an appropriate 'relatedness'. Thus, *no new concrete entities emerge*; all that exists is still the basic physical objects, events, processes, and their

aggregates; it's only that some of these entities come to be characterized by novel characteristics not had by their constituents." (KIM [1992] p 123) In the following passage examples are given to illustrate the emergentists agglomeration theory. One has to place first that, for instance, aggregations of strings form particles like sub^n-quarks etc. forming quarks and electrons which themselves give rise to atoms, let's say, of hydrogen and oxygen shaping water. So we push a little bit deeper into space as the emergentists do in starting with elementary particles or space-time (ALEXANDER - for instance - says explicitely: "...and ultimately the [stuff] of everything is a piece of Space-Time..." (ALEXANDER [1920] Vol. II, p 69)); KIM deals with the problem. "Thus, certain aggregates of water molecules have such emergent properties as transparency, its characteristic viscosity and taste, properties not had by individual water molecules. Vitality, or the phenomenon of life, is thought to emerge from certain highly complex physicochemical conditions; and mentality, or the phenomenon of consciousness, emerges in biological systems when they reach a certain level of complexity."

[10] BECKERMANN [1997] p 312.

[11] or property dualism concerning only the relation of body to mind. Cf KIM [1997] p 293.

[12] KIM [1997] p 293]

[13] The authors quoted and others naturally discuss ways out at the level of a metatheory of emergence.

[14] Cf. HOYNINGEN HUENE [1992] in BECKERMANN et al. [1992] furthermore BECKERMANN [1997] p 314 or BARTELS [1996] p 107 seq.

[15] The respective lower levels or reduction bases of the S_G, S_{G+1}, S_{G+2}, S_{G+3}, S_{G+4} and so on here always in square brackets.

[16] THOM [1975] p 63, p 111. In the original catastrophe the path goes horizontally to the cusp.

[17] Cf. for example GLENDINNING [1994], chs. 8 and 9, especially p 202, p 216.

[18] For a more detailed discussion of that relation of discontinuities as known from the theories of dynamical systems on the one side and the problem of emergence on the other cf. especially EISENHARDT, KURTH, STIEHL [1988], pp 166-184

[19] HAKEN [1983], p 32.

[20] HAKEN 'asserts his entitlement' to a "General Emergence Theory" ("Allgemeine Emergenztheorie", HAKEN [1989], p 231) reflecting (ibid.) on some considerations in a book of the authors of this paper (cf. EISENHARDT, KURTH, STIEHL [1988], p 148) about emergence and the limits of science. There we tried to point out the limits of Synergetics as a "mathematische Darstellung des Emergenzproblems" (mathematical formulation of the problem of emergence).

[21] HAKEN [1983], p 34.

[22] EISENHARDT,KURTH [1990]; HAKEN, WUNDERLIN [1991], ch. 5.6.

[23] I.e. that an infinitesimal hole in a topological context is of course not less of a pointlike entity as e.g. an infinitesimal line segment or linelet in analytical geometry.

[24] EISENHARDT, KURTH, WALDECK [1996]

[25] Cf. EISENHARDT, KURTH, WALDECK [1996], EISENHARDT, KURTH, WALDECK [1999a] and EISENHARDT, KURTH, WALDECK [1999b]

Philosophical Aspects of Spin Networks: An Alternative Einstein Memorial[1]

by

Rainer E. Zimmermann, Wolfram Völcker, Can Yurtöven[2]

I

Since the advent of general relativity theory, the strive for a unification of science (philosophically speaking: for grasping the foundations of physics and hence the foundation of the world) has entered a new phase of intensity. For Einstein himself, it appeared appropriate to look for a unification in terms of gravitation and electromagnetism, the two fundamental forces (or interactions) at the time. To be more precise: Einstein's idea was to actually include electromagnetic fields in his gravitational field equations by introducing them (by means of their potentials) into the metric components of space-time, in first place. This can be interpreted in terms of *visualizing substance as space-time geometry* in the sense of general relativity, representing gravitation, at the same time. (Indeed, as Lewis Feuer has pointed out earlier[3], it is quite certain that Einstein developed his theory under the impression of discussing Spinoza's philosophy when in Prague.) The subsequent advent of quantum theory however, to which Einstein himself contributed some fundamental insight, led to a much more complex situation, because not only did two more forms of interaction (weak and strong fields) turn up, but the basic interpretations of relativity on the one side and quantum theory on the other, differed according to their respective (worldly) domains: on the one hand, relativity being defined in terms of a four-dimensional space-time manifold with a Lorentz metric (of signature -2),

---

[1] Part of this paper has been presented as a talk to ANPA 21, Cambridge (UK), September 1999. Another part presents a preliminary and reduced version of chapters IV through VI of R.E.Zimmermann: Spinoza Revisited, Strings, Loops, and Knots as Topoi of Substance (1999) which will be published elsewhere and has been circulated as a preprint in November and December of 1999. What is presented here is also part of a research project to be established at the University of Kassel and embedded into a European Cooperation Project (INTAS/NIS) together with the universities of Vienna/TU (leader), Kiev, and Moscow. Members of the Kassel group are also Annette Schlemm, Sabine Ley, Doris Zeilinger, and Michael Weh. The authors thank them for their cooperation. One of us (R.E.Z.) would like to thank John Baez, Julian Barbour, Richard Bell, Mary Hesse, Lou Kauffman, Chris Isham, Lee Smolin, John Spudich, Steve Vickers, and Paola Zizzi for illuminating discussions and kind advice on various versions of this (and the other) paper. He also thanks Basil Hiley and his theory group for the occasion of giving a talk on this topic at Birkbeck College London.

[2] IAG Philosophische Grudlagenprobleme, Fachbereich Erziehungswissenschaft und Humanwissenschaften, Universität-Gesamthochschule, Nora-Platiel-Str.1, D - 34127 Kassel. - R.E.Z. is presently on leave of absence to the University of Cambridge: Clare Hall, UK - Cambridge CB3 9AL.

[3] L.S.Feuer: Einstein and the Generations of Science, 2nd ed. 1982.

quantum theory on the other, showing up as being defined in terms of a high-dimensional Hilbert manifold with a positive-definite (Euklidean) metric. Hence, one ended up with two disjoint domains claiming to explain the same universe. So, although in the meantime various intermediate successes have been celebrated (such as the unification of the electroweak force, or the approach towards a *grand unification* of the latter with the strong forces: GUT), the ultimate goal, the unification of all forces (and matter) with gravitation in a *theory of everything* (TOE) has not yet been achieved.

During the last three decades, a basically different approach to unification has been put forward going back to an old idea of John Wheeler's: If it is not possible to unify the competitors within the world, it might be possible to unify them *outside* the world, the basic idea being to introduce an abstract mathematical structure from which space and time (and matter) as fundamental categories of the world could be eventually *derived*. It is quite straightforward in fact, to notice the connotation of substance here: If we define our world in terms of fundamental categories such as space and time (and matter), then everything outside the world from which we might be able to derive these fundamental categories is the *foundation* of the world and as such it is *non-being*. Hence, the idea is to visualize the world as a variety which has become out of a primordial (actually pre-worldly) unity. If so, then the next step, namely to formulate this the other way round: that the world *is* in fact this primordial unity as being observed as a becoming variety by its members who have restricted means of perception, is relatively small. Contrary to what Einstein thought, space-time-matter would not be substance itself, but only the latter's worldly attribute. And what is „before" (and external to) the world, *pre-geometry*, would gain the connotation of a substance.

The question is how to reasonably approach such a conception in more detail. Because, technically, this would point to achieving a unified context for both gravitation and quantum theory anyway. A useful motivation for „quantum gravity" of that sort has been given by John Baez recently[4]: There are three fundamental length scales, defined in terms of three physical constants coming from both regions in question ($\hbar$, the Planck constant divided by 2π, Newton's gravitational constant G, and the velocity of light c, respectively), which are important for relativity and quantum theory, at the same time. One is the *Planck length* $l_P = (\hbar G/c^3)^{1/2}$, of roughly the order of magnitude of 10^{-35} m. For length scales smaller than this, quantum gravity would be the appropriate tool to describe the physics there. Another one is the *Compton wavelength*, $l_C = \hbar/mc$, which basically indicates that the measuring of the position of a particle of mass m precisely within one Compton length, requires energy which is able to create

[4] J.C.Baez: Higher-Dimensional Algebra and Planck-Scale Physics, in: C.Callender, N.Huggett (eds.), Physics Meets Philosophy at the Planck Scale, Cambridge University Press (1999), in press. (We are quoting according to the manuscript version of 28th January 1999, here: 4.)

another particle of the same mass. Hence, this length scale is characteristic for effects in quantum field theory. Finally, there is the*Schwarzschild radius* which basically defines the horizon of a black hole which has been formed by a collpasing star of mass m: $l_S = Gm/c^2$. (All constants up to numerical factors.) If m is now the Planck mass itself ($m_P = (\hbar c/G)^{1/2}$), then $l_C = l_S = l_P$. Hence, we would suppose that at the Planck scale, both domains of physics should show up in a somewhat unified way.

Thinking of the traditional division of classical relativity, and quantum theory, respectively, it is quite natural to ask whether the continuum picture of space-time is only an approximation which inevitably breaks down when approaching the Planck scale. And if so, whether there are constituents of space and time which show up according to some scheme of quantization, such as to construct quantum operators with discrete eigen-values. The microscopic structure of space and time would be determined then by eigen-values and eigen-vectors of purely geometric operators, and the macroscopic superposition of these would show up as the well-known space-time continuum (as a limit for large values). But this would also mean to abandon any underlying space-time structure we have got accustomed with (both in relativity*and in quantum theory* which is also based on a classical space-time background). As Ashtekar and Krasnov have pointed out, „ ... to probe the nature of quantum geometry, one should not begin by *assuming* the validity of the continuum picture; the quantum theory itself has to tell us, if the picture is adequate ...'[6]

For their approach, referred to as „loop quantum gravity", it is the goal therefore, to find a background-free quantum theory with local degrees of freedom propagating causally. This is also true for related approaches (e.g. topological quantum field theory (TQFT) etc.), but it is *not* for the celebrated *theory of superstrings*. The latter has the same objective though, in general terms, namely looking for a *pre-geometric TOE* of some kind. It is however, a chiefly *perturbative* theory meaning that it starts from a given space-time background (very much in the tradition of quantum theory), performing a perturbation expansion as power series expansions of the string coupling constants similar to those used in quantum field theory, and extracting perturbational modes corresponding to physical particles (or particle fields). To that end, all superstring theories contain a massless scalar field called the*dilaton* that belongs to the same supersymmetry multiplet as the graviton. It determines a suitable string coupling constant on which the perturbation expansions can be based.[6] However, the big problem is with the background (whose existence contradicts the principle of diffeomorphism invariance which is one of the central results of general relativity). As Carlo Rovelli has put it, keeping a background means to

[5] A.Ashtekar, K.Krasnov: Quantum Geometry and Black Holes, gr-qc 9804039 v2 (4/2/99), 4.

[6] J.H.Schwarz, N.Seiberg: String Theory, Supersymmetry, Unification, and All That, hep-th 9803179 v2 (22/4/98), 12.

describe the motion of physical entities on a non-dynamical stage instead of describing the dynamics of the stage itself.[7]

There are also other problems of which the dimensionality of the appropriate superstring spaces might be the most relevant. Although the Kaluza-Klein tradition in physics has its merits (and actually impresses by its beauty), and although the problem of ending up with five ten-dimensional theories rather than with one fundamental theory seems to be resolved now by introducing certain dualities among them, the idea being that there is some large moduli space of consistent vacua of a single underlying theory (called M theory by now) containing also a theory of eleven dimensions, it is nevertheless very unsatisfactory - to say the least - that a mechanism is to be (practically arbitrarily) postulated in order to explain why the dimensions should curl up to microscopic spaces (which is actually a flagrant breach of the rule as given in terms of the theorem of sufficient reason) so as to leave us with the four dimensions as we know them. In fact, as we have seen already, this could be accounted for in a much more straightforward way, thinking of our world as an attribute of substance (or: alternatively: as a visualization of the world's foundation). Finally: leaving aside the derivation of the black hole entropy with the exact Bekenstein-Hawking coefficient, which is certainly a success in its own right, also the *experimental evidence* according to the expectations determined by the standard model, is not very well supported at all, because, as Rovelli mentions in his review paper, most testable predictions have all failed to be confirmed so far. Rovelli argues that comparing this lack of success with the considerable success the standard model had in the past, leads to the assumption that there might be some principal limit to further following this path.[8] Things have not been improved since then, as the recent Minnowbrook symposion has shown.[9] This is the reason why within this paper we will concentrate on the other approach (loop quantum gravity and related) rather than on superstrings or M theory. But it should be noted that in principle, the latter approach is also on the line with the concept of substance, and so it is worth being mentioned here.

The approach put forward by Ashtekar and others is mainly based on introducing a new kind of connection which replaces the Levi-Civita connection of general relativity. The idea is to simplify procedures and to bring out more clearly the kinematical similarity of this ansatz with Yang-Mills theory. (Recall that according to the standard theory, Yang-Mills fields show up as gluons in their condensed state being responsible for the quark confinement. In this model then, matter is being visualized as consisting of quarks and leptons together with their Yang-Mills components, namely showing up as 36 quarks (coming in six

[7] C.Rovelli: Strings, loops, and others: a critical survey of the present approaches to quantum gravity, gr-qc 9803024 v3 (7/4/98), 19.

[8] Ibid., 5.

[9] J.C.Baez: This Week's Finds in Mathematical Physics, week 134, math.ucr.edu/home/baez.

flavours and three colours - times two for their anti-particles), six leptons (caring primarily for weak interactions), and twelve Yang-Mills components (eight for the gluons plus four for electroweak interactions). These families come typically in three generations. Usually, this can be alternatively described in terms of group symmetries, for the GUT of the characteristic form U(1) x SU(2) x SU(3), the idea being that the subsequent breaking of symmetries would have actually produced the interactions within the universe as we know them. In particular, within string theory, symmetries may be visualized as side-effects of (string) oscillations in hyperspace, and it is tried to combine all these components as metric components on this very hyperspace - which is nothing but the original idea of Kaluza, only applied to higher dimensions.[10])

II

Coming back to loop quantum gravity: Starting with the 3+1 split of the metric, the phase space of general relativity can be described in terms of a three-dimensional manifold M which is compact and without boundary, and a smooth real SU(2) connection $A^i_a(x)$, as well as a vector density $E^a_i(x)$. Here x refers to co-ordinates on M, the letters a,b, ... = 1,2,3 to spatial, the letters i,j, ... = 1,2,3 to internal indices. The relation to the conventional symbols is secured by the equations

$$g\, g^{ab} = E^a_i\, E^b_i,$$

where $g = \det g_{ab}$, and

$$A^i_a(x) = \Gamma^i_a(x) + \gamma\, k^i_a(x),$$

Γ being here the spin connection associated to the local triad for whose labelling the internal indices can be visualized as being used, k referring to the extrinsic curvature of the constant time 3-surface.[11]

---

[10] Note that Michio Kaku uses this picture for explicitly formulating a relationship to the concept of substance: „To put it quite clearly, the definition of the word „universe" is not anymore „all what exists", but is now „all what can exist". (Hyperspace, re-translated from the German version, Byblos, Berlin, 1995, 307) Obviously, this generalized universe refers to substance as non-being (to the field of possibilities that is), but it is not the same physical universe anymore we are used to deal with. This can only be recovered in terms of the attribute of this substance (which is actually a point Kaku does not realize as far as we can see).

[11] Note that the Einstein summation convention does not apply to internal indices. The parameter γ, called the Immirzi parameter, is usually chosen to be equal to the imaginary unit for recovering the Ashtekar standard notation. We are following here the presentation in C.Rovelli: Loop Quantum Gravity, Living Reviews in Relativity, www.livingreviews.org/Articles/Volume1/1998-1rovelli, Max-Planck-Gesellschaft (Potsdam), 18sqq. See also: Ashtekar, Krasnov, op.cit.

Note that this space-time-split establishes already a theoretical choice which is very relevant for the philosophical implications as to the significance of time: What changes in general relativity in dynamical terms is not the 4-distances within space-time, but rather the 3-distances within spaces as being nested in space-time. Hence, the dynamics is essentially one of 3-dimensional Riemannian spaces. This idea going back to John Wheeler in the sixties is discussed in Julian Barbour's new book[12], where Barbour points out that the „key geometric property of space-times that satisfy Einstein's equations reflects an underlying principle of best matching built into the foundations of the theory.“[13] The time separation of spatial slices shows up here as what Barbour calls a *distinguished simplifier*, as an ordering principle for making unfoldings simple.[14] If time is being visualized as a mere ordering principle, then, in philosophical terms, we are left with space as an attribute. Note however, that the *dimensionality* of space is only a finite representation then, which does not reflect the true nature of space, but only our modal attitude towards it with a view to spatial ordering.

The Ashtekar ansatz is also invariant under local SU(2) gauge transformations, three-dimensional diffeomorphisms of the manifold on which the fields are defined, as well as under (coordinate) time translations generated by the Hamiltonian constraint. The full dynamical content of general relativity is also captured by the three constraints that generate these gauge invariances.[15] So what we can do now is to compare the configuration variable of general relativity as known from gauge theories with the SU(2) connection A on a spatial 3-manifold, and the canonically conjugate momentum E with the Yang-Mills „electric“ field. Physically, the latter is essentially the triad and carries all information about space. This is where in quantum theory, the gauge invariant *Wilson loop functionals* are coming in: They are the path-ordered exponentials of the connections around closed loops. Hence, the name for the theory (loop quantum gravity). We will come to that in a moment.

Note however another relationship first: Usually, the kinematics of quantum theory is defined by an algebra of operators on a Hilbert space. The outcome of the physical theory will depend on the connection which can be uncovered between operators and classical variables, and on the interpretation of the Hilbert space as a space of quantum states. The dynamics is determined by a Hamiltonian (in general relativity called quantum constraint) constructed from the operators. The idea is now to express quantum states in terms of the

---

[12] J.Barbour: The End of Time (The Next Revolution in Our Understanding of the Universe), Weidenfeld & Nicolson, London, 1999. - Note that for the Einstein vacuum equations $R_{ab} = 0$, with space-time as $M = R \times S$, where S is the (t = 0)-slice of M, $R_{0b} = 0$ are the constraints on the initial data, and the remaining equations give the evolution in time. The physical states are given as the subspace of diffeomorphism invariant states that are annihilated by the constraint corresponding to R_{00}. The equations expressing this fact are the Wheeler-deWitt equations, which, in the book of Barbour's, take a central position therefore.

[13] Ibid., 176.

[14] Ibid., 180.

[15] Rovelli, op.cit. (n.61), 19.

amplitude of the connection, given by some functional of the type $\Psi(A)$ in the sense of Schrödinger. Functionals of this kind form a space which can be made into a Hilbert space provided a suitable inner product is being defined. Having done this, we can express the amplitude as $\Psi(A) = \langle A | \Psi \rangle$ in Dirac notation. Is now H our Hilbert space, then we define H_0 as its subspace formed by states invariant under SU(2) gauge transformations. Then it can be shown that an orthonormal basis in H_0 is actually a *spin network* basis.

What is a spin network? Basically, spin networks are graphs with three-valued nodes and spin values on the edges, their states being denoted by $|\Gamma\rangle$. To take the norm refer to the mirror image of a graph and tie up the ends, forming a closed spin network $\Gamma\#\Gamma^*$ of value V such that

$$\langle \Gamma | \Gamma \rangle = V(\Gamma\#\Gamma^*)$$

with

$$V(...) = \Pi\, (1/j!)\, \Sigma\, \varepsilon(-2)^N,$$

here j being the edge label, N the number of closed loops, and ε referring to the intertwining operation taking care of permutations of signs. The product is to be taken over all edges, the sum over all routings. Hence, the networks can be visualized in diagrammatic form such as to represent the underlying „spin dynamics". For instance, a diagram with vertices a, b, c, and i, j, k within the region of intertwiners such that $i + j = a$, $i + k = b$, $j + k = c$, can be interpreted as two particles with spin a and b which produce (create) a particle with spin c. Spin interactions of this kind can lead to the creation of new structures: Take a large part of the network (effectively representing a part of the universe) and detach from this small m-units, and n-units, respectively, as „free ends". The outcome of their tying up to form a new structure can be estimated in terms of a probability for the latter having spin number P, say. This turns out as being basically the quotient of the norm of the closed network and the norm of the network with free ends (times some intertwining operations). The *spin geometry theorem* tells us then that when repeating this procedure and getting the same outcome, then the new quotient is proportional to $(1/2)\cos\theta$, where the angle is one which is taken between the axes of the large units. Hence, it is possible to show that angles obtained in this way satisfy the well-known laws of Euklidean geometry. Or, in other words: This purely combinatorial procedure can be used to actually approximate space from a *pre-spatial* structure which is more basic. This idea has been due to Penrose who originally tried to base his concept of *twistors* on spin networks. In looking for primary concepts of an abstract structure from which space (and eventually space-time) could be approximated,

starting from purely combinatorial elements, he began with looking more closely to the implications of angular momentum as to the re-construction of space.[16] The consequences of this approach have remained relevant until today, although twistors are not in fashion nowadays.[17]

A generalization of spin networks and a connection with knot theory has been achieved more recently by Carlo Rovelli and Lee Smolin referring to their concept of quantum gravity: They start with loops from the outset and show that since spin network states $<S|$ span the loop state space, it follows that any ket state $|\psi>$ is uniquely determined by the values of the S-functionals on it, namely of the form

$$\psi(S) := <S|\psi>.$$

To be more precise, Rovelli and Smolin take embedded spin networks rather than the usual spin networks, i.e. they take the latter plus an immersion of its graph into a 3-manifold. Considering then, the equivalence classes of embedded oriented spin networks under diffeomorphisms, it can be shown that they are to be identified by the knotting properties of the embedded graph forming the network and by its colouring (which is the labelling of its links with positive integers referring to spin numbers).[18]

When generalizing this concept even further, a network design may be introduced as a conceptual approach towards pre-geometry based on the elementary concept of *distinctions*, as Louis Kauffman has shown.[19] In particular, space-time can be visualized as being produced directly from the operator algebra of a distinction. If thinking of distinctions in terms of 1-0 (or yes-no) decisions, we have a direct link here to information theory, which has been discussed recently again with a view to holography.[20] Ashtekar and Krasnov have noted this already when deriving the celebrated Bekenstein-Hawking formula in applying loops to black holes. Referring to punctured horizons they can show that each set of punctures gives actually rise to a Hilbert space of (Chern-Simons) quantum states of the connection on the horizon. Be P

---

[16] R.Penrose: Angular Momentum: An Approach to Combinatorial Space-Time, in: T.Bastin (ed.), Quantum Theory and Beyond, Cambridge University Press, 1970, 151-180.

[17] See however S.A.Huggett et al. (eds.): The Geometric Universe, Science, Geometry, and the Work of Roger Penrose, Oxford University Press, 1998. - For an interpretation of twistors within the framework of substance metaphysics see R.E.Zimmermann: Initiale Emergenz und kosmische Evolution. Zur Rekonstruktion der Substanz-Metaphysik. (System & Struktur I/1, 1993, 39-55) and: Twistors & Substance. On Some Metaphysical Aspects of Science. (in: C.W.Kilmister (ed.), Alternatives, ANPA 15, Cambridge (UK), London, 1994, 131-141)

[18] C.Rovelli, L.Smolin: Spin Networks and Quantum Gravity, gr-qc 9505006 (4/5/95). - If $D(\alpha)$ is the representation of the loop α, then the notation of Penrose can be recovered by $P(\alpha) = (-1)^{n(\alpha)+1}D(\alpha)$, where n is the number of single loops. In knot theoretic language, Penrose's spinor identity takes on the modified form $>< + X + {}^{\vee}_{\wedge} = 0$.

[19] L.H.Kaufmann: Knots and Physics, 2nd ed., World Scientific, Singapore etc., 1993, 459sq.

[20] P.A.Zizzi: Holography, Quantum Geometry, and Quantum Information Theory, zizzi@pd.astro.it.

$= \{j_{p1} \ldots j_{pn}\}$ the set of punctures, and H_p the respective Hilbert space. Then $\dim H_p \sim \Pi_{jp \in P}(2j_p + 1)$, and the entropy of the black hole is simply given by $S_{bh} = \ln \sum_P \dim H_p$. The edges of spin networks can be visualized then as flux lines carrying area. With each given configuration of flux lines, there is a finite-dimensional Hilbert space describing the quantum states associated with curvature excitations initiated by the punctures of the horizon. Because the states that dominate the counting (for S) correspond to punctures all of which have labels $j = ½$, each microstate can be thought of as providing a yes-no decision or an elementary „bit" of information.[21] Hence, the reference to Wheeler's „It from Bit". (Paola Zizzi has tried to generalize this within the conception of inflationary cosmology, and terms this „It from Qubit".[22])

Generalized spin networks can be used now for lattice gauge theory and for non-perturbative quantum gravity. In the former case, they turn out as products of Wilson loops. In the latter case, it can be found that the space of diffeomorphism invariant states is spanned by a basis which is in one-to-one correspondence with embeddings of spin networks.

Note that according to the standard terminology, a loop in some space Σ, say, is a continuous map γ from the unit interval into Σ such that $\gamma(0) = \gamma(1)$. The set of all such maps will be denoted by $\Omega\Sigma$, the loop space of Σ. Given a loop element γ, and a space $\mathbf{X}$ [23] of connections, we can define a complex function on $\mathbf{X} \times \Omega\Sigma$, the socalled *Wilson loop* such that

$$T_A(\gamma) := (1/N)\, \mathrm{Tr}_R\, P \exp \oint_\gamma A.$$

Here, the path-ordered exponential of the connection $A \in \mathbf{X}$, along the loop γ, is also known as the holonomy of A along γ. The holonomy measures the change undergone by an internal vector when parallel transported along γ. The trace is taken in the representation R of G (which is the Lie group of Yang-Mills theory), N being the dimensionality of this representation. The quantity measures therefore the curvature (or field strength) in a gauge-invariant way.[24]

In Topological Quantum Field Theory and in Conformal Field Theory, the path integral definitions are mainly based on (quantum) Chern-Simons theory. (Recall that if A is a connection 1-form for a gauge group, then the quantum Chern-Simons theory has the path integral

$$Z = \int d\mu(A) \exp (i\hbar/4\pi\, S^{cs}(A)),$$

[21] Ashtekar, Krasnov, op.cit., 14.

[22] Zizzi, op.cit., 13.

[23] Pronounce like „En" (which means „cycle").

[24] We are following here the terminology as given by Renate Loll, in: J.Baez (ed.), Knots and Quantum Gravity, Clarendon Press, Oxford, 1994, 6sq.

where $S = \int Tr\ (A \wedge dA + 2/3\ A \wedge A \wedge A)$ is the Chern-Simons action on a compact 3-manifold Σ.) Over a given loop γ, the expectation value $< T(\gamma)>$ turns out to be equal to a knot invariant (the „Kauffman bracket") such that when applied to spin networks, the latter shows up as a deformation of Penrose's value $V(\Gamma)$. This is mainly due to the fact that

$$< T(\gamma)> = K^k(\gamma) = (1/Z) \int d\mu(A) \exp(...)\ T(\gamma, A).$$

So, for any spin network Γ (replace γ by Γ), the old relation holds up to regularization. Hence, spin networks are deformed into quantum spin networks (which are essentially given by a family of deformations of the original networks of Penrose labelled by a deformation parameter $q = \exp(4\pi/(k+2))$ for the Chern-Simons case). The latter may be understood as built up from the representation theory of quantum groups which are deformations of Lie algebras (namely Hopf algebras).[25]

Similar results can also be shown in terms of a categorial approach introduced by Segal and Atiyah.[26] Note also that the CS invariant is important when having a non-zero cosmological constant Λ, because there is an exact physical state of quantum gravity given by $\Psi_{cs}(A) = \exp(k/4\pi\ S^{cs}(A))$, where k is actually related to Newton's constant by $G^2\Lambda = 6\pi/k$. This state can be shown to reproduce $K^k(\Gamma)$ above.[27]

There is also a simplicial aspect to this: Loop quantum gravity provides for a quantization of geometric entities such as area and volume. The main sequence of the spectrum of area e.g., shows up as $A = 8\pi\gamma\hbar G\sum_i(j_i(j_i + 1))^{1/2}$, where the j's are half-integers labelling the eigenvalues. (Compare this with the remarks on black holes above.) This quantization shows that the states of the spin network basis are eigenvalues of some area and volume operators. We can say that a spin network carries quanta of area along its links, and quanta of volume at its nodes. A quantum space-time can be decomposed therefore, in a basis of states visualized as made up by quanta of volume which in turn are separated by quanta of area (at the intersections and on the links, respectively). Hence, we can visualize a spin network as sitting on the dual of a cellular de-composition of physical space.[28]

---

[25] We will not comment here on any mathematical complications of which there are some, e.g. due to divergences of the integral.

[26] See e.g. M.Atiyah: The geometry and physics of knots, Cambridge University Press, 1990, 52sqq. - See also L.Smolin: The future of spin networks, gr-qc 9702030 (16/2/97), 19-21.

[27] See Smolin, op.cit., 21. - See also J.C.Baez (n.79), 36-39, for s-deformations and the cosmological constant in BF theory. - In particular, Rovelli and Smolin (op.cit.) note that loop states are in 1-1 correspondence with knots in space.

[28] C.Rovelli: Strings, loops, and others: ... , op.cit., 8.

As far as the dynamics of spin networks is concerned, there is still another, more recent approach, which appears to be promising as to the further development of topological aspects of quantum gravity (referred to as TQFT). In setting out to develop this new ansatz[29], John Baez notes that there are basically only two new ideas involved in loop quantum gravity. One is the insistence on a background-free approach. The other is to base the theory on the aspect of parallel transport rather than on the metric. Spin networks are at the basis of this approach. But although originally, Penrose thought of them in terms of describing the geometry of space-time, they really turn out to describe the geometry space much better. The idea of Baez is therefore, to supplement loop quantum gravity with an appropriate path-integral formalism. While in traditional quantum field theory, path integrals are calculated using Feynman diagrams, he would like to introduce two-dimensional analogues of the latter, called *spin foams*.[30] Basically, a spin foam is a two-dimensional complex built from vertices, edges, and polygonal faces, with the faces labelled by group representations, and the edges labelled by intertwining operators. If we take a generic slice of a spin foam, we get a spin network. The first explication of spin foams Baez performs with a view to BF theory (as a first simple step).

For this, choose as a gauge group any Lie group G whose Lie algebra is equipped with an invariant non-degenerate bilinear form. Take as space-time any n-dimensional oriented smooth manifold M, and choose a principal G-bundle P over M. The basic fields are the connection A on P, an ad(P)-valued (n-2)-form E on M, where ad(P) refers to a vector bundle associated with P via the adjoint action of G onto its Lie algebra, and the curvature of A, which is an ad(P)-valued 2-form F on M. The Lagrangian for BF theory is simply $L = tr\ (E \wedge F)$. Setting the variation of the action zero gives the field equations: $\delta \int_M L = 0 \Rightarrow F = 0$, $d_A E = 0$. Here, d is the exterior covariant derivative. The first equation tells us that the connection A is flat. BF theory shows up as a *topological field theory*, and locally all solutions look the same, because it does not have any local degrees of freedom. The second equation determines the gauge symmetries. The configuration space of BF theory is the space A of connections on P.[31] The corresponding classical phase space is the cotangent bundle T^*A . Be A_0 the moduli space of flat connections on P, the physical phase space, and be G the group of gauge transformations of the bundle P. Then the *canonical quantization program* can be visualized by the following diagram:

---

[29] This idea goes back to Atiyah e.g., cf. n.75.

[30] J.C.Baez: An Introduction to Spin Foam Models of BF Theory and Quantum Gravity, preprint: 20/5/99. Here: 1sq. par. - We are following here the general outline of this paper, partially paraphrasing details.

[31] To be more precise: on P/s which is the restriction of the bundle P to the time-zero slice {0}xS being identified with S, where S is the spatial part of space-time M = R x S.

$$
\begin{array}{ccc}
T^*(A) & \rightarrow \text{quantize} \rightarrow & L^2(A) \\
\downarrow & & \downarrow \\
\text{constrain} & & \text{constrain} \\
\downarrow & & \downarrow \\
T^*(A/G) & \rightarrow \text{quantize} \rightarrow & L^2(A/G) \\
\downarrow & & \downarrow \\
\text{constrain} & & \text{constrain} \\
\downarrow & & \downarrow \\
T^*(A_0/G) & \rightarrow \text{quantize} \rightarrow & L^2(A_0/G)
\end{array}
$$

The problem is that typically, A and A/G, are infinite-dimensional, making it difficult to define the related L^2-spaces.[32] But, as Baez points out, the great achievement of loop quantum gravity is that it gives background-free definitions of these Hilbert spaces by leaving the traditional Fock space formalism and taking holonomies along paths instead as basic variables. Hence, the basic excitations are not particles anymore (0-dimensional entities), but 1-dimensional spin networks. It turns out, in fact, that $L^2(A/G)$ is actually being spanned by spin network states. Call such a state $\Psi \in$ Fun (A/G) so that any spin network in S (= space) defines such a function. Because Fun is an algebra (namely consisting of all functions on A of the form $\Psi(A) = f(T \exp \int_{\gamma 1} A \ldots T \exp \int_{\gamma n} A)$, where f is a continuous complex-valued function of finitely many holonomies which are represented here by the integral expressions), multiplication by Ψ defines an operator on Fun. We call this operator *spin network observable.* (It actually extends to a bounded operator on L^2, because Ψ is bounded.) In fact, any product of Wilson loop observables can be written as a finite linear combination of spin network observables. Hence, the latter can be used to measure correlations among holonomies of A around a collection of loops. Moreover, it can be shown that a spin network edge labelled by the spin j contributes a length $(j(j+1))^{1/2}$ (s.a.) to any curve it crosses transversely. Hence, length has a discrete spectrum of possible values in quantum gravity.

Note that there is a very promising aside concerning triangulations: Given (n-1)-dimensional space S and any triangulation of it, choose a graph called „dual 1-

[32] Baez, op.cit. (n.71), 3-7.

skeleton", having one vertex at the centre of each (n-1)-simplex and one edge intersecting each (n-2)-simplex. We can express now any state in $Fun(A_\gamma/G)$ as a linear combination of states *coming from spin networks* whose underlying graph is this dual 1-skeleton. Using the holonomy picture we already know, we can define Hilbert spaces $L^2(A_\gamma)$ and $L^2(A_\gamma/G_\gamma)$ as before such that any spin network with γ as its underlying graph defines a function $\Psi \in L^2(A_\gamma/G_\gamma)$. Spin network states are functions of the type $\Psi(A)$. They span the respective L^2-space, and we can therefore choose an orthonormal basis of them. Now, if γ is a graph in S, trivializing the principal G-bundle with which S is equipped at the vertices of γ, gives a map $A \to A_\gamma$ and a homomorphism $G \to G_\gamma$ such that $L^2(A_\gamma) \subset\!\!\to L^2(A)$ and the same for A/G, $\subset\!\!\to$ referring here to the inclusion mapping. When γ is the dual 1-skeleton of a triangulation of S, for 3-dimensional Riemannian quantum gravity, it is always trivalent and particularly easy to visualize (having actually a hexagonal network pattern superimposing the triangulation). Spin numbers specify the lengths of the edges, with spin j corresponding to length $(j(j+1))^{1/2}$. The same can be performed for 4-dimensional BF theory using 4-valent graphs now and tetrahedra.[33]

The definition of a *spin foam* now is very much alike the one for a spin network, only one dimension higher. A spin foam is essentially taking one spin network into another, of the form $F: \Psi \to \Psi'$. Just as spin networks are designed to merge the concepts of quantum state and geometry of space, spin foams shall serve the merging of concepts of quantum history and geometry of space-time.[34] Very much like Feynman diagrams do, also spin foams can be used to evaluate information about the history of a transition of which the amplitude is being determined. Hence, if Ψ and Ψ' are spin networks with underlying graphs γ and γ', then any spin foam $F: \Psi \to \Psi'$ determines an operator from $L^2(A_\gamma/G_\gamma)$ to $L^2(A_{\gamma'}/G_{\gamma'})$ denoted by o such that

$$\langle \Phi', o\,\Phi \rangle = \langle \Phi', \Psi' \rangle \langle \Psi, \Phi \rangle$$

for any states Φ, Φ'. The evolution operator Z(M) is a linear combination of these operators weighted by the amplitudes Z(o). Obviously, we can define a category with spin networks as objects and spin foams as morphisms.

So what we essentially do is the following: Given the (n-1)-dimensional S and a triangulation of S, choose a graph called the dual 1-skeleton. Express any state in Fun as a linear combination of states coming from spin networks whose underlying graph is this dual 1-skeleton. Define now space-time as a compact

[33] Ibid., 16-20.
[34] Ibid., 32-34.

oriented cobordism M: $S \rightarrow S'$, where S, S' are compact oriented manifolds of dimension n-1. (Recall that two closed (n-1)-manifolds X and Y are said to be cobordant, if there is an n-manifold Z with boundary such that ∂Z is the disjoint union of X and Y.[35]) Choose a triangulation of M such that the triangulations of S, S' with dual 1-skeletons γ, γ' can be determined. The basis for gauge-invariant Hilbert spaces is given by the respective spin networks. Then the evolution operator Z(M): $L^2(A_\gamma/G_\gamma) \rightarrow L^2(A_{\gamma'}/G_{\gamma'})$ determines transition amplitudes $<\Psi', Z(M)\Psi>$ with Ψ, Ψ' being spin network states. Write the amplitude as a sum over spin foams from Ψ to Ψ': $< , > = \sum_{F:\Psi\rightarrow\Psi'} Z(F)$ plus composition rules such that $Z(F') \circ Z(F) = Z(F' \circ F)$. This is a discrete version of a path integral. Hence, re-arrangement of spin numbers on the „combinatorial level" is equivalent to an evolution of states in terms of Hilbert spaces in the „quantum picture" and effectively changes the topology of space on the „cobordism level". This can be understood as a kind of *manifold morphogenesis* in time: Visualize the n-dimensional manifold M (with $\partial M = S \cup S'$ - disjointly) as M: $S \rightarrow S'$, that is as a process (or as time) passing from S to S'. This the mentioned*cobordism*. Note that composition of cobordisms holds and is associative, but*not commutative*. The identity cobordism can be interpreted as describing a passage of time when topology stays constant. If there is no change of topology (due to the action of the identity cobordism), then there is no change of state, because we do not have any local degrees of freedom here. Visualized this way, TQFT might suggest that general relativity and quantum theory are not so different after all. In fact, *the concepts of space and state turn out to be two aspects of a unified whole, likewise space-time and process*.

Note that „time" shows up here not as a function, but as a manifold (although arrows are used). This is particularly interesting, because with a view to what Barbour tells us about the „absence" of time, this means that the concept of time is intrinsically included here as a pragmatic ordering principle for localizing topology changes. This is similar to what Prigogine calls the „age of a system", which is roughly a frequency of formations of new structures in a system making the latter more complex. Time as a convention then, would be an approximate „average" over such ages. (It has been commented on this in more detail at another place.[36]) Hence, time shows up as being associated to a kind of measuring device for local complexity gradients. So what we have in the end, is a rough (and simplified) outline of the foundations of emergence, in the sense that we can localize the fine structure of emergence (the re-arrangements of spin numbers in purely combinatorial terms being visualized as a motion-in-itself) and its results on the „macroscopic" scale (as a change of topology being visualized by physical observers as a motion-for-itself). This is actually what we

[35] C.R.F.Maunder: Algebraic Topology, Van Nostrand Reinhold, London etc., 1970, 194.

[36] R.E.Zimmermann: Selbstreferenz und poetische Praxis, Junghans, Cuxhaven, 1991.

would expect of a proper theory of emergence. But note also that space and time, in the classical sense, are obviously absent on a fundamental level of the theory, but can be recovered as concepts when tracing the way „upward" to macroscopic structures. In other words: even as a gross average feature for „shortsighted" human scientists (as Penrose indicates it at the end of his first twistor paper), space and time would nevertheless turn up as (philosophical) categories of concepts, simply, because the meaning of these categories is well-adapted to what humans actually perceive of their world (and communicate to other humans). This is in fact, a point, where Barbour's argument seems to break down (if discussed within this philosophical perspective): What he essentially shows in his book is that quantum theory, *in so far as it is foundational*, describes partly what was called non-being (or substance) in former times. Hence, there is neither space nor time in *real* terms (= realiter, i.e. with respect to what there is in an absolute sense of the world's foundation), but there *is* space and time in *modal* terms (= modaliter, i.e. with respect to what humans perceive of their world). The former refers to substance, or, alternatively, to what *speculative* philosophy is all about. The latter refers to the physical world, or, to what *sceptical* philosophy is all about. The one relies on theoretical speculation according to what we know - speculating about the foundation of the world, which is outside (logically „before") the world, and of which we are not a part therefore, and hence, about which we cannot actually know anything. The other refers to the empirical world, about which, with the help of experiments, we can obtain knowledge, in fact. Obviously, in terms of physics, the first (speculative) aspect is corresponding to physical theory, in so far as it is foundational. The second (sceptical) aspect corresponds to physical theory, in so far as it is empirical.[37]

These results can also be formulated in the language of category theory: As TQFT maps each manifold S representing space to a Hilbert space Z(S) and each cobordism M: S $\rightarrow$ S' representing space-time to an operator Z(M): Z(S) $\rightarrow$ Z(S') such that composition and identities are preserved, this means that TQFT is a functor Z: nCob $\rightarrow$ Hilb. Note that the non-commutativity of operators in quantum theory corresponds to non-commutativity of composing cobordisms, and adjoint operation in quantum theory turning an operator A: H$\rightarrow$ H' into A*:

[37] Do not think that both things would be the same for physics: In general relativity, we can clearly recognize that there is a well-defined conceptual part (in other words: a foundational part) which is actually clarifying the attitude of the approach, but without giving a pathway towards empirical results. But when the Einstein field equations are being introduced, then testable hypotheses can be formulated and checked in observational experiments. Probably, the same is true for quantum theory: In so far it is foundational, one would not expect any directions for experiments. Hence, Barbour would not have to defend his approach against Fay Dowker's objection of not having to offer any testable prediction. A fundamental theory is not actually obliged to make such an offer. Essentially, Barbour is telling us that there is no time at a fundamental level. But then, this is not a revolution at all, because in physics we learnt this already within the development of our century. In philosophy, already Spinoza has formulated this quite clearly, a long time ago.

H' → H corresponds to the operation of reversing the roles of past and future in a cobordism M: S → S' obtaining M*: S' → S.[38]

III

We have already noted the significance of categories for the approaches we are interested in. Louis Crane has announced that category theory would probably be a unifying principle in physics.[39] But this is not simply a point of formal conceptualization. It is also an important aspect of the process of thinking itself. For the first time, Vladimir Trifonov has made this aspect explicit, as a consequence of applying topos-theoretic concepts to physics.[40] He introduces topoi (toposes)[41] as abstract worlds which represent universes of mathematical discourse whose inhabitants can utilize non-Boolean logics for their argumentation (i.e. their propositional structures). In contrary to the *sensory space* which mainly describes the observations of observers (= researchers), the space of motions is the *set of actions*. Be F a partially ordered field. Then, an F-*xenomorph* is a category A(F) of linear algebras over F. The objects of A(F) are called *paradigms* of an F-xenomorph, the arrows are called *actions*. Essentially, a paradigm then, is the set of states of knowledge. A paradigm is called *rational*, if the space of motions M(A) is a monoid. In particular, it can be shown that the set of all possible actions of a researcher is a topos whose arrows are those mappings which preserve realizations of the monoid (of the space of motions). It can also be shown that, if A is a rational paradigm, and the topos of all possible actions is Boolean (non-Boolean), then the paradigm A is classical (non-classical). For a xenomorph F = R of a generic type of the observer's psychology, Trifonov can finally show that an R-xenomorph implies a classical Einstein paradigm, i.e. dimension 4 and signature 2 of the space-time metric.

---

[38] For a somewhat different approach to spin network evolution see also F.Markopoulou, L.Smolin: Causal Evolution of Spin Networks, Nucl.Phys. B 508, 1997, 409, and id.: Quantum Geometry with Intrinsic Local Causality, gr-qc 9712067 (16/12/77). See also more recently F.Markopoulou: Dual Formulation of Spin Network Evolution, gr-qc 9704013, id.: The Internal Logic of Causal Sets: What the Universe Looks Like from the Inside, gr-qc 9811053, and also id.: Quantum Causal Histories, hep-th 9904009 v4 (4/6/99).

[39] As quoted according to J.C.Baez: This Week's Finds in Mathematical Physics, op.cit., week 31, p.1.

[40] V.Trifonov: A Linear Solution of the Four-Dimensionality Problem, Europhys. Lett. 32 (8), 1995, 621-626.

[41] We keep to the original plural of „topos" used by Saunders MacLane, Goldblatt, and others. Besides being more correct in linguistic terms (because although used in French for the first time, and eventually being thought of as an abbreviation, its connotation is in fact a Greek one - which was also intended, by the way - hence, the Greek plural), it is also implying a nice double meaning, because in philosophical terms it has the meaning of characteristic, fundamental concepts (or categories). In ancient Greek, the word „category" is actually originating from *legal* language: Categorize (kathgore‹n) means „to accuse", and the categories are the actual points according to which a person is being accused in front of a judge and which are read from a list of such points. (Practically, humans are accusing nature, because apparently, it is of another *mode of being* than they themselves are.) Hence, categories are „topoi" of conflict.

Also: If A is a non-trivial Grassmann algebra, then the paradigm is the Grassmannian of an R-xenomorph. Because A has a zero divisor, M(A) cannot be a group. Hence, the logic of a Grassmannian paradigm is always non-Boolean, and the mathematics is non-classical. As has been discussed at another place[42], it is very likely that the category of negators (essentially operators acting upon „world states" in order to produce complexity, which can also be visualized as chaotic self-compositions of some suitable, unfolding „ground state" of the world) forms a topos. (They may even turn out to be basically identical with the functor Z: nCob → Hilb, discussed in the preceding section in terms of TQFT.)
The interesting point in the conception of Trifonov's is that (the logically formal part of) *thinking itself* is directly related to the physical process of unfolding the worldly structure as it can be described in terms of cosmological evolution. The basic idea in this is to define a self-referent cycle in the sense that the physical process is producing observers who choose their explicit logic for evaluating what they actually observe. We recognize the idea again, of nature exploring itself by means of human research (or telling its own story to itself, as a kind of self-narration which is modelling its own self-unfolding).
Note however that this time, we have a sound mathematical base for giving a consistent foundation for such a view. In fact, the speculative part of the theory (as far as the aspect of substance and its relationship to its own attributes is concerned) is at least partially formalized, with respect to two important points: On the one hand, it is formalized in being integrated into the description of the physical process itself. On the other hand, it is formalized according to the process of thinking such as to give an explicit choice of logical types which can be utilized for interpreting observations made. The crucial aspect here is in particular that the appropriate logical type can be used to reproduce the phenomenological structure of the worldly observations actually being made (as we know them). Take irreversibility for instance: Types of logic which have a modified law of negation, of the kind $\neg (\neg x) \neq x$, if x is a given proposition, determine the phenomenology as it is actually being observed according to the fact that processes are irreversible in the sense that their logical representation cannot reproduce initial propositions, independent of the number of negation operations acting repeatedly on such propositions. In other words: Recursive operations of this type have no fixed points. In a sense, we can say that temporality is coming in explicitly where earlier the logic remained static all the time (and created considerable difficulties when comparing theory with praxis, as e.g. Lacan has shown in some detail[43]). Hence, the advantage of topoi: They operate in terms of an intrinsic concept of time which can be visualized as a kind

[42] R.E.Zimmermann: The Klymene Principle, Kasseler Philosophische Schriften, Materialien & Preprints, IAG Philosophische Grundlagenprobleme, UGH Kassel, 1999, III.D.4.6 & 11.
[43] J.Lacan: Die logische Zeit und die Assertion der antizipierten Gewißheit. Ein neues Sophisma. In: id., Schriften III, Walter, Olten, Freiburg, 1980, 101-121. (French edition: du Seuil, Paris, 1966.)

of generic concept, unifying the object level of a theory (that about which the theory is actually speaking) with the subject level (which determines the logic of the observer who actually speaks - in terms of the theory). This is another indication as to the phenomenological necessity of time on a macroscopic level of worldly perception and reflexion. (But, as said before, this does not alter the fact that on a fundamental level, time as a concept, may be absent altogether.)[44]
To this end, we note that the process of the concrete, physical unfolding of the world (as it can be assessed in empirical terms) is essentially identical with the process of reflecting about it, in a cyclic manner which secures that the egg comes before the hen. Following Sandkühler here[45], we call this aspect „onto-epistemic", in the sense that both the ontological and epistemological components of the human mode of grasping the world operate very much on the same footing (also very close to Spinoza's argument given in his 2p7, as we have seen earlier). But in the meantime, we can do a little more, thanks to recent research undertaken in terms of topos theory. We will shortly elaborate on this in the following. We begin with recent work of Isham and Butterfield on topos theory and quantum physics.[46]
This ansatz is particularly interesting, because it starts from a propositional view-point, in first place: Basically, the Kochen-Specker theorem states the impossibility of assigning values to all physical quantities, when (at the same time) preserving the functional relations among them. Or, in more technical terms: If dim H is greater than two, there are no global valuations. Reconceive of a valuation now, as giving truth values to *propositions about the values* of a physical quantity rather than assigning a value to the quantity itself. Here, a certain amount of contextuality is involved, in which a value ascribed to a quantity cannot be part of a global assignment of values, but must instead depend on some context. The idea is then, to eventually re-introduce globality, but for the price of ending up with *partial* truth values, which means that the truth value of a proposition belongs to a logical structure that is larger than {0,1}, and these target-logics are context-dependent. In particular, the space of contexts is the category of all Boolean subalgebras of the projection lattice (rather than the category of self-adjoint operators).

---

[44] There is however, a kind of „perceptive obligation" to actually deal with time, because even in the case of the spontaneous re-arrangements of spin networks, on a very fundamental, and truly combinatorial level, we are obliged to characterize them in terms of a sequential process (which „re-arrangement" means indeed). The latter is for the reflecting of our perceptions a necessary ordering principle without which we would be unable to communicate. This is meant when quoting the famous dictum that „time is there for not letting everything happen once and for all." In fact, this is actually the case, if visualizing space and time in terms of substance, which is non-local and eternal: Everything is everywhere once and for all.
[45] H.J.Sandkühler: Onto-Epistemologie, in: id. (ed.), Europäische Enzyklopädie zu Philosophie und Wissenschaften, Meiner, Hamburg, 1990, vol.3, 608-615.
[46] C.J.Isham, J.Butterfield: A Topos Perspective on the Kochen-Specker Theorem, part I: quant-ph 9803055, part II: quant-ph 9808067.

There are two basic aspects to this ansatz: The*first* one is that a theorem can be proven[47] which states that to each generalized valuation v, there is a natural transformation $V: \Sigma \rightarrow \Omega$ for which, at each stage of truth, the component with respect to ©A, is defined by $V_A(a) := v(A = a)$. Here, ©A refers to the spectral representation of bounded self-adjoint operators on a Hilbert spaceH. They are the objects of a category o for which the morphisms are maps ©B → ©A, if there is a Borel function $f: \sigma(©A) \rightarrow R$ such that ©B = f(©A), when σ is the spectrum, and $©A = \int_\sigma \lambda \, dE$, with E being the spectral projection operators. Then, *a generalized valuation on propositions* in a quantum theory is a map v that associates with each group of the form $A \in \Delta$, Δ being a Borel subset of σ(©A), a sieve $v(A \in \Delta)$ on ©A in o. This is actually the crucial point: to associate a sieve, which is the reason for Ω in the theorem to be a subobject classifier.

Recall that in fact, parallel to set theory, where subsets of some set are in 1-1 correspondence with characteristic functions whose target space can be visualized as giving the simplest „false-true"-Boolean algebra, in topos theory, subobjects of an object are similarly in 1-1 correspondence with „characteristic" morphisms having a special object, called the*subobject classifier*, as their target space which is analogous to {0,1}. Take a poset C then: A function that assigns to each $p \in C$ a set X_p, and to each pair $p \leq q$ a map $X_{qp}: X_q \rightarrow X_p$ such that $X_{pp} = \mathrm{id}(X_p)$, and whenever $p \leq q \leq r$, then $X_{rp} = X_{qp} \circ X_{rq}$, is called a *pre-sheaf* X on C. A *subobject* K of a pre-sheaf X is then essentially another pre-sheaf with a similar map, K_{qp} say, which is the restriction of X_{qp}. The collection of all pre-sheaves on a poset C forms a category, denoted Set^C. This can be shown to be a topos then, because pre-sheaves can be defined via *sieves*, which are collections of morphisms acting on objects of C such that compositions are preserved. The crucial property of sieves is that there are subobject classifiers which have the structure of a *Heyting algebra.* (To see this note with the above that, given a pre-sheaf $\Omega: C \rightarrow \mathrm{Set}$, and an object A of C, then Ω(A) is the set of all sieves on A, and if $f: B \rightarrow A$, then $\Omega(f): \Omega(A) \rightarrow \Omega(B)$ is defined as the pull-back to B of the sieve S on A by the morphism f:

$$\Omega(f)(S) := \{h: C \rightarrow B \mid f \circ h \in S\},$$

for all $S \in \Omega(A)$.) The existence of a subobject classifier can be taken as a defining property of a topos. The former does turn out to be an object of possible truth values such that there is a characteristic morphism in the above mentioned sense, which, at each „stage of truth" A in C, can be written like

[47] Ibid., Th.4.3 of part I, cf. Def. 4.1.

$$\chi_A(x) := \{f: B \to A \mid X(f)(x) \in K(B)\},$$

for all $x \in X(A)$. Hence, each stage of truth serves as a possible context for an assignment to each proposition x of a generalized truth value, which is a sieve belonging to the Heyting algebra Ω. (This being the result of finding out that a valuation on propositions must be some sort of structure-preserving function from the set of propositions to the set of truth values of logical algebra. Within this language, the Kochen-Specker theorem can be rephrased, saying that there is no global section of pre-sheaves arising in quantum theory. Defining generalized valuations then, whose values are sieves of operators, can be used to show that each quantum state actually generates such a valuation.)

The basic ideas for this go back to Lawvere who in the late seventies developed a very elegant concept of motion as derived from logical properties of topoi.[48] As a short aside note that in particular, the states X of a body B should be sufficient to determine their own evolution provided the general law of motion ι (in the Lagrangian sense) is known. In fact, X may involve histories of motion, but there is always a morphism $X \to Q$ (the configuration space of B), expressing the fact that each state involves a specific underlying configuration. Classically, the state space will be $X = Q^D$ (the tangent bundle) meaning that infinitesimal histories are all that is necessary. The configuration space $Q \subset\!\!\to E^B$ (subspace of admissable placements) implies that $X \subset\!\!\to Q \times V^B$ (space of velocity fields on B). The dynamically possible motions of B can be singled out from the kinetically possible (Q^T) by knowing the Lagrangian ι: $X \to W$, where the latter refers to the work needed to add to the potential energy of q in order to obtain the kinetic energy of the velocity field v. Note that q a morphism $T \times B \to E$, describing the motion of B in E, where E is usually equipped with a Euklidean metric $d: E \times E \to R$. Then, the action is given by

$$I = \int \iota\,(q, \partial q/\partial t)\, dt,$$

which is a morphism $\varrho \to A = W \otimes T$, $\varrho \subset\!\!\to Q^T$. The object of the ι-possible motions is then the subobject of ϱ such that grad (I) vanishes. This gives the usual Lagrangian equations of motions.[49] To the aspect of having a concept of „intrinsic motion" we will come back later again.

The *second* aspect in Isham and Butterfield is more in the propositional field[50]. Although discussed in somewhat intuitive terms, the explication that truth values

[48] F.W.Lawvere: Variable Quantities and Variable Structures in Topoi, in: A.Heller, M.Tierney (eds.), Algebra, Topology, and Category Theory (Papers in Honour of Samuel Eilenberg), Academic Press, New York, 1976, 101-131. Also id.: Toward the Description in a Smooth Topos of the Dynamically Possible Motions and Deformations of a Continuous Body, in: Colloque Charles Ehresmann, Cahiers Topologie & Geometrie Differentielle, XXI - 4, 1980, 377-392.

[49] Cf. also F.W.Lawvere, S.H.Schanuel: Conceptual Mathematics, Cambridge University Press, 1998 (1997, 1991), especially section 33.

[50] Isham, Butterfield, op.cit., part II, 31sqq.

are essentially sieves, turns out to be of a significant relevance: Given a category C, and to each object A∈C, a set P(A). For each A and d∈P(A), [A,d] shall be thought of as a proposition. If there is a morphism f: B→ A, then there is also a function f*: {[A,d] | d∈P(A)} → {[B,e] | e∈P(B)} acting on the d's. Hence, given f, then [B, f*(d)] is the B-proposition that corresponds to the respective A-proposition by f. Note that if now the composition f: B→ A, g: C → B; f o g: C → A with g*(f*(d)) = (F o g)*(d) is *not* satisfied, then the *-operation is actually *path-dependent*: If a morphism k: C→ A can be factored as $C \to_g B \to_f A$, then the pull-back k*(d) of d∈P(A) *may not be equal to* g*(f*(d)) obatined by factoring through B. While the authors refer to this situation in physical terms as being clearly „pathological", it may be exactly this case which is common for most hermeneutic systems of propositional structures. We will come back to that in the final section.

The general idea is then to assign truth values to each of the propositions. Obviously, [B, f*(d)] is logically weaker than [A,d], because it is its consequence (≤). If now v(A,d) is the truth value assigned to [A,d], then if the latter is totally true, so are all of its consequences. If it is partially true, it is more true the more its consequences are totally true. The truth value v(A,d) is to be determined by which of the consequences [B, f*(d)] of [A,d] is totally true. So there is actually the possibility to define „truth distances", and the semantic value (contents) of a proposition is actually being determined by the set of those of its consequences that are true. Hence, v(A,d) is the set of morphisms f: B→ A such that the associated [B, f*(d)] is totally true. (Recall that „totally" and „partially" refer to the cases according to whether the characteristic object is {0,1} only, or larger than that, respectively.) The proposal is then, that for any [A,d], total truth is just v(A,d) being the set of all morphisms of type f (which is actually the principal sieve on A) - underlying the idea that v(A,d) is a sieve, indeed.

Obviously, these aspects of propositional semantics are of some importance within the field of computer science. Although computer logic might turn out to be not equal to human logic, it is nevertheless very interesting to have a look for possible applications of the above said to this domain.

So, some time ago, starting from a somewhat different perspective, Abramsky and Vickers have begun to place notions of *observing and testing processes* within an algebraic framework in which observations effectively constitute a quantale, and the propositions of geometric logic are related to the logic of finitely observable properties.[51] To this purpose they define *topological systems* to be essentially topological spaces and locales, at the same time, the latter being homomorphisms from the frames of open sets (the complete Heyting algebras)

[51] S.Abramsky, S.Vickers: Quantales, Observational Logic, and Process Semantics, Math. Struct. Comp. Sci. 3, 1993, 161-227.

to the set 2. The idea is then that finitely observable properties closed under propositional connectives of geometric logic only, give a computational interpretation of topology and domain theory in some logical form.

A *quantale* is basically a sup-lattice (a complete join semilattice) equipped with a monoid structure satisfying distributive laws. Hence, quantales are to linear logic as frames are to intuitionistic logic.[52] The programme is then, to use the algebraic framework of modules over quantales to analyze a process equivalence. One of the interesting results is the following: Take typed semantics with two types which are objects of a *quantaloid.* This is a small category such that each homomorphism set is a sup-lattice, and the morphism composition distributes over all joins. (In fact, Hom(quantaloid) is a functor that preserves all joins.) The types are called live (*) and dead (⊥). Actions are observed of a live process, refusals and acceptances are postmortem observations. Introduce ♠ to mark the transition from life to death. Then we get a directed graph which generates the *ready* quantaloid.

What does that actually mean? Basically, the idea deals with the testing of equivalences of processes, defining the respective types of equivalence by quantales with a testing preorder, using Act as the set of atomic actions. Then, a *transition system* labelled over Act is a set Proc equipped with a transition relation $\rightarrow \subseteq$ Proc x Act x Proc. Quantales can be shown to actually generalize both topological spaces and transition systems. Technical refinements of this[53] do yield a whole family of testing equivalences on processes, as instances of an algebraically formulated axiomatic framework. The diagram deals actually with three of them: Ready (R), Failure (F), and Acceptance (A), which are closely related to their respective traces RT, FT, and AT, but with the property that after refusal or acceptance, no more pure actions are possible. They are represented by a suitable quantaloid, and according to a suggestion of Abramsky[54], that different process equivalences represent equivalence in behaviour under different notions of how the processes can be tested or observed, certain fundamental observations are formalized as generators of quantales. They are called subbasic (in analogy with topology), and are defined with respect to a fixed set Act of process actions. Hence, α is the observation that action α has been effectively performed, along with any associated change of state. Then α^{x} is the *refusal* of α, meaning the observation that the process has signalled its inability to perform α. There is no change of state (although the state of knowledge improves). Finally, $\alpha^{\surd}$ is the *acceptance* of α, i.e. the observation that

[52] Ibid. 12sq. - These aspects are related to work by Lambeck of 1958 on categorial grammars with a view to non-communicative linear logic.

[53] For the terminology see ibid., 15, incl. the references quoted therein. See also sections 9.1 through 9.3 for a detailed technical discussion of failures semantics, ready semantics, and acceptance semantics, respectively, ibid., 71sqq.

[54] Ibid., 18.

the process has signalled its ability to performα, although it has not yet done so. The latter two are propositional in nature.
Take e.g. the failures semantics F: Let Act be a set. Then we may present the quantaloid of the form

$$Q = \{\alpha: * \to * \ (\alpha \in \mathrm{Act}),\ \spadesuit: * \to \bot,\ \alpha^x: \bot \to \bot \mid \alpha^x \bullet \alpha^x \leq 1_\bot,\ \alpha^x \bullet \beta^x = \beta^x \bullet \alpha^x\}$$

and the testing order by

$$1_* \leq \spadesuit,$$
$$\spadesuit \bullet X^x \leq \alpha \vee \spadesuit \bullet (X \cup \{\alpha\})^x.$$

The first inequality tells us that if a process is live, then it can die. For the second one, suppose that p is a live process, and that after death, a postmortem examination reveals that it would refuse the actions in X: $p \bullet \spadesuit \bullet X^x \neq 0$. Consider then, whether p could have done α. If so, then $p \bullet \alpha \neq 0$; if not, then a more careful postmortem examination would reveal that p would refuse the actions in $X \cup \{\alpha\}$: $p \bullet \spadesuit \bullet (X \cup \{\alpha\})^x \neq 0$. Inherent in this is the notion of the meaning of a process: The *meaning of a process* is given by the set of its capabilities: Be $a \in Q$ a quantale such that $\{p\} \bullet a \neq \varnothing$. Then construct the semantic domain for processes out of Q. The question is now when two elements of Q might be equivalent as process capabilities: Given a, b in Q, when do we have that for every process p, $\{p\} \bullet a \neq \varnothing$ iff $\{p\} \bullet b \neq \varnothing$? This is basically what this kind of „algebraic semantics" is all about.

IV

We collect now the most important results from the last two sections and relate them to what we have said in the first section as to the underlying intention of our enterprise. We will go backward for this, and take the last aspects first: Hence, as the initial concept, the *meaning of a process* following its observation shows up, especially with a view to classifying equivalent processes visualized as operations on quantities which in turn classify transitions (between states of systems). This may be interpreted as the actual beginning of the process of scientific research which is also the process of human reflexion (in particular: of reflecting about nature). The formalization of this in terms of typed semantics (leading up to the definition of quantaloids) means basically that reflexion has to rely on the abstract mapping of its results, by including them within an algebraic

(or topological, or geometric) framework. Obviously, this is due to semiological conditions which are boundary conditions for human reflexion. On the other hand, these boundary conditions are self-imposed by the cognitive system of humans, because they simply express the human capabilities of „working with" what is perceived. Hence, the close relationship between algebraic, topological, and geometric structures on the one hand, and logic on the other. The relevance of Trifonov's work lies in the fact that it can be shown that formally, the actual choice of human logic (for the practise of reflexions) is nothing but a product of the process which is the object of research in terms of this very logic. The relevance of the work of Abramsky and Vickers is to show that in terms of topos theory, this logic can be explicitly related to the (mathematical) structures mentioned above. (And in fact, for a quite generalized framework, because we would expect that typically, the computer logic (of simulations e.g.) turns out to be different from human logic.)[55] Hence, categories, and especially topoi, provide a formalized (mathematical) structure for computing purposes, for the modelling of processes, and for propositional structures (of modelling the modelling), at the same time. This is the chief advantage of categories: that they make explicit *the onto-epistemic character* of research (i.e. reflexion). Moreover, the intrinsic logic shows that observers who interpret their world in a non-Boolean manner, are of generic type (or: may be „more generic" than Boolean observers).

Reverse the order of argument now: When we think of humans with a logic as products of nature with a physics, then the choice of logic is actually*an outcome* of the physical process, in first place, rather than something which is imposed upon nature by some „external" (or: independent) physical (human) observer. This is a somewhat stronger rephrasing of Trifonov's argument, in fact. The propositional aspect as derived in the paper by Isham and Butterfield then, deals mainly with the problem of mapping theory-languages onto each other: The point is to actually discuss „translational" aspects of semantics. It is important now to notice the role of the composition rules: If they are path-independent operations (acting on propositions), they refer to what Trifonov calls a*rational paradigm* of logic. If however, they are path-dependent, they refer to a *hermeneutic paradigm* instead.[56] Hence, the latter can be referred to the

[55] For the characteristic relationships of algebraic and geometric theories within the framework of topoi see P.T.Johnstone: Topos Theory, Academic Press, London etc., 1977. - The semiological aspects of this, and in particular, the meaning of metaphors with a view to theories, has been discussed in more detail in R.E.Zimmermann: The Klymene Principle, op.cit., section III D. - In his „Toposes *pour les nuls*", Steve Vickers, starting from categories of sheaves as generalized universes of sets, comments on geometric constructions, and refers this more to model theoretic aspects. He shows in which sense a classifying topos can be interpreted as a space of models of a geometric theory.

[56] Note that we would speak of „hermeneutic" rather than of „irrational", because although there is a good deal of irrationality in what is at stake in hermeneutics, the „space of free play" associated with a more flexible amount of „excess meaning" typical for hermeneutic situations, can also be attributed to a mere lack of information under simultaneously performed rational reflexion.

production of excess meanings, which are characteristic for „non-formalized" situations.

This can also be discussed in more physical terms: In fact, for Vickers (though he mentions this within another context), the category of sheaves is essentially a generalized universe of sets. In this sense, Trifonov speaks of „abstract worlds" which are „universes of mathematical discourse". But, on the other hand, given any category C, the pre-sheaf on C is a contravariant functor F: C→ Set. If, in particular, C is the category of shapes, then the morphisms correspond to ways the shapes can be glued together to give one shape (or to be more precise: there are morphisms of the type f: x → y such that x can be included one way or the other in y). Then a pre-sheaf on C can be thought of as a geometric structure built by gluing together these shapes along their common pieces. (Similarly, a pre-sheaf on the category of simplices turns out to be a „kind of space".)[57] Note that there is a choice of configurations in a way, because when we introduce a special pre-sheaf, the subobject classifier, then we actually deal with generating truth values (in the sense of Isham and Butterfield). In particular, we can think of TQFT as a pre-sheaf of Hilbert spaces on the category nCob whose morphisms are n-dimensional cobordisms. Hence, TQFT is a Hilbert space object in the topos of pre-sheaves on nCob. So, it is a quantum theory, because of it's being a Hilbert space object, while its peculiar variability (in assigning different Hilbert spaces to each (n-1)-dimensional manifold representing space) is due to it's being an object in a topos, because this property is expressing the *aspect of intrinsic motion* (= change/variation).

Hence, we can clearly recognize the close interrelationship between the physical states and their representation in terms of „form" (morphology) on the one hand, and their logical conceptualization on the other. This does not actually mean however, to explicitly *derive* logic from physics. But it may demonstrate (for a start) both the intrinsic equivalence and the strict parallelism of the two, very much in the sense of our modern re-interpretation of Spinoza's original intention. And there are many examples by now, which can illustrate the general tendency of what is actually being done: Note e.g. Kauffman's remarks on knot theory and DNA, clearly displaying their semiological relevance.[58] See also the

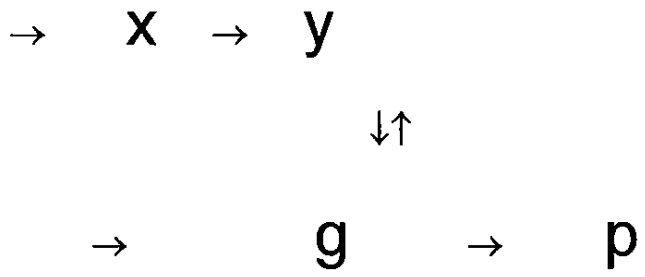
---

[57] This example is mentioned in J.C.Baez, This Week's Finds ..., op.cit., week 115.

[58] L.H.Kaufmann: Knots and Physics, op.cit., 423sqq. - Here molecular processes, knot theoretic operations, and the toplogical information of knots are closely connected to each other. Hence, the DNA shows up as the fundamental carrier of morphogenetic fields: the knotted form functioning as a receiver of field information, knots forming a kind of alphabet of the field language, generating a lexicographic order. This is actually very much in the sense of what Thom expected as a result coming from the semiological interpretation of his catastrophe theory. Cf. R.E.Zimmermann: René Thom - Semiologie des Chaos, in: G.Abel (ed.), Französische Nachkriegsphilosophie, Autoren und Positionen, in press. Note also that according to Rovelli and Smolin, loop states are in 1-1 correspondance with knots in space, and that spin networks (lying at the basis of loops) are simplicial quantum gravity (Haslacher, Perry, 1981). Hence, the relationship between the latter and morphogenetic information. This might be the key for solving the „genetic problem" of Smolin's theory of cosmological natural selection, cf. R.E.Zimmermann: The Klymene Principle, op.cit., ch. VI.

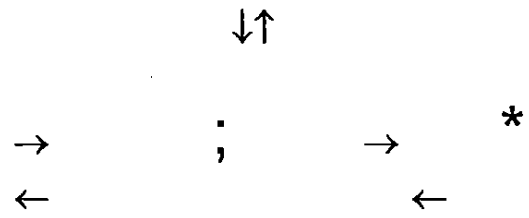
work of Casati and Varzi who give an explicit history of discontinuities[59] modelled according to a parallel with macroscopic morphogenesis. Although the authors do not refer to it, there is an obvious analogy with the perceptive and cognitive processes which are at the basis of human reflexion, as Patricia Churchland has discussed them.[60] On the lowest (or highest, but in anyway most fundamental) level then, we have spin foams which turn out to be equivalent to a „microscopic" picture of an evolution operator acting on Hilbert spaces, determining transition amplitudes of spin network states. Note the abstract character of this underlying foundation of the world, from which physical structures and forms of matter are eventually emerging (the human mind being one of them). Hence, the connotation of substance. Note also that *production* means here differentiation rather: The world can be visualized as a *self-differentiation* of the ground, gradually unfolding the complete hierarchy of worldly structures.

In order to represent the hierarchy so achieved we take „really important" categories and express them in terms of Chinese characters: We start with the foundation of all, given by the abstract structures of spin networks and spin foams, or, alternatively, by Hilbert space operations, going forward to concrete cobordisms, producing changes of topology, then to the geometric picture implying loops and connections, and curvature (or forces) in the end. This is the physical level of worldly processes. At the same time, the fundamental level implies the level of reflexion in terms of logic and hermeneutic, the latter being also produced in the physical sense, but they themselves „acting back" again on what produced them, logic acting upon physics, in attempting to actually represent it, hermeneutic acting *on logic and physics*, respectively, in attempting to represent representations:



[59] R.Casati, A.C.Varzi: Holes, MIT Press, 1994.

[60] P.S.Churchland: Neurophilosophy, Toward a Unified Science of the Mind/Brain, MIT Press, 1992 (1986). Note in particular: 428-447. We find here the remarkable formulation: „There is no reason to imply that the mind might be a nonphysical substance." (317)



We start from the left-hand side with the foundation itself, which is pronounced *tchi* (.) and means „ground". The physical line of production is the upper one, beginning with *en* (cycle) (x), and producing *getsu* (sickle moon = curvature) (y) and hence concrete physics in macroscopic terms. The physical structures emerge from the left out of their foundation, but they themselves also produce one of their forms which is rational mind (logic), expressed by *bun* (sentence/proposition) (g). This in turn produces, through its spaces of free play, excess meanings, and thus hermeneutic, expressed by *wu'a* (harmony) (;). The one produces knowledge laid down in various disciplines, expressed by *gaku* (subject disciplines) (p). The other produces understanding and insight, expressed by *dan* (the rising sun) (*). It is the latter only which can act back onto the ground (= foundation).

Independent of this somewhat metaphorical representation of a systematic approach to what we call *transcendental materialism* (because of obvious reasons), we may note that, in principle, what we have here is a category whose objects are categories, and whose morphisms are functors. Recall that „the theory of categories" itself is a category with finite limits whose models are categories. And in this case, models are functors to Set preserving finite limits.[61] If we could show this for each of our categories, functors of this kind would be *world models*, actually representing (and expressing) the foundation of the world. The interesting point is that part of the foundation itself (remember the spin foams) can be modelled in terms of (mathematical) physics. But in so far this is a part of the foundation of the world rather than of the world itself, the physical theory associated to it, is a *foundational theory*, in the sense that it precludes empirical confirmation. (This is basically true for all conceptual theories of a unifying character.) Hence, it is the formalized counterpart of speculative philosophy. Whilst empirical theories which are subject to experimental testing describe the physics within the world. Unless they are not added to the foundational ingredients of a theory, the latter is not testable. It is

[61] J.C.Baez: This Week's Finds ..., op.cit., week 136.

not its objective to be testable in fact, because it is designed in order to represent worldly foundation. Hence, empirical theories are the formalized counterpart of sceptical philosophy. This relationship is essentially discussed here in terms of an analogy which draws its expressive potential from the classical Spinozist relationship between substance and attribute. Although today, we would not really think of Spinoza's philosophy as a solution to our present problems, it proves nevertheless useful to actually re-construct the latter within the context of a modernized version of the general framework Spinoza's theory laid down as a guideline for further orientation. This is in fact, what philosophy is all about. And in this sense, philosophy might have its own (heuristic) merits.

First Steps into the Particle Zoo

Geoffrey Constable

102 Kings Road
Berkhamsted
Herts HP4 3BP

The Quantised Structure Model for Fermions is used, with Schrödinger's Equation, to explain the existence of the muon and tauon, the masses of these particles being calculated with reasonable accuracy. It is also shown that the masses of zero-spin unflavoured and flavoured mesons can be calculated with good accuracy and that the different particle flavours of such mesons can be explained.

1) Introduction

Previous papers submitted to ANPA (ref 1) describe the derivation of the Quantised Structure Model, a model that defines an extended structure for particles with half-integral spin. This derivation is now summarised in order to provide a basis for the arguments that follow.

Scattering experiments indicate that the electron is small in size, probably less than 10^{-17}cm. On the other hand, the electron has an angular momentum of $\hbar/2$. If such a particle were a spinning sphere, it would have to have a radius at least as large as its Compton Radius (10^{-11}cm) because, if the radius were smaller, the tangential speed at the particle's equator have to be greater than the speed of light – an impossibility. We are driven, therefore, to the alternative concept that the electron is a mass that is very small in size (as indicated by experiment) that orbits around a central point.

Such a concept can be developed further. Imagine a mass 'm' that orbits in a circular path of radius 'r' with speed 'v'. Its momentum-based energy 'E' is given by

$$E = mv^2/2 \qquad (1)$$

If this particle possesses half integral spin and orbits as above, its angular momentum is given by $\hbar/2 = mr^2 \times v/r = mvr$

Thus, $$E = \hbar^2/8mr^2$$

Schrödinger's equation in polar form is

$$-\hbar^2/2m\left(d^2u/dr^2\right) + V(r)u = Eu \qquad (2)$$

where V(r) is the potential energy of the particle when located at a distance 'r' from the central point and the Schrödinger wave function ψ is given by $\psi = u/r$.

Equation (2) can be used to describe the location of a particle in free space when $V(r) = 0$ for all values of r. Equations (1) and (2) can now be combined to yield

$$-\hbar^2/2m\left(d^2u/dr^2\right) = \hbar^2u/8mr^2$$

or $$d^2u/dr^2 + u/4r^2 = 0 \qquad (3)$$

An obvious solution to (3) is $u = A\sqrt{r}$

where A is a constant. To consider the uni-dimensional location (x) we use the substitution $\psi = u/r$. This gives, writing x for r,

$$\psi = A/\sqrt{x} \quad \text{and (in the belief that A is real)}$$

$$|\Psi|^2 = A^2/x$$

Thus the probability that an electron in free space is located at a distance x along the X axis from its central location is reciprocally proportional to x.

The Theory of Quantised Variables (ref 2) is now invoked to propose, as an assumption, that the orbital velocity 'v', as referred to in equation (1), is reciprocally quantised and

can only have values as given by the sequence $c/2, c/3 c/4 \ldots$. This assumption leads immediately to the Quantised Structure Model, as shown in fig 1. In this model, the electron orbits about the 'Z' axis and has a sequence of locations limited to points along the 'X' and 'Y' axes (along either of which the 'quantum observer' observes). The separation between one point and the next is half the electron Compton Radius $R/2$. The tangential speed and probability of location of the electron decrease as 'x' or 'y' increases, but angular momentum is conserved.

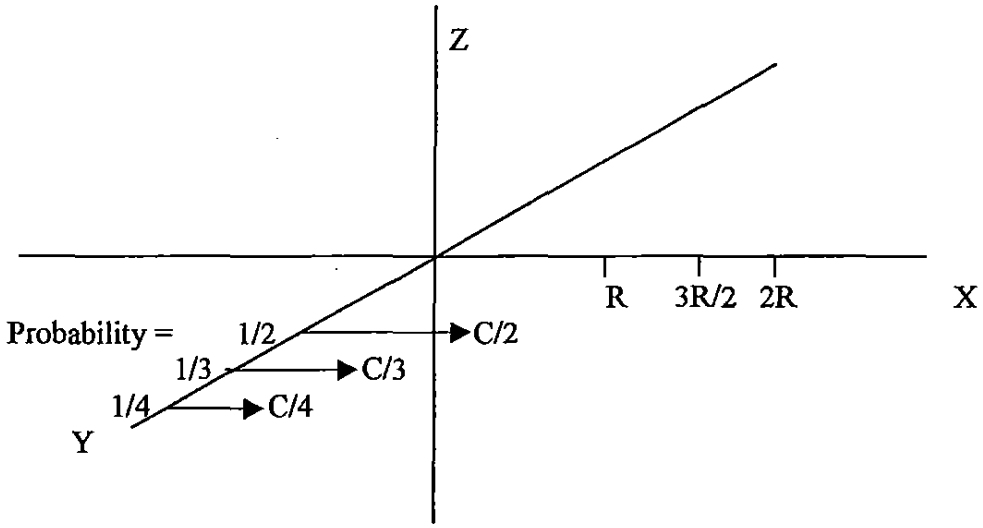


Fig I The Quantised Structure Model

Previous papers (refs 1 and 2) have shown that this model can be applied to explain the forces of electrostatics, electromagnetism and gravity, and the magnetic moments of the electron, proton and neutron. A further deduction (ref 2) is that the age of the universe is ca13.7 billion years. This leads to the interesting (but empirically determined) relationship

$$cT_0/2R = 2^{127} \qquad (5)$$

where T_0 is the age of the universe, R is the Compton Radius of the electron, and 'c' is the speed of light.

As $R = \hbar/M_e c$, equation (5) can be written as $\dfrac{M_e}{\hbar/c^2 T_0} = 2^{128}$, which defines the mass of the electron in terms of fundamental constants.

2 The Expansion of an Electron into a Series of 'Effective Masses'.

The constant in equation (4) can be determined by applying the condition that probabilities must sum to zero, ie that $\Sigma\psi^2 = 1$

To simplify matters, we rewrite equation (4) as

$$\psi^2 = R/2Bx$$

where B is a new constant, referred to subsequently as the 'normalisation constant'.

Hence, the sum of probabilities for any one axis is given by

$$R/2B\,\Sigma(1/x)$$

or

$$R/2B\left(2/2R + 2/3R + 2/4R + \ldots 2/nR\right)$$

If 'n' is infinitely large, the sum of this series will be infinite. Fortunately, as indicated by the Theory of Quantised Variables, all primary variables have maximum values. The maximum value for distance is cT_0. Since we are dealing here with a radius, not a diameter, it is reasonable to guess that the maximum value of the distance *nR/2 is* $cT_0/2$, Whence

$$n = cT_0/R$$

We denote this number as 'N' which, by equation (5) has the value 2^{128}, or 3.4028×10^{38}.

In order to calculate the Normalisation Constant 'B', we make use of the formula that defines Euler's Constant, 0.5772157 ie,

Lim (n tends to infinity) $1 + 1/2 + 1/3 + \ldots 1/n - \ln(n) = 0.5772157$, or

$$1/2 + 1/3 + \ldots 1/n = \ln(n) - 0.4228 \qquad (6)$$

Using the value $n=N$ as given above yields

$$\sum_{2}^{N} 1/n = 88.2998 = B \qquad (7)$$

Note: this series has to start where n = 2, not where n = 1. An element of an electron cannot be located at one half Compton Radius from the mean position of the electron unless the tangential velocity at the 'equator' is 'c' – an impossibility.

Thus, the probability distribution of electron location is as shown in fig2.

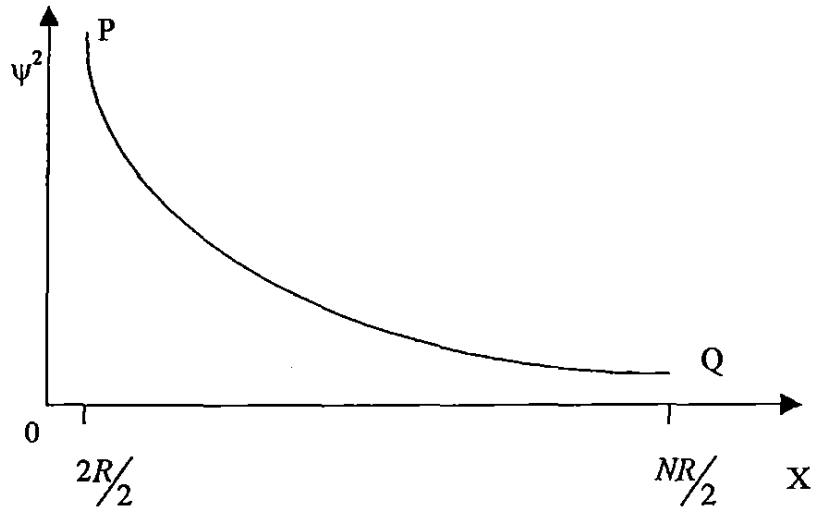


Fig 2 Graph showing distribution of location probability ψ^2 of an electron as a function of 'x'.

The location probability at P (separated by one Compton Radius from the origin) is given by

$$\psi^2 = 1/2B$$

The location probability of the electron when located at Q (separated by N half Compton Radii from the origin) is given by

$$\psi^2 = 1/NB$$

The ratio between these two probabilities is N/2 or 2^{127}

The 'effective mass' of an electron at a particular location can be defined as the mass of the electron multiplied by the probability that the electron will be so located. Thus, the effective mass of the electron when at Q (denoted as M_q) is given by

$$M_q = M_e \times 1/NB$$

But, $$N = cT_0/R \quad \text{and} \quad R = \hbar/M_e c$$

Thus $$M_q = \hbar / c^2 T_0 B$$

The effective mass of the electron M_p when located at P is given by

$$M_p = M_e / 2B$$

The ratio between M_p and M_q is given by

$$M_p / M_q = M_e / \frac{2\hbar}{c^2 T_0} = M_e / 2M_{min}$$

where M_{min} is the minimum mass $\hbar / c^2 T_o$ as predicted by the Theory of Quantised Variables.

3 The Muon and Tauon

The electron, as is well-known, is one of a family of leptons that includes the muon and tauon. These particles resemble the electron but have masses that are different. A test of the theory developed so far is whether it sheds any light on the existence of such particles.

The muon has a mass of 105.7 MeV/c^2 and is more massive than the electron by a factor of 206.85. The tauon has a mass of 1784.1 MeV/c^2 and is more massive than the electron by a factor of 3491.

We restate equation (3) on which the analysis so far rests.

$$d^2u / dr^2 + u / 4r^2 = 0 \qquad (3)$$

Our experience with the electron suggests that it would be helpful to express this equation in a non-dimensional form using the particle half-Compton Radius as the unit of length.

We substitute $r = n/2$ into (3) to obtain

$$d^2u/dn^2 + u/4n^2 = 0$$

To which an obvious solution is $u = A\sqrt{n}$

There is, however, a further solution $u = A'\sqrt{n}\ln(n)$ (A and A' being constants)

Using the substitution $\psi = u/n$

We obtain $\psi = A/\sqrt{n}, or \psi = A'\ln(n)/\sqrt{n}$

When a linear differential equation has two different solutions, it is customary to add these together to give the most general solution.

Thus,

$$\psi = A/\sqrt{n} + A'\ln(n)/\sqrt{n} \quad \text{(believing that A and A' are real)}$$

$$|\Psi|^2 = A^2/n + 2AA'\ln(n)/n + A'^2\ln^2(n)/n$$

Such a probability function differs significantly from that considered previously with respect to the electron on its own. The obvious route forward would be to assume that this function describes the behaviour of a single particle. However, the existence of three separate terms suggests that this function describes three separate sub-particles that collectively possess zero spin. We choose to explore this possibility.

On this basis, normalisation constants can be assigned for each distribution as follows, when constants such as A and A' disappear.

$$\psi^2 = \frac{1}{n\int \frac{dn}{n}}, \frac{2\ln(n)}{n\int \frac{2\ln(n)dn}{n}}, or \frac{\ln^2(n)}{n\int \frac{\ln^2(n)dn}{n}} \qquad (8)$$

Note: these normalisation constants are approximate. To be accurate they should be expressed as the sums of series, not as integrals of continuously varying functions. However, we proceed with this approximation in order to simplify the resulting calculations and because the resulting errors should be slight.

The first term has been shown in earlier papers (refs 1 and 2) to describe the electron. We explore whether the other terms represent the muon and tauon.

To progress we need to make some assumptions. First, as both the muon and tauon possess electronic charge, we assume that both have extended structures similar to that of the electron. In other words, we assume that both particles possess probability distributions that are discrete in form and have probability points separated by half-Compton Radii.

Second, when calculating the properties of the electron, we assumed the maximum distance (as a radius) is $cT_0/2$. This yields the ratio $cT_0/R_e = 2^{128}$, the large number referred to previously as 'N'. It seems reasonable to explore whether the extended structures of the muon and tauon are limited in the same manner by the same large number. On this basis we can proceed.

Taking the first term of (8) $$\psi^2 = \frac{1}{n\int_2^N \frac{dn}{n}} = \frac{1}{n\ln\left(N/2\right)} \quad (9)$$

And the second term $$\psi^2 = \frac{\ln(n)}{n\int_2^N \frac{\ln(n)dn}{n}} = \frac{2\ln(n)}{n\ln^2\left(N/2\right)} \quad (10)$$

And the third term $$\psi^2 = \frac{\ln^2(n)}{n\int_2^N \frac{\ln^2(n)dn}{n}} = \frac{3\ln^2(n)}{n\ln^3\left(N/2\right)} \quad (11)$$

We ask, first, what minimum values are produced by equations (9) – (11) when $n = N$.

From (9) we obtain $$\psi^2 \approx \frac{1}{N\ln\left(N/2\right)}$$

From (10) we obtain $$\psi^2 \approx \frac{2}{N\ln\left(N/2\right)}$$

From (11) we obtain $$\psi^2 \approx \frac{3}{N\ln\left(N/2\right)}$$

These results display apparent quantisation of probability – a result that the Theory of Quantised Variables might lead us to expect and one, therefore, that is to some degree supportive of our assumptions.

We ask, next, what maximum values of effective mass are yielded by equations (9) – (11) when n = 2, the masses of the respective particles being denoted as M_e, M_x, and M_y.

From (9) we obtain $$m\psi^2 = M_e / 2\ln(N/2) \quad (12)$$

From (10) we obtain $$m\psi^2 = 2M_x \ln(2) / 2\ln^2(N/2) \quad (13)$$

From (11) we obtain $$m\psi^2 = 3M_y \ln^2(2) / 2\ln^3(N/2) \quad (14)$$

It will be helpful to sketch these distributions as shown in fig 3.

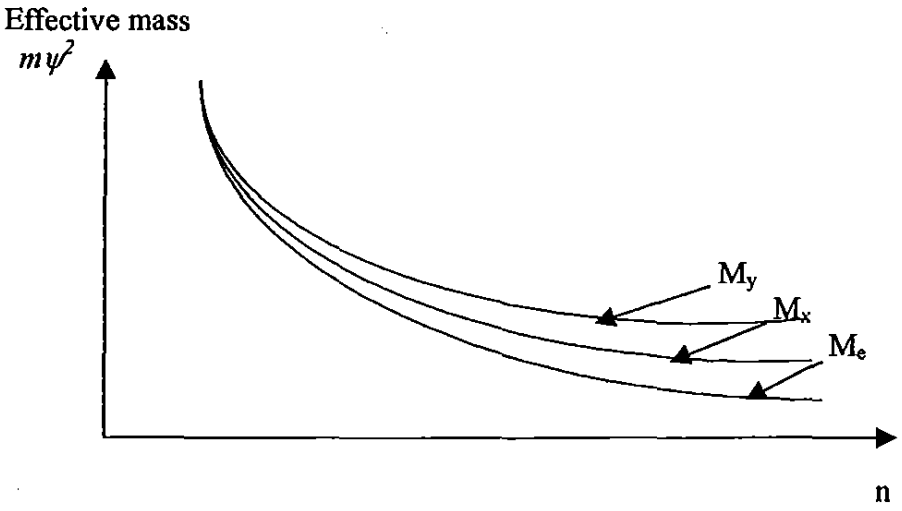


Fig 3 Diagram illustrating the variation of effective mass with distance of the three sub-particles described in equations (9) – (11).

These distributions have been sketched on the assumption that the maximum values of effective mass for these three sub-particles are related. Each sub-particle has a common central point and it is reasonable to guess that, at such a point, one sub-particle can transpose into another. (It is known, for example, that the muon decays into the electron and that the tauon decays into the muon). However, at such an event, there are four degrees of freedom. As was shown in the Quantised Structure Model, the elements of a lepton are located on any of four axes. Thus, at the point of transposition, the element of one sub-particle can transpose into an element of a second sub-particle that can be located in any one of four different ways.

Taking this factor of four into consideration, we can equate the effective masses of the first two particles (M_e and M_x) at the point where n = 2.

$$4M_e \Big/ 2\ln\left(N/2\right) = M_x \ln(2) \Big/ \ln^2\left(N/2\right), \qquad (15)$$

or

$$M_x = 4M_e \ln N/2 \Big/ 2\ln(2)$$

$$= 129.8 \text{ MeV/c}^2$$

This mass is of the same order as that of the muon (105.7 MeV/c^2) – the difference being a factor of 1.23 - which suggests that the second term of equation (9) might account for the muon's existence.

From equations (13) and (14) we write – once more taking account of four degrees of freedom

$$2M_x \ln(2) \Big/ n\ln^2\left(N/2\right) = 4 \times 3M_y \ln^2(2) \Big/ n\ln^3\left(N/2\right)$$

From which

$$M_y = 2M_x \ln\left(N/2\right) \Big/ 12\ln(2)$$

Writing the true muon mass for M_x,

$$M_y = 2237 MeV/c^2$$

which is in fair agreement with the true value of tauon mass of 1784.1 Mev/c^2, the difference being a factor of 1.25 – very similar to that for the muon calculation above.

Thus, although some refinement to the present theory is needed in order to improve alignment between prediction and measurement, it seems likely that the theory described thus far may account for the existence of the electron, muon and tauon. What is not clear thus far is whether the theory describes the full extent of the lepton family. Further leptons have yet to be revealed by experiment and it is possible, therefore, that the three terms of equation (8) define the only three leptons that exist. That said, it is not impossible that equation (8) will be discovered to be the first three terms of a sequence and to possess further terms of the form

$$m\ln^{m-1}(2) \Big/ n\ln^m N/2 \quad \text{(n and m being integers)}$$

that lead to predictions of further leptons.

For example, we calculate the approximate mass of a fourth lepton by equating the third and fourth terms of such a sequence (making, as before, allowance for four degrees of freedom). Thus,

$$3M_\tau \ln^2(2) \Big/ 2\ln^3\left(N/2\right) = 4 \times {}^{4M_p \ln^3(2)} \Big/ 2\ln^4\left(N/2\right)$$

where M_τ is the mass of the tauon and M_p is the mass of a fourth lepton.

This equation can be simplified

$$M_p = {}^{3M_\tau \ln\left(N/2\right)} \Big/ 16\ln(2)$$

$$= 42{,}615 \text{ MeV/c}^2$$

However, we note that the predictions of muon and tauon masses have been too large by a factor of ca 1.25, and guess that this calculated mass should be reduced by the same factor. This change yields a predicted mass of

$$34{,}092 \text{ MeV/c}^2$$

Whether such a prediction is realistic will be discussed later in this paper.

Note: the proposal on which equation (15) is based can be refined to include the other terms of the lepton series. Thus:

$$\frac{4M_e}{2\ln N/2} = \frac{2M_\mu \ln(2)}{2\ln^2 N/2} + \frac{3M_\tau \ln^2(2)}{2\ln^3 N/2} + \frac{4M_p \ln^3(2)}{2\ln^4 N/2} + \ldots.$$

re-arranging,
$$M_\mu = \frac{2M_e \ln N/2}{\ln(2)} - \frac{3M_\tau \ln^2(2)}{2\ln N/2} - \frac{4M_p \ln^2(2)}{2\ln^2 N/2}$$

The mass of the muon can now be calculated in terms of the other lepton masses.

$$M_\mu = 130.2 \quad - \quad 21.00 \quad - \quad 4.2 \quad = \quad 105\text{MeV/c}^2$$

(In this calculation $\ln N/2$ has been calculated as $\sum_2^N 1/n = 88.30$).

This mass agrees well with the true mass of the muon. This agreement depends upon the mass of the fourth lepton being taken into account - a further indication, perhaps, of the existence of such a particle.

4 Investigation of Spin-zero Particles (the charged and uncharged pion)

Thus far the proposed theory has dealt with leptons - particles that possess half-integral spin. However, some particles possess integral spin, and others possess no spin at all. Prominent examples of spin-zero particles include 'unflavoured' mesons such as the charged pion (mass 139.57 MeV/$c^{2)}$ and uncharged pion (mass 134.97 MeV/c^2), and 'strange' mesons such as the charged kaon (mass 493.65 MeV/c^2) and uncharged kaon (mass 497.67 MeV/c^2).

Spin-zero particles that possess both charge and gravitational attraction seem to present a puzzle. It has been argued earlier (refs 1 and 2) that the fundamental forces known to physics (including the electro-static force) arise from particles that possess angular momentum and an extended structure. How, then, can a particle with no angular momentum possess charge?

The answer is that we should not be confined to regarding a particle as a solid spinning sphere. Imagine, for example, a particle with a more complicated structure, one part an extended orbiting structure (similar to that of an electron) that exhibits charge, and another part a small massive spinning sphere that has no extended structure and hence exhibits no charge. If the angular momentum of one part is equal but opposite to that of the other, the particle will possess no 'spin' but will possess charge. There are, of course, other particle models that would exhibit such characteristics. We rely upon the theory described thus far to guide us forward.

Let us imagine that, as for leptons, a meson possesses an extended and probabilistic structure and that the precise location of the particle is uncertain. If the particle has no angular momentum and is stationary, its 'momentum-based' energy will be zero.

If the particle is in free space, it possesses zero potential energy at all locations.

Under these circumstances, Schrödinger's Equation in polar form, see equation (2), reduces to

$$d^2u/dr^2 = 0$$

From which $u = Ar + B$ (A and B being real constants)

Using the substitution $u = \psi/r$ and writing x for r to mark the fact that the equation is now uni-dimensional, we obtain

$$\psi = A + B/x$$

and

$$|\Psi|^2 = A^2 + 2AB/x + B^2/x^2$$

This equation can be expressed in terms of the half Compton Radius by substituting $x = nR/2$

$$\psi^2 = A^2 + 4AB/nR + 4B^2/n^2R^2$$

As before, as when considering the lepton, we note that this expression contains three terms and speculate that these describe the probability distributions of three separate sub-particles – a speculation that is reinforced by the need to regard the meson as a particle with a complicated structure. With normalisation constants added (when the constants shown above disappear),

$$\psi^2 = 1/N,\ 1/n\sum_2^N 1/n,\ or\ 1/n^2\sum_2^N 1/n^2$$

We speculate further that these three terms represent three different sub-particles of mass M_x, M_y and M_z.

Thus

$$m\psi^2 = M_x/N,\ M_y/n\sum_2^N 1/n,\ or\ M_z/n^2\sum_2^N 1/n^2 \qquad (16)$$

These distributions are illustrated in fig 5.

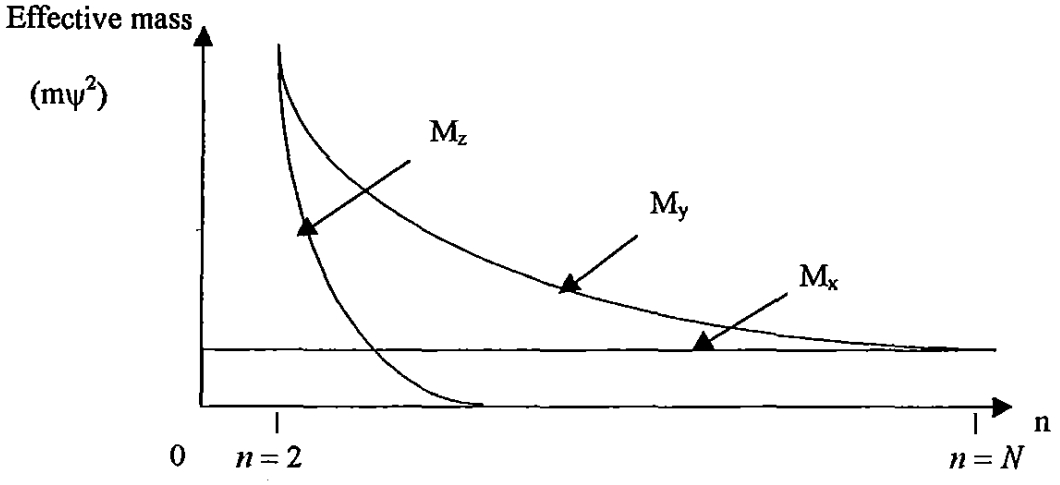


Fig 5 Distributions of spin-zero sub-particles.

We examine first the point where the effective masses of the M_z and M_y sub-particles are assumed to coincide at n = 2.

Taking account of four degrees of freedom and equating the second and third terms of equation (16) yields

$$M_y/M_z = 4 \times 2\sum_2^N 1/n \Big/ 4\sum_2^N 1/n^2 \qquad (17)$$

Writing $\sum_2^N 1/n = 88.30$ and $\sum_2^N 1/n^2 = 0.6449$ yields

$$M_y/M_z = 273.84$$

We guess that M_z possesses electronic mass. This assumption yields

$$M_y = 139.93 \text{ MeV/c}^2$$

in good agreement with the experimentally determined mass of the charged pion (139.57 MeV/c^2).

We now need a further adjustment in order to cope with the spin problem. One possibility is that the charged pion should consist of the M_y sub-particle with its spin and extended structure plus the M_z sub-particle with a far less extended structure and opposite spin. Such a combination would possess charge but no net spin.

Thus we explore the thought that the charged pion is the sum of both sub-particles. But, as we assume the mass of the M_z sub-particle to be that of the electron, we must consider whether this mass should be added to or subtracted from that of the M_y sub-particle. The concept of a negative energy electron has been developed by Dirac and others, and the necessity that this electronic mass should have negative spin suggests that M_z might be such a particle.

This guess requires that the mass of the electron should be subtracted from our recalculated mass to yield

$$M_y \text{ mass} = 139.42 \text{ MeV/c}^2$$

in improved agreement with experiment.

We examine next the implications of equating the first and second terms of equation (16) at the point where n = N. However, at this point the number of degrees of freedom are reduced to three, because the four degrees referred to above are now reduced by one. (It is not possible to extend along the length axis that would take us beyond the $cT_o/2$ point). Taking account of this restriction, we find

$$M_y = 3M_x \sum_2^N 1/n \qquad (18)$$

We guess that M_x is the mass of an electron, which yields

$$M_y = 135.36\,\mathrm{MeV/c^2}$$

in good agreement with the experimentally determined mass of the uncharged pion (134.97 $\mathrm{MeV/c^2}$).

If, as reasoned above, it is necessary to subtract one electron mass, we find that the mass of the uncharged pion is 134.85 $\mathrm{MeV/c^2}$ – in improved agreement with measurement.

5 Unflavoured Mesons of high mass

Equation (16) reads

$$m\psi^2 = M_x/N,\ M_y \Big/ n\sum_2^N 1/n,\ \text{or}\ M_z \Big/ n^2 \sum_2^N 1/n^2 \qquad (16)$$

These three terms appear to be the commencement of a sequence. Further terms of the form

$$M_m \Big/ n^m \sum_2^N 1/n^m$$

might be expected, where 'm' is an integer. Such a term, if equated with

$$M_y \Big/ n\ln(N/2)$$

might lead to predictions of other zero-spin particles. On the assumption that n =2, and $M_m = M_z = M_e$ (M_e is the electron mass)

$$M_y = AM_e \ln N/2 \Big/ 2^{m-1} \sum_2^N 1/n^m \qquad (19)$$

where A is a constant that reflects the number of degrees of freedom and should, therefore, be an integer.

Table 1 lists unflavoured mesons of zero spin and accurately measured mass, and tests the validity of (19). Columns 1 and 2 show the measured masses of all such mesons. Column 3 indicates the accuracy of each mass measurement (σ being the standard deviation provided by the researcher in question). Column 4 gives the value of two equivalent functions (as derived from equation 19). Columns 5 to 9 show the values of these functions when multiplied by the appropriate summed series $\sum_{2}^{N} 1/n^m$, thereby yielding a set numbers that approximate closely to integers or straightforward fractions. Column 10 expresses these numbers as such fractions with denominators 2^{m-1}. Column 11 shows the of the appropriate meson mass calculated using such fractions, to be compared with the measured mass and quoted accuracy.

1 Meson type	2 Mass MeV/c^2 (M_y)	3 σ	4 $A/2^{m-1}\sum_{2}^{N}1/n^m$ or $M_y/M_e\sum_{2}^{N}1/n$	5 $\frac{A_{m=2}}{2}$	6 $\frac{A_{m=4}}{2^3}$	7 $\frac{A_{m=6}}{2^5}$	8 $\frac{A_{m=8}}{2^7}$	9 $\frac{A_{m=10}}{2^9}$	10 $\frac{A}{2^{m-1}}$	11 Calc Mass (M_{ycalc} - M_e)
π^+	139.57	.0007	3.103	1.995					$4/2^1$	139.42
η	547.75	0.2	12.176		1.002				$8/2^3$	547.61
η'	957.75	0.14	21.29		1.753				$14/2^3$	958.70
f_0'	974.1	2.5	21.65			0.3753			$12/2^5$	975.85
η	1295	4.0	28.78			0.4989			$16/2^5$	1301.3
η	1420	20	31.57				0.128		$16/2^7$	1388.0
f_o	1587	11	35.28				0.1433		$18/2^7$	1562
f_o	1709	5	37.99					0.0369	$19/2^9$	1721

Table 1 Calculated masses versus measured masses for spin-zero unflavoured mesons.

The values of constants used in table 1 have been calculated using the series

$$1+\frac{1}{2^{2k}}+\frac{1}{3^{2k}}+\ldots..\frac{1}{n^{2k}}=\pi^{2k}2^{2k-1}/(2k)!B_k$$

where B_k is the appropriate Bernoulli Number. Thus

$$\sum_2^N \frac{1}{n^2}=0.6449, \qquad \sum_2^N \frac{1}{n^4}=0.08232 \qquad \sum_2^N \frac{1}{n^6}=0.01733,$$

$$\sum_2^N \frac{1}{n^8}=0.004062, \qquad \sum_2^N \frac{1}{n^{10}}=0.0009727$$

These data point towards the following observations.

a) The constant 'A' has integral values which, when used to calculate the mass of the particle concerned, provide predictions of remarkable accuracy. This is a strong indication that the theory proposed thus far may have substance.

b) No theory has been formulated thus far to account for the values of 'A'. However, it seems reasonable to guess that such values are related to the number of 'degrees of freedom'.

c) The accuracy of the predictions provided by the various values of 'A' suggest that equation (17) does indeed possess further terms of the type outlined in equation (18).

d) The existence of such terms may require us to modify our views of Schrödinger's equation, from which equation (17) is derived.

6 Spin-zero 'Flavoured' Mesons

Many spin-zero mesons exist that possess specified flavours – for example strange, charmed, charmed-strange, and bottom. These groups have something in common in that each (in company with unflavoured mesons) commences with a triplet of particles, two of plus or minus charge and identical mass, and a third of similar (but not identical) mass but no charge. These triplets are listed in Table 2.

We restate equation (16)

$$m\psi^2=M_s/N,\ M_s/n\sum_2^N 1/n,\ or\ M_s/n^2\sum_2^N 1/n^2 \qquad (16)$$

The mass of the uncharged pion was calculated earlier by equating the first and second terms of this equation at the point where n =N, and assuming that M_x possesses the mass of the first lepton - the electron. Thus

$$M_y = AM_e \sum_2^N 1/n, \qquad (20)$$

where, if the constant A is given the value 3 to denote the number of degrees of freedom available (as explained earlier),

$$M_y = 3 \times 0.511 \times 88.30 = 135.36 \text{ MeV/c}^2,$$

the mass of the uncharged pion.

Unflavoured		
The charged pion	mass 139.57, σ = 0.0007 MeV/^{2}	
The uncharged pion	134.97	0.0008
Strange		
The charged kaon	493.65	0.009
The uncharged kaon	497.67	0.031
Charmed		
$D^{\pm}$	1869.3	0.5
D^0	1864.5	0.5
Charmed/strange		
$D_S^{\pm}$	1968.8	0.7
D_S^0 yet to be observed		
Bottom		
$B^{\pm}$	5278.6	2.0
B^0	5278.7	2.1

Table 2 A list of spin zero meson triplets.

We try the approach outlined above for the uncharged kaon and obtain, as before

$$M_y = AM_e \sum_2^N 1/n$$

If the constant A now becomes the integer 11,

$$M_y = 11 \times 0.511 \times 88.3 = 496.33\,\text{MeV/c}^2$$

or (to good accuracy) the mass of the uncharged kaon.

On the assumption that this result is not mere coincidence, we presume that the integer eleven refers again to the number of degrees of freedom – although a theoretical justification for this assumption has yet to be derived.

We now try a similar approach with the uncharged 'charmed' D^0 particle. A simple arithmetical test reveals that there is no integral or straightforward factional value that can be ascribed to the constant A when using an analogue of equation (20) to yield the correct mass for this particle. However, are we correct to assume in this case that M_x has the mass of an electron? We try using the mass of the muon instead and obtain

$$M_y = AM_\mu \sum_2^N 1/n = A \times 105.658 \times 88.30 = A \times 9329.6 \ \text{MeV/c}^2$$

By inspection, we see that this number is exactly five times too large. Setting A = 1/5 yields, therefore,

$$M_y = 1865.92\,\text{MeV/c}^2$$

a surprisingly accurate outcome.

Turning now to the D_S particle, the uncharged element of the triplet – if it exists – has yet to be discovered. We assume that such a particle exists, and that it has a mass similar to that of the other particles in the supposed triplet. However, following the approach used thus far fails to reveal a suitable value for the constant A, irrespective of whether we use the mass of the electron or that of the muon. We try the mass of the tauon and obtain

$$M_y = AM_\tau \sum_2^N 1/n = A \times 1784 \times 88.30 = A \times 157527 \ \text{MeV/}^2$$

Again by inspection we see that this number is too large by a factor 80. Putting A =1/80 yields, therefore,

$$M_{\gamma} = 1969\,\text{MeV/c}^2$$

once more, a surprisingly accurate outcome.

The B^0 particle presents us with a problem. Following our approach once more, we discover that the constant A cannot be ascribed a suitable value even though we try the masses of the electron, muon and tauon. We speculate that this particle is derived from the mass of a fourth lepton – as yet undiscovered and of mass unknown.

We can try, however, to predict the value of the constant A. When predicting the masses of the electron-based pion, we found that

$$A_{\pi} = 3$$

The kaon was found to be another electron-based particle and, as such, we choose to ignore the value of A used the calculation of its mass.

The D^0 particle was discovered to be muon-based where

$$A_{\mu} = 1/5$$

We note that these two constants are differentiated by a factor of 15.

The D^0_S particle was discovered to be tauon-based where

$$A_{\tau} = 1/80$$

We note that A_{μ} and A_{τ} are differentiated by a factor of 16.

If the constant for the B_0 calculation is denoted by A_p, we speculate that it might be differentiated from A_{τ} by a factor of approximately 15.

This leads to the prediction that

$$A_p \approx 1/15 \times 1/80 = 1/1200$$

Using an analogy to equation (20) we propose the following calculation

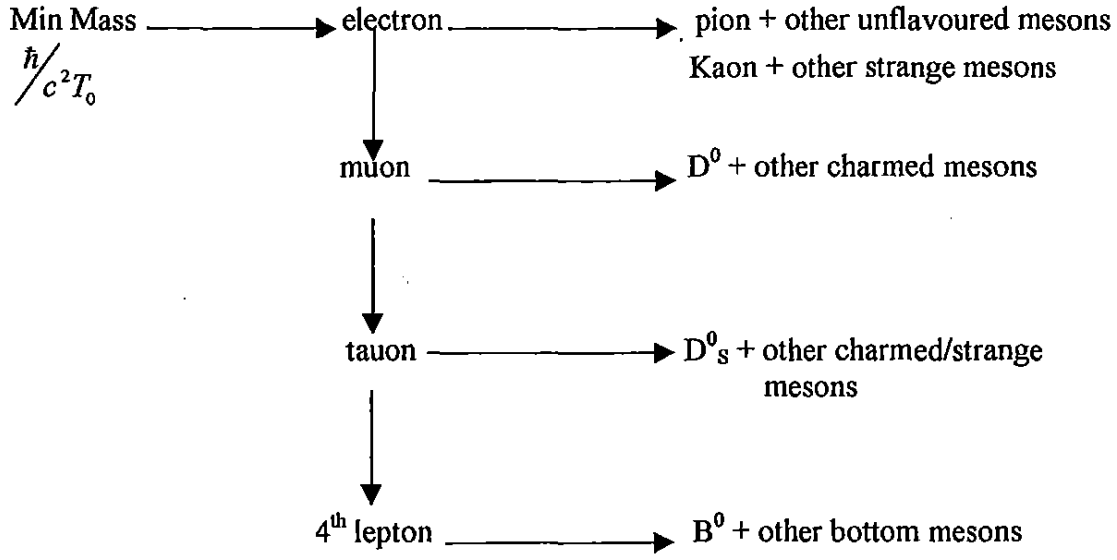
$$5278.6 = 1/1200 \times M_p \times 88.3$$

or $$M_p = 5278.6 \times 1200/88.3 = 71{,}736\ \text{MeV/c}^2$$

This estimate of fourth lepton mass is compared with the equivalent mass estimate produced earlier in this paper of 34,092 MeV/c^2. These two estimates are of the same order and provide some support, therefore, for the speculation that led to their derivation.

8 Summary

This paper proposes how leptons and spin-zero mesons are related to the minimum mass and, hence, to fundamental constants.



The content of this paper is speculative and may well require extension and/or revision. Nevertheless, the arguments put forward seem to fit the facts and may, therefore, have substance. If this proves to be so, there are some important implications for particle physics and for physics as a whole.

In the light of further terms existing for equations (8) and (17), it may be necessary to reconsider the structure or application of Schrödinger's equation.

The apparently universal relevance of the constant $\sum_{2}^{2^{128}} 1/n = 88.30$ confirms the validity of $M_e = \hbar/c^2T_0 \times 2^{128}$ and, hence, that the age of the universe is 13.7 billion years.

Finally, consideration may now have to be given to adopting a comprehensively discrete view of all physical variables.

Geoffrey Constable 3 August 1999

Ref 1 ANPA 18 'An investigation, arising from the Theory of Quantised Variables, into inter-particle forces' - Geoffrey Constable
Ref 2 ANPA 19 'The Theory of Quantised Variables - an investigation into the magnetic moments of elementary particles' - Geoffrey Constable

Operator Formulation of Relativistic Kinematics

P. D. Mountcastle
6382 Brown Circle
Huntington Beach, CA 92647
USA

Abstract

We analyze the transformation which is induced in a signal when it reflects from a moving body, for the restricted case of two or more bodies moving uniformly along a line. This is an elementary step in a program to replace the classical geometrical/kinematical model of motion with a framework based directly on transformation theory. A formalism is developed which describes the relationship between a pair of uniformly moving bodies in close analogy with the familiar transformation properties of electromagnetic signals, instead of geometrically in terms of distance and velocity. The observable physical effects, contraction of a moving rod, time dilation and the one-way doppler effect are worked out and shown to be in agreement with the usual formulation of the theory. The purpose of the reformulation is to eliminate the dependence on the formal concepts of particle trajectory and field in favor of the transformation properties of signals, which are local to the observer and need only be correlated with a local clock.

1 Introduction

First, we give an elementary treatment of the transformation which a light signal undergoes in reflecting from a moving body and returning to the observer. The technique is that of the Green's function or impulse response[1]

[1]Papoulis A.,*The Fourier Integral and its Applications.*, McGraw-Hill, New York (1962).

of a linear channel. The result is familiar, but we only need the principle of linear superposition and the constancy of the speed of light to obtain it. In that way, we demonstrate that the transformation law applies to any kind of signal which obeys the superposition principle and propagates at the speed of light. The signal model is a real scalar time function which might represent, for example, the (unpolarized) electric field magnitude or the output of a polarizer followed by a detector. Anyway, we ignore the effects of polarization for the purposes of the present chapter.

A simple system of labeling is introduced for the operators which represent kinematical signal transformations in the present theory. This notation enables us to formally express the returning signal for N bodies moving along a line as a sum over paths of elementary two-body transformations. According to our analysis, these signal transformations belong to a two-parameter Lie group, which is the affine or $ax + b$ group of wavelet theory[2]. The group properties are demonstrated and formulas are worked out for the product, inverse and square root.

Next, the Lorentz contraction of a rigid rod is worked out by a simple argument which employs only the round trip operators[3]. This is especially satisfactory since it requires only those operators whose parameters can be measured directly by an observer using his own clock, and are therefore observable in a very direct sense.

One-way operators are formally introduced as the the factors of two way operators, and in this way the upstairs or clock index is introduced, along with the whole set of atomic operators which entered into the abstract sum over paths. Explicit formulas are found for all the operators in the theory.

2 Form of the Two-Way Transformation

The direct way of finding the transformation induced on an arbitrary signal when it reflects from a body moving at speed v is to decompose the signal into the limit of a series of narrow pulses. To do it, we write

$$T(t) = \int_{-\infty}^{\infty} T(\tau)\, \delta(t - \tau)\, d\tau. \tag{1}$$

[2]A. Grossman and J. Morlet, *Decomposition of Hardy functions into wavelets of constant shape*, SIAM J. Math. Anal. **15** (1984) 723-736.

[3]This argument was communicated to the author by Professor P.K. Lam.

The returning signal is

$$R(t) = \Gamma |T(t)\rangle = \int_{-\infty}^{\infty} T(\tau)\ [\Gamma\delta(t-\tau)]\ d\tau \tag{2}$$

where I have used the linearity of the operator Γ to bring it into the integral. Then it is sufficient to define the action $\Gamma\,\delta(t-\tau)$ of reflection on an impulse.

Imagine a pulse transmitted from body A, assumed to be at rest, at time τ. Since it propagates at the speed of light, the distance of the pulse from A at any later time is

$$x_{pulse}(t) = c(t-\tau) \tag{3}$$

a fact which only depends on the postulate of the constancy of the speed of light. The position of the body varies linearly with time, which is a statement of the fact that we are restricting our discussion to uniformly moving bodies.

$$x_B(t) = vt + x_0. \tag{4}$$

We may assign to the constant x_0 the physical interpretation that it is the distance between the two bodies at the time t=0 according to the clock of the emitting body A. This constant has no absolute significance, since it can be made to assume any value by a suitable redefinition of the origin of time. Of course, when there are two or more pulses, the differences between these distances takes on absolute significance, and this leads to the observable transformation of signals.

The time t_R of reflection is obtained by equating the position of the pulse with the position of the reflecting body

$$c(t_R - \tau) = vt_R + x_0$$
$$t_R = \frac{c\tau + x_0}{c - v}.$$

So the time for the pulse to propagate from transmit at body A to reflection at body B (according to a clock comoving with body A) is

$$\Delta_{1/2} = t_R - \tau = \left(\frac{v}{1-v}\right)\tau + \left(\frac{x_0}{1-v}\right) \tag{5}$$

Now body A is, according to our arbitrary convention, the body at rest. Therefore the entire delay incurred is, according to the constancy of the speed

of light for the outgoing and return trips, twice the one-way delay $\Delta_{1/2}$ above, which gives the needed two-way transformation on the impulse

$$\Gamma\,\delta(t-\tau) = \delta(t-\tau-2\Delta_{1/2})$$

Substituting for $\Delta_{1/2}$ from equation 5 above, and combining the coefficients of like terms, we have

$$\Gamma\,\delta\left(t-\tau\right) = \left(t-\left(\frac{c+v}{c-v}\right)\tau - \frac{2x_0}{c-v}\right) = \delta\left(t-\frac{1}{\lambda^2}\tau - 2\Delta\right) \qquad (6)$$

This result may now be substituted back into Equation 2:

$$R(t) = \int_{-\infty}^{\infty} T\left(\tau\right)\delta\left(t-\frac{1}{\lambda^2}\tau - 2\Delta\right)\,d\tau$$

which upon making the substitution $u = 2\Delta + \tau/\lambda^2$ gives

$$R(t) = \int_{-\infty}^{\infty} T\left(\lambda^2\left(u-2\Delta\right)\right)\delta\left(t-u\right)\,\lambda^2 du = \lambda^2 T\left(\lambda^2\left(t-2\Delta\right)\right) \qquad (7)$$

We will often ignore the constant multiplier λ^2. The justification is that we consider the signals as vectors in a Hilbert space, therefore multiplying one by a constant reproduces the same vector. It is possible to adjust the definition of the operator in such a way that it preserves the norm of the ingoing vector. The required condition on the two-way operator Γ is then

$$\langle R \mid R\rangle = \langle T \mid T\rangle = \langle T|\,\Gamma^\dagger\Gamma\,|T\rangle$$

which can only hold for arbitrary transmit signal $|T\rangle$ if $\Gamma^\dagger\Gamma = 1$, that is, if Γ is unitary. The proper normalization to satisfy this requirement is easy to find. Let

$$\Gamma\,s(t) = A\,s(\lambda^2(t-2\Delta))$$

$$\begin{aligned}
\langle s|\,\Gamma^\dagger\Gamma\,|s\rangle &= A^2\int_{-\infty}^{\infty} dt\,\left|s\left(\lambda^2\left(t-2\Delta\right)\right)\right|^2 \\
&= \frac{A^2}{\lambda^2}\int_{-\infty}^{\infty} du\,\left|s\left(u\right)\right|^2 \\
&= \frac{A^2}{\lambda^2}\,\langle s \mid s\rangle
\end{aligned}$$

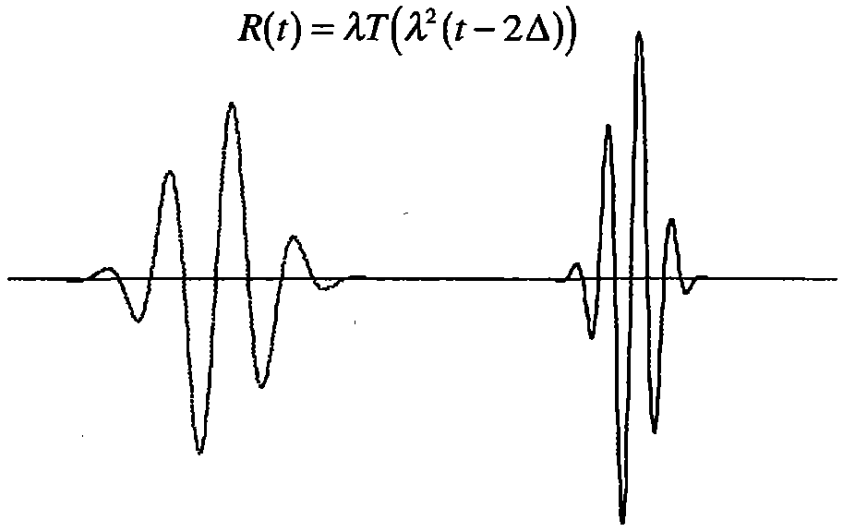


Figure 1: *A wave packet is transformed by reflection from a body approaching at near the speed of light. The transformation consists of a contraction of the time scale and a delay. If the body were receding, the returning packet would be linearly dilated or red shifted.*

therefore $A = \lambda$ is the normalization. We have thus shown that the transformation of an arbitrary signal upon reflection from a moving body, normalized to preserve the magnitude of the signal, is

$$R(t) = \Gamma T(t) = \lambda T\left(\lambda^2 (t - 2\Delta)\right) \tag{8}$$

with the two constants λ^2 and Δ given by

$$\lambda^2 = \frac{c - v}{c + v} \tag{9}$$

$$\Delta = \frac{x_0}{c - v} \tag{10}$$

Figure 1 illustrates the transformation of a signal when it reflects from a body moving near the speed of light. The wave packet is delayed in time and contracted (or dilated) by the factor λ^2 as indicated in Equation 8. The usual classical approximation to the transformation is that it consists of a frequency shift and a delay.

To connect with the conventional kinematical description, the position of the body is given for all time by the values of these two constant parameters

$$x_B(t) = \left(\frac{1-\lambda^2}{1+\lambda^2}\right) ct + \frac{2\lambda^2 c\Delta}{1+\lambda^2} \tag{11}$$

which, along with Equation 4, completes the connection between the signal description and the space-time description of motion along the radial line joining a pair of bodies.

It is good to point out the arbitrariness of the constant Δ for the case of a single signal exchange involving two bodies. It may be made to assume any value by an appropriate choice of the time $t = 0$, however this choice can only be made once. Any additional observers must then synchronize their clocks; They cannot choose Δ arbitrarily.

The physical elements of the scheme are the observer A (or B) and his own proper clock from which the time t is reckoned, and against which the two signals are measured. This much is accepted uncritically for now. So we can think of the two signals $T(t)$ and $R(t)$ as being available to either observer. The definition is in agreement with experience for bodies in uniform motion.

The relationships above specify the connection between signal transformations and the standard kinematical framework. Notice that, whereas the fundamental variable λ in the present scheme ranges from zero to infinity, meaning from infinite redshift to infinite blueshift, respectively, the quantity v ranges from 1 to -1 (in units of c), which is sensible. If the parameter λ is greater than one, a sinusoidal signal will return as a sinusoid with higher frequency, hence this characterizes a body approaching. Likewise, λ between zero and one characterizes a body receding. The condition $\lambda = 1$ indicates relative rest.

Figure 2 shows the relationship between λ and v. As we will see later, the doppler factor for collinear motion is multiplicative under composition, so a useful parameter is its logarithm, which is additive under composition. Indeed the quantity $u = -\ln(\lambda)$ is precicely the 'rapidity' parameter $c\tanh(u) = v$ which is sometimes introduced for just this reason in advanced treatments of special relativity theory[4].

The use of real-valued time functions as signal representatives is arbitrary in a certain sense. From a general point of view, these are representations of

[4]e.g. R.P. Feynman, *The Theory of Fundamental Processes*, Benjamin, Reading MA (1961), p.23

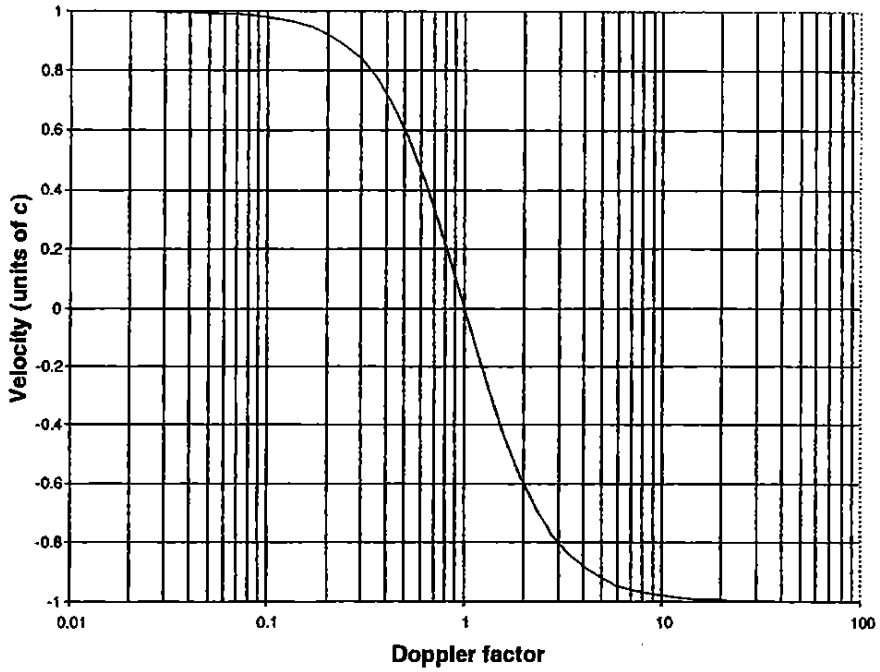


Figure 2: *The relation between the doppler factor λ and the velocity v. doppler factor approaching zero represents a large redshift, indicating a body departing at velocity +c. doppler factor approaching infinity represents a large blueshift, indicating a body approaching at velocity -c.*

continuous information, and the choice is not unique. For example, vector signals of any dimension could be employed. Real-valued signals are apparently the simplest objects which will accommodate the present problem. From the point of view of electromagnetism, is natural to look to polarized signals. This connection will be investigated further in the future, as we attempt to extend this theory to three dimensions. The use of continuous time and, perhaps more important, signals of arbitrary bandwidth, must ultimately come into question. But these assumptions adequately reflect the classical view. They will be accepted uncritically in the present paper. It should be clear that the dynamical interaction between the signal and the body is ignored, or else the concept of uniform motion would have no definite meaning. This can be roughly understood to mean that the masses of all bodies are large in comparison with the momentum (in natural units) carried by the signals.

The quantity Δ plays a central role in this theory, as one of the parameters of the fundamental signal transformation. It is good to try to develop a correct physical sense and consistent interpretation for this quantity. The situation is not quite as straightforward as it is with the doppler factor λ; There we have Equation 9 which connects it with the relative velocity in a simple way. The analogous formula for Δ is $x_0 = (c-v)\Delta$, where x_0 denotes the position of the moving body at time $t = 0$. The case of relative rest is easy to interpret. For that case $x(t) = \Delta$ (a constant), so Δ is just the distance between the bodies. When there is relative motion, we can say that Δ gives the distance between the bodies at the (arbitrary) time $t = 0$, but only indirectly because of the factor $(c - v)$. This does not provide a very satisfactory mental picture of the quantity Δ.

For a more physical interpretation, notice that

$$x_B(\Delta) = v\Delta + (c - v)\Delta = \Delta.$$

Therefore Δ represents the displacement between the two bodies at the moment when a light signal, emitted by body A at time $t = 0$ reaches body B. That is why, for the case of two bodies moving along a line, Δ may be made to assume any value by shifting the arbitrary origin of time. In particular, it is reasonable to choose as the origin of time the unique moment at which both bodies coincide. In this respect the case of relative rest is an exception, since for this case there is no time at which the bodies coincide, and Δ is nothing but the (invariant) proper distance between the two bodies.

This shift of the origin of time can only be taken once, therefore when several bodies are involved, the factor Δ for any pair of bodies can be set to zero without any loss of generality, but then the Δ-factors for the remaining pairs are fixed. This situation is easy to understand: To completely specify the motions of several particles it is necessary not only to give their trajectories but also their relative phases along those trajectories.

3 Abstract Theory of Signal Operators

We concentrate on a hypothetical world which consists of three particles in linear uniform motion, exchanging scalar signals. The example is illustrative. At this stage of complexity, the *world operator*, defined as the complete operator which carries a signal transmitted into one received, already exhibits its main characteristics. It is given formally by an infinite sum over paths

of products of elementary two-point transfer functions depending on a finite number of kinematical parameters. In this section these one-way operators are introduced only as formal concepts.

We introduce two-way operators, the parametric form of which has already been found, and one-way operators which are their factors, since these one-way operators turn out to be the atomic elements in terms of which the world operator or sum over paths is expressed. Finally, we discuss the observability of the one-way operators. We observe that they are all observable in the general sense that we have parametric forms for them, the parameters of which are directly observable. Then every observer can obtain the parameters of any labeled operator by a combination of direct measurement and calculations.

3.1 Elementary two-way operator symbols

We begin by analyzing the simple case of two particles in uniform radial motion. If the particles are labeled as A and B, then we write

$$R^A(t) = \Gamma^A_{ABA} T^A(t) \tag{12}$$

$$R^B(t) = \Gamma^B_{BAB} T^B(t) \tag{13}$$

The labeling of the operators is explicit to the point of considerable redundancy. This redundancy anticipates further developments in notation, but for now the main consideration is to provide precise physical meanings for each of the symbols. The symbols $R^A(t)$ and $T^A(t)$ represent, respectively, scalar signals transmitted and received by particle A. They are functions of time as recorded by a regular clock which is at relative rest and colocated with body A. The same interpretation applies to the signals labeled by B.

The operator symbols Γ^A_{ABA} and Γ^B_{BAB} represent transformations which connect the transmitted and received signals. Since there are only two bodies, any returning signal may be considered to be the reflection from the other body of what was transmitted. That is the interpretation of the subscripts ABA and BAB. They denote the *path* or sequence of bodies visited by the signal. Our convention is to read the path from right to left (sinistrally), but of course these two example sequences are the same in either order. The signals are represented mathematically by functions of time. It is convenient to restrict our representations to the class of square integrable time functions,

that is to the set

$$L^2 = \left\{ f(t) : \int_{-\infty}^{\infty} f^2(t)\, dt < \infty \right\}.$$

This restriction has the meaning that the total energy in each of the signals is finite, which is reasonable from a physical point of view.

The upstairs index is meant to indicate the coordinate system of the observer, which in the present case means the clock. For now, this is the same as the initial and final letter of the path, since we have only introduced the idea of round trip measurements made by a single observer with his own clock.

3.2 One-way operator symbols

We consider a simple world consisting of three particles labeled {A,B,C} moving uniformly along a line. Each of these particles is characterized by a transmitted signal, which we may denote as $|A\rangle$, $|B\rangle$ and $|C\rangle$ respectively, and by a returning signal which is a sum over a countable infinity of paths of the same signal transformed by a series of reflections from the other bodies. The aim is to exhibit the operator which connects the transmitted signal to the received signal for one of the particles. It is reasonable to call this operator the *world* to the given body, since it contains all the information available to the body with respect to its external influences.

From physical considerations, it is clear that the operator in this instance is a sum of terms, each of which depends on just three kinematical parameters. The inverse problem of deducing these parameters from the signals is investigated. Finally, some of the results are extrapolated to the situation with arbitrarily many bodies moving in a line.

We introduce a simple graphical representation of the exchange of signals among bodies, and establish a convention for using symbols to denote the elements such graphs.

Figures 3 and 4 are signal flow graphs [5]. In the first case, the three bodies are drawn in a straight line, which reminds us that the motion is all along a line. It is an aid to visualization. The second form of the diagram is better conceptually because it emphasizes the formal symmetry among the particle labels. Interpreted as signal flow graphs, the two diagrams are equivalent.

[5] Yutze Chow and Etienne Cassignol *Linear Signal Flow Graphs and Applications*, John Wiley and Sons, New York (1962).

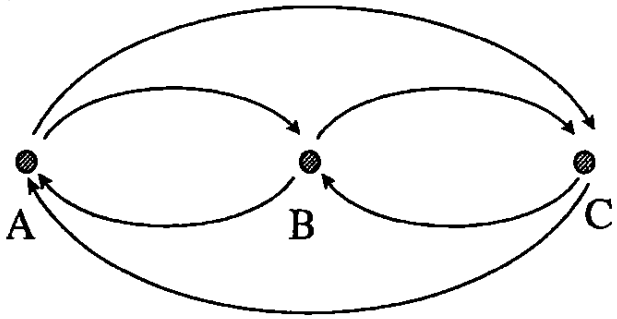


Figure 3: Signal flow diagram for three bodies. In the first figure the bodies are arranged along in a line.

Each dot or vertex represents a particle, and each arrow represents an operator or transformation which connects a signal transmitted at one body to the corresponding signal received at the other body.

There are exactly six arrows in the figure, so it may seem sensible that there are six distinct signal transformations or operators, each of which may be labeled by its two endpoints. But this method of counting operators turns out to contain a subtle error, which has its origin in the original signal concept.

A signal has two essential aspects. First, it is transmitted or received *by* a particular body, therefore it must be distinguished by the index of that body. Second, it is a function of a clock which is associated with some particular body. Then the signal function depends implicitly on the clock which is used to measure it, and we need to include an additional index to designate the clock or reference system employed. It would be possible to attach the index of the clock to the time argument, designating one of the arrows or operators in the signal flow diagram, for example as $\Gamma_{AB}(t_C)$ to indicate the transformation of a signal passing from A to B with reference to the clock of C. We utilize a different convention by attaching the index of the reference body to the operator as a superscript, writing $\Gamma^C_{AB}(t)$. The difference in point of view is like the distinction between the various pictures in quantum mechanics. The viewpoint we adopt is most similar to the Heisenberg picture, since the kets or signals $|A\rangle$, $|B\rangle$ and $|C\rangle$ are considered to be commensurate functions of a universal parameter t, labeled only by the transacting body,

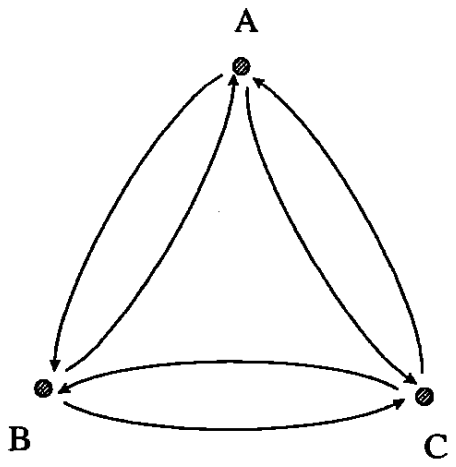


Figure 4: Signal flow diagram for three bodies in a more symmetrical arrangement. The diagram is equivalent to the previous one, but it emphasizes the general feature of symmetry among the three labels.

whereas the clock dependence is carried by the operators.

So we have a system for denoting the signals, and also the arrows or operators which appear in the signal-flow graphs. In the present case of three bodies, there are 18 elementary operators $\Gamma^{\omega}_{\mu\nu}$ which will form a basis for all operators which have a physical significance. The elementary operators are elements of the 2-parameter $ax+b$ or affine group of linear signal transformations. They are evidently not all independent, since we need 36 parameters to specify them, whereas only three are needed to fix the physical situation. *The complete set of relationships which exist between the operators constitute the kinematical theory in the present operator context.*

3.3 Operators Representing Paths

We denote a signal path by a sequence of letters. For example ABCBA denotes a signal which visits body A, then B, then C, then B then returns to A. The sequence of letters or character string denoting the path should be read from right to left, or sinistrally. The complete transformation taken

over the path is denoted by the operator symbol with the character string as a subscript, hence for example

$$\Gamma^X_{ABCBA} = \Gamma^X_{AB}\Gamma^X_{BC}\Gamma^X_{CB}\Gamma^X_{BA} \tag{14}$$

We can state the foregoing definition succinctly by writing

$$\Gamma^A_{\Phi B\Psi} = \Gamma^A_{\Phi B}\Gamma^A_{B\Psi} \tag{15}$$

where A and B are letters and Φ and Ψ are letters or sequences of letters. This definition allows us to express the operator for any path in terms of the elementary two-body operators which have already been introduced. It is now necessary to construct a list of character strings which corresponds to the set of signal paths.

This is easy to do because, aside from being signal flow diagrams, Figures 3 and 4 may equally be interpreted as the graphs, called finite state diagrams, which are used to represent Markov processes in discrete communication theory[6]. In such diagrams, the vertices distinguished with letters represent states, and the arrows represent transitions between states, to which are ordinarily assigned transition probabilities. In the context of communication theory, the purpose of the graph is to assign probabilities to all allowable character strings. They are convenient for our purpose since the character strings they generate are precisely the ones we require for labeling the paths. For the present case of three states with transitions as indicated, we require all the strings beginning and terminating with the designated observing body, say A, and for which no letter ever appears twice in a row (such a repetition would appear in the state transition diagram as a loop beginning and ending with the same body, so these transitions are not allowed by the diagram as drawn). The required strings are the set

$$\Phi = \left\{ \begin{array}{cccc} ABA & ABCA & ABCBA & ABCBCA\ldots \\ ACA & ACBA & ACBCA & ACBCBA\ldots \end{array} \right\} \tag{16}$$

We are now in a position to express the world operator from the point of view of our representative body A.

$$\Omega\,|A\rangle = \sum_{\phi\in\Phi} \Gamma^A_\phi\,|A\rangle \tag{17}$$

[6]Claude Shannon "The Mathematical Theory of Communication", Bell System Technical Journal, July 1948

Here Ω is the required operator which carries an outgoing or transmitted signal into an ingoing or recieved signal: A coherent sum over paths of transformed signals as stated. The set of character strings or sequences is the one given explicitly above, and the operator with a string for a subscript is to be interpreted as a product of elementary two-point operators according to Equation 15. Then the world operator has been given explicitly in terms of the 18 two-point operators. It should be noted that the formalism is quite general. For example, the exact nature of the 2-point operators will certainly depend on the assumptions of uniform linear motion, and the set Φ of subscripts as given by Equation 16 depends on the assumption that there are only three bodies, but the form of Equation 17 does not depend on any of these simplifying assumptions.

3.4 Observability of Operators

The physical interpretation of the operator kinematics requires that the elementary operators (18 operators $\Gamma^{\omega}_{\mu\nu}$ in the present restricted case) have physical definitions. They must be observable, and indeed physically available to each of the observers. We have stated that even these 18 operators are not independent, but must be determined, for physical reasons, by the kinematical parameters of the problem, which are the velocities of bodies B and C relative to body A, and the position of body C along the line AB at the moment when A and B coincide, or else some equivalent set of three independent quantities. Therefore one way to settle the question of observability is constructively, by exhibiting an algorithm for obtaining the physical parameters mentioned from knowledge of the world operator, or from a transmitted and a received signal.

In the present restricted problem of linear uniform motion, we are helped by the knowledge that the received signal consists of a sum of a countable infinity of terms, each of which is a scaled and delayed copy of the transmitted signal. Then the (continuous) wavelet transform of the returned signal $\Omega|A\rangle$ is not a continous function in scale and delay but a sum of delta functions, analogous to a discrete line spectrum in Fourier analysis. This statement holds true for any finite number of particles in uniform relative motion along a line.

The wavelet transform of a recieved signal $R(t)$, analyzed in terms of transmitted signal $T(t)$ as the basic wavelet, provides a decomposition of the signal function into a sum of scaled and delayed copies of the transmitted

signal. Then this decomposition is exactly what is required to extract kinematical parameters from the world operator. The wavelet transform is given by Ingrid Daubechies[7] in the form

$$\phi_{Wav,f}(a,b) = \frac{1}{|a|^{1/2}} \int dx\, h\left(\frac{x-b}{a}\right) f(x) \tag{18}$$

In this conventional notation, the parameter a is called the scale factor and the parameter b is called the delay. The function $h(x)$ is called the basic wavelet, and for this one we are of course choosing the transmitted signal function. Our assertion regarding the discrete character of the wavelet transform is then

$$\phi_{Wav,R}(a,b) = \frac{1}{|a|^{1/2}} \int dt\, T\left(\frac{t-b}{a}\right) R(t) = \sum_i c_i \delta(a-a_i)\,\delta(b-b_i) \tag{19}$$

where the parameters (a_i, b_i) appearing on the right constitute a discrete and countable set of spectral features in the (a,b) plane of scale and delay. The formal analogy with the analysis of atomic line spectra may be carried further. The different scales a_i occuring in the wavelet transform of the signal will be all the products of a set of doppler factors corresponding to the relative velocities of the particles. If we work in terms of the logarithms of these doppler factors (conventionally called *rapidities*) then the problem of obtaining these basic parameters from the set (a_i, b_i) of features in the wavelet transform is only the problem of finding a finite set of terms, given the list of all their sums, which is elementary. The analogous problem in atomic theory is to deduce the energy levels of a Bohr atom from its spectrum, and this is likewise simple.

As to the question of the observability of the operators in the present kinematical theory, we can now answer it. The set of elementary kinematical parameters are available to any observer by using the wavelet transform as just described. Therefore, provided formulas can be found which relate the parameters of these operators to those of observable round-trip operators, we may say that all of the operators in the theory are observable either directly or indirectly, through explicit calculations. These formulas will be given.

An object such as Γ^C_{AB} which describes the transformation of a signal passing from A to B as referred to the clock carried by a third external

[7] I. Daubechies, "Orthonormal bases of compactly supported wavelets", Comm. Pure & Appl. Math. **49** (1988) 909-996.

observer C appears at first to suffer from the logical difficulty that it admits no operational definition. This is not the case. For example, observer C has the operators Γ^C_{CAC}, Γ^C_{CBC} and Γ^C_{CABAC}, all of which correspond directly to terms in the wavelet decomposition of his returning signal. He can easily obtain the needed operator by making the required operator products, as

$$\Gamma^C_{ABA} = \left(\Gamma^C_{CA}\right)^{-1} \Gamma^C_{CABAC} \left(\Gamma^C_{AC}\right)^{-1}$$

We have concentrated on the case of three bodies moving uniformly along a line, since it is easy to visualize and it expresses most of the important features of the N-body problem. The result expressed by Equation 17 applies to the case of N particles as long as we interpret the set Φ of character strings as being generated by the corresponding finite state diagram, as indicated. The restriction to uniform linear motion can be removed, but then the two-point operators are no longer affine transformations, so the analysis by means of the wavelet transform is not correct. However, the form of Equation 17 still holds. This case will be investigated in Chapter 4.

4 Algebra of Γ-Operators

It is convenient to focus on the class of operators which act by inducing a linear transformation on the argument an ingoing function, since this is precisely how our two-way operators act, according to Equation 8. These operators are known in the theory of wavelets[8] as affine or $ax + b$ transformations.

We will introduce the affine operators with the following notation

$$\Gamma(a,b) f(t) \equiv \sqrt{a} f(a(t-b)) \tag{20}$$

The fact that the letter Γ has been overloaded with a second meaning will not cause any confusion, since the appearance of the symbol with subscripts and superscripts always indicates a kinematical operator as defined by equation 12, whereas the appearance of the symbol with two arguments implies a generic affine operator as in Equation 20. The usefulness of this choice of symbols is that a Γ-operator, whether it appears in the first form or the second, is recognizable as an element of the set of affine transformations. These operators form a two parameter linear group. The following useful properties about affine operators are easily proved by iterating Equation 20:

[8]A. Grossman and J. Morlet,"Decomposition of Hardy functions into wavelets of constant shape", SIAM J. Math. Anal. **15** (1984) 723-736.

- Product Rule

$$\Gamma(a',b')\Gamma(a,b) = \Gamma\left(aa', b + \frac{b'}{a}\right) \tag{21}$$

- Inverse

$$\Gamma^{-1}(a,b) = \Gamma\left(\frac{1}{a}, -ab\right) \tag{22}$$

- Square Root

$$\sqrt{\Gamma(a,b)}_{+} = \Gamma\left(\sqrt{a}, \frac{b\sqrt{a}}{1+\sqrt{a}}\right) \tag{23}$$

- Triple Product

$$\Gamma(a'',b'')\Gamma(a',b')\Gamma(a,b) = \Gamma\left(a''a'a, \frac{b''}{aa'} + \frac{b'}{a} + b\right) \tag{24}$$

Operator multiplication is associative but not commutative, and the set of operators forms a group. In fact, the restricted set $\{\Gamma(a,b) : a \geq 0\}$ with non-negative scale factor forms a subgroup which is sufficient for our discussion of relativistic kinematics. We do not consider the possibility of negative scale factors.

5 Contraction of a Moving Rod

The Lorentz contraction of a moving rod is a relationship among observable quantities, that is, among the parameters of the elementary round trip operators which have already been introduced. This leads to an attempt to calculate the formula without reference to any auxilliary quantities or operators, and in particular without the introduction of the variable upper index representing a clock or reference system.

As shown in Section 2 two-way operator has the following general form, which employs the two kinematical parameters (λ, Δ)

$$R^A(t) = \Gamma^A_{ABA} T^A(t) = \lambda T^A\left(\lambda^2(t - 2\Delta)\right) \tag{25}$$

The kinematical parameters are related to the relative speed v of the two bodies and to the distance x_0 between them at time t=0 by

$$v = \frac{1-\lambda^2}{1+\lambda^2} c \tag{26}$$

and

$$x_0 = \frac{\lambda^2 \Delta c}{1+\lambda^2}. \tag{27}$$

So the connection between our new kinematical parameters (λ, Δ) and the conventional description of motion by means of a linear trajectory is

$$x(t) = vt + x_0 = \frac{1-\lambda^2}{1+\lambda^2} ct + \frac{2\lambda^2 c\Delta}{1+\lambda^2} \tag{28}$$

To investigate the moving rod, consider three bodies, A and B the two ends of a rod of rest length l, and C a third body moving with speed v relative to the other two. We are free to define the origin of time once for the problem as a whole. A convenient choice is to set the zero of time at the moment when A and C coincide. This condition is an expression of the synchronization of the two clocks. Of course, it is not strictly necessary to set the moment of coincidence as the zero point of time. More generally the two clocks could be set ahead or backward by an arbitrary fixed interval of time and still be synchronized. The choice is convenient here because it simplifies the appearance of the operators. Writing them out explicitly

$$\Gamma^A_{ACA} = \Gamma^C_{CAC} = \Gamma\left(\lambda^2, 0\right) \tag{29}$$

in which the parameter λ is related to velocity v by Equation 26.

For a signal transacted between the two (relatively stationary) ends A and B of the rod

$$\Gamma^A_{ABA} = \Gamma^B_{BAB} = \Gamma\left(1, 2l\right) \tag{30}$$

Now the operator $\Gamma^C_{CBC} = \Gamma^B_{BCB}$ determines the distance (as measured by moving observer C) to the far end of the rod, whereas the operator Γ^C_{CAC} which is already explicitly given, determines the distance to the near end of the rod as measured by observer C. Their difference is the relativistic length of the rod. So it is reasonable to ask whether the operator Γ^C_{BCB} can be written down as a function of the other two known operators.

To do it, define a new operator Π_{AC} by the equation

$$\Gamma^C_{CBC} = \Pi_{AC} \Gamma^A_{ABA} \Pi_{AC} \tag{31}$$

A special case is one in which A and B are the same body. From this particular case, we have

$$\Gamma^A_{ACA} = \Pi_{AC} \Pi_{AC} = \Pi^2_{AC}$$

then

$$\Pi_{AC} = \sqrt{\Gamma^{A}_{ACA}} \tag{32}$$

but the operator Π_{AC} can only depend on the kinematical parameters connecting A and C, and we conclude that Equation 32 is true for general Γ^{B}_{ABA}. In particular, we have

$$\Gamma^{C}_{CBC} = \sqrt{\Gamma^{A}_{ACA}}\Gamma^{A}_{ABA}\sqrt{\Gamma^{A}_{ACA}} \tag{33}$$

which is the needed relationship. The operators on the right are given explicitly by Equations 29 and 30. Substituting

$$\Gamma^{C}_{CBC} = \Gamma(\lambda, 0)\,\Gamma(1,\, 2l)\,\Gamma(\lambda, 0) = \Gamma\left(\lambda^2, \frac{2l}{\lambda}\right) \tag{34}$$

Now according to Equation 28, the position of body A at any time is given by

$$x_A(t) = vt = \frac{1-\lambda^2}{1+\lambda^2}t$$

The position of body B at any time is given by

$$x_B(t) = \frac{1-\lambda^2}{1+\lambda^2}t + \frac{2\lambda^2}{1+\lambda^2}\frac{l}{\lambda}$$

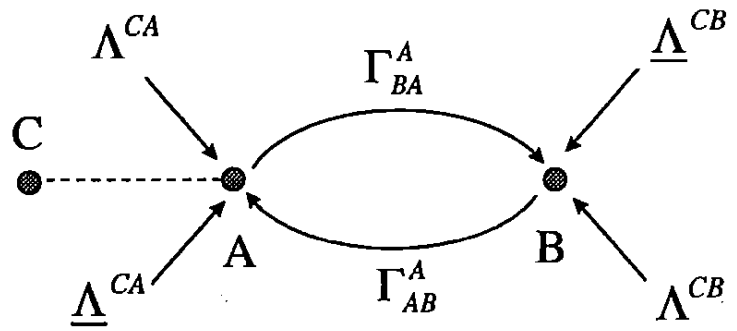
and their difference, by definition the length of the moving rod as measured by body C, is

$$l' = x_B(t) - x_A(t) = \frac{l}{\frac{1}{2}\left(\frac{1}{\lambda}+\lambda\right)} = \frac{l}{\gamma}.$$

where

$$\gamma = \frac{1}{2}\left(\frac{1}{\lambda}+\lambda\right) = \frac{1}{2}\left(\sqrt{\frac{c+v}{c-v}} + \sqrt{\frac{c-v}{c+v}}\right) = \frac{1}{\sqrt{1-v^2/c^2}}.$$

So we have obtained the contraction of a moving rod and found the correct Lorentz factor without the introduction of any auxilliary operators, by taking advantage of simplifying assumptions that were appropriate to the thought experiment at hand (namely the simplified synchronization condition). In the next section, we connect the operators which appear in this thought experiment to those that appear in a formulation of the theory in terms of one and two-way signal transformations. We demonstrate the connection between the operator symbols in each approach, and derive time dilation and the one-way (relativistic) doppler effect.



$$\Gamma^{C}_{ABA} = \underline{\Lambda}^{CA}\Gamma^{A}_{AB}\Lambda^{CB}\underline{\Lambda}^{CB}\Gamma^{A}_{BA}\Lambda^{CA}$$

Figure 5: *The transaction of a signal between bodies A and B is observed using the clock or coordinate system belonging to C. We denote by Λ^{CB} the rescaling of the time coordinate which is required to make the two clocks agree.*

6 Formalism of Coordinate Transformations

Round trip signal transformations like Γ^{A}_{ABA} were introduced and connected to the conventional kinematical picture by means of explicit formulas (Equations 8,9 and 10). A whole formal or notational scheme was then introduced, with indices up (representing clocks or coordinate systems)and down (representing paths). The proposition that the operator representing a path could be written as a product of operators representing legs in a graph led to more exotic operator symbols like Γ^{C}_{AB}, meaning the one-way transformation on a signal propagating from A to B, but measured with the clock of C, but we have not shown how to construct the parametric forms of such operators in terms of kinematical parameters.

Figure 5 depicts a transaction between bodies A and B, but measured using the clock of body C. Any clock adjustment is an affine transformation, since it consists of equalizing the rates and possibly resetting the zero (synchronizing) the two clocks. Then it amounts to a delay (or advance) combined with a rescaling of the time coordinate. We may formally intro-

duce a new kind of object Λ^{CA} called a coordinate transfer operator, which has the property of changing the upper index, which means resetting the rate of a clock. The notation is that Λ^{CA} converts the clock of A to the clock of C. Formally, then, the action of the clock or coordinate transfer operator on a round trip is

$$\Gamma^{C}_{ABA} = \Lambda^{CA}\Gamma^{A}_{ABA}\Lambda^{AC} \tag{35}$$

Indicating that body C must switch into the coordinates of body A, observe the transaction, and then switch back into his own system of coordinates. This definition of the coordinate transfer operator formalizes the idea of the upper index in the way we have been using it, namely that it indicates the clock or coordinate system. The new operator must be symmetric in the two indices. If there is any clock operation to be taken, it must be the same going from A to C as from C to A, otherwise we have an observable fact which distinguishes the absolute state of motion of one of the bodies.

Figure 5 illustrates the concept of the Λ-operator construction in analogy with a Feynman diagram. From the viewpoint of body A, the operators Λ^{CA} and $\underline{\Lambda}^{CA}$ are vertex factors to be associated with the outgoing and ingoing signal lines, respectively. They transfer reference between the clocks of A and C. The operators Γ^{A}_{BA} and Γ^{A}_{AB} are propagators which carry the signal from A to B and from B to A, from the reference point of body A.

To find expressions for the Λ operators, consider A and B to be the ends of a rigid rod as before.

$$\Lambda^{CA} = \underline{\Lambda}^{CB} = \Gamma(\alpha, 0) \tag{36}$$

$$\underline{\Lambda}^{CA} = \Lambda^{CB} = \Gamma(\alpha', 0) \tag{37}$$

where α and α' are unknowns. Reading off the operator factors from the diagram above, we have

$$\Gamma^{C}_{ABA} = \underline{\Lambda}^{CA}\Gamma^{A}_{AB}\Lambda^{CB}\underline{\Lambda}^{CB}\Gamma^{A}_{BA}\Lambda^{CA} \tag{38}$$

Alternately, we can write the coordinate transformation of the round trip operator

$$\Gamma^{C}_{ABA} = \underline{\Lambda}^{CA}\Gamma^{A}_{ABA}\Lambda^{CA} = \underline{\Lambda}^{CA}\Gamma^{A}_{AB}\Gamma^{A}_{BA}\Lambda^{CA} \tag{39}$$

Comparing Equation 38 and Equation 39, we have

$$\Lambda^{CB}\underline{\Lambda}^{CB} = 1$$

$$\begin{aligned} \Lambda(\alpha\alpha', 0) &= \Lambda(1, 0) \\ \alpha' &= 1/\alpha \end{aligned}$$

The outcome of the previous section was to establish the Lorentz contraction of the moving rod. But this provides us with the formula for the two-way operator according to body C

$$\Gamma^{C}_{ABA} = \Gamma\left(1, \frac{2l}{\gamma}\right) \tag{40}$$

Using the forms of the Λ-operators

$$\Gamma^{C}_{ABA} = \underline{\Lambda}^{CA}\Gamma^{A}_{ABA}\Lambda^{AC} = \Gamma\left(\frac{1}{\alpha}, 0\right)\Gamma(1, 2l)\,\Gamma(\alpha, 0) \tag{41}$$

Finally, we can equate the right hand sides of Equations 40 and 41 to get the explicit parametric forms of the clock or coordinate transfer operators, namely

$$\Lambda^{AC} = \Lambda^{CA} = \Gamma(\gamma, 0) \tag{42}$$

$$\underline{\Lambda}^{CA} = \underline{\Lambda}^{AC} = \Gamma\left(\frac{1}{\gamma}, 0\right) \tag{43}$$

Now the entire group of operators has been specified, because we have an explicit parametric form for every operator which appears in the theory. The paths or character strings which form the subscripts of the operators can be formed by concatenating or multiplying together smaller operators , and ultimately they can all be built up in this way from the atomic two-point operators. The upper or coordinate index of an operator is transferred with the help of the Λ-operators, and of course these are also given in terms of the relative velocity between a pair of bodies, through the Lorentz factor γ, as above.

7 Information Theory

In this paper, we systematically shifted the emphasis in the description of uniform motion from the geometrical trajectories of bodies to the parameters which describe the transformation of signals. Light signals were our basic model, but of course we really only used those features of light which

also model information in a more general sense. Signals are the central objects in the theory of information. This theory was created by Shannon in 1948 to handle problems related to telephone communication[9]. The theory is presented in a very lucid way in the famous monograph of Woodward on radars[10], a subject which has much in common with the present investigation. In the language of information theory, we chose the model of continuous information when we described our signals by means of functions of continuous time. This choice is reasonable from the point of view of classical physics. As we remarked, information may be represented by real or complex, scalar or vector signals, so we chose only the simplest case. The analysis of motion in three dimensions appears to depend on the larger machinery of complex spinor-valued (or polarized) signals, in which case the needed transformations are Pauli spinors or complex-valued quaternions. That analysis will be taken up elsewhere.

We did not go as far as we might have done with respect to the concept of time, as this was not necessary for the problem we wanted to address, but information theory is equipped with a very satisfactory method for handling physical time, through the analysis of the (information) rates of continuous and discrete processes. This means that within information theory, the concepts of time and signal are interrelated in a way which seems to be very attractive from the point of view of physics.

It seems to be possible to carry out the steps of the present analysis using a discrete model of information. For one obvious example, we can make an algorithm mapping continuous functions to discrete character strings. A more profound form of discreteness is provided by quantum mechanics, in which we may focus on the sequence of states visited by an atomic system. The sequence itself falls under Shannon's category of discrete information.

8 Summary

The kinematics of several bodies moving on a line was taken up from the point of view of light signaling. The advertized advantage of this method is that the dependence of the descriptive framework on exterior space and time is

[9]C. Shannon, *The Mathematical Theory of Communication*, Bell Systems Technical Journal, **27**, 379-423, July 1948.

[10]P.M. Woodward, *Probability and Information Theory, with Applications to Radar*, MacMillan, NY (1953).

greatly reduced once the transformation law on signals has been established. In this way one may hope to establish a manifestly local and operational version of kinematics which involves the signals alone. A study was made of linear motion, which is the case in which a transmitted signal is delayed and linearly dilated or contracted in time.

This program was carried through for the problem of several bodies moving on a line, in which case the operators were found to be elements of the 'affine' or $ax + b$ group of wavelet transformations. We started by demonstrating this with a Green's function method. Next, an abstract formalism for operators was built up, based on the notational principle of upstairs indeces representing clocks and downstairs character strings representing paths. Arguments were given with respect to the observability of the operators, with the conclusion that all of the operators in the theory could be either observed or calculated by any one observer if formulas could be found to connect the operators to a certain minimum set of parameters. This program was carried out, first by an analysis of the rigid rod, and then by a discussion of operators representing changes of clock, which could be related to the observable two-way operators in the theory.

It is hoped that the general method can be extended to two and three dimensions without all of its attractive features being lost. The analysis of this extension is ongoing. Additional difficulties already appear with just two bodies in general two-dimensional uniform motion which includes transverse motion, because the signal transformation is no longer of the simple affine variety. The solution appears to turn on promoting the signals, as is reasonable, from scalar to polarized. We hope to be able to report the success of this program in the near future.

Remarks on the Signal Approach to Relativistic Kinematics

P.D. Mountcastle
6382 Brown Circle
Huntington Beach, CA 92647
USA

1 Introduction

In the last paper, relativistic kinematics were recovered (for the case of N bodies moving along a straight line) from a world picture in which the observer, his clock and the signals at his transmit and recieve terminals were the only constructive elements. For this to constitute a physical theory, it must be possible to extend the ideas to three dimensions, and to consistently introduce the photonic nature of light and matter. It seems to be possible to do so by use of thought experiments which are direct extensions of those employed in the one-dimensional case.

2 Extension to Three Dimensions

The program to derive angular space from inner space, or what Einstein called extension, appears to be inconsistent with experience. The electron (and perhaps more significantly, because it so readily exhibits relativistic behavior, the muon), have no measurable extension and no internal structure. Yet they behave like three-dimensional mass points with regard to kinematics. This must lead us to try extending the signaling theory to three dimensions in a different way. The classical development which I presented in the previous paper [1] assumed the use of a real-valued scalar signal, and this was sufficient

[1]Mountcastle,P.D., *Operator Formulation of Relativistic Kinematics*, Proceedings of ANPA 21, September 1999.

to construct the relativistic theory for motion along a line. On the other hand, a classical electromagnetic signal is not scalar, on account of polarization.

In the most general case it is described as a Fourier superposition of harmonic components, each of which is in turn a coherent sum of right-circular and left-circular aspects. The following bit of notation may convey the sense of this classical generalization of the signal model.

$$|s\rangle = \int [c_+(\omega)\,|+,\omega\rangle + c_-(\omega)\,|-,\omega\rangle]\,d\omega = \int d\omega \begin{bmatrix} c_+(\omega) \\ c_-(\omega) \end{bmatrix} \qquad (1)$$

It is necessary to express the signals in the frequency representation in order to formulate the polarization degree of freedom correctly, since only a monochromatic signal can be in a pure state of polarization. The effect of reflection on a harmonic signal is a phase shift and a scaling of the frequency, so all of the operator formalism of the theory which has been expressed in the time domain can in principle be carried over to the frequency domain, but I have not yet worked out the translation.

Even so, it is useful to examine in some detail the analogy between the development of the kinematical theory from scalar signals in one dimension and the proposed development in three-dimensions, begining with polarized signals.

In the problem of motion along a line, which was successfully tackled, there were two alternative parametric descriptions. The first one was geometrical, with each of the bodies characterized by a pair of coordinates (v, t_0) denoting speed and the time of coincidence with the observer. Each body posessed a local degree of freedom (the zero of it's local clock) which could not affect any observed quantity, and thus played a role roughly analogous to the local gauge or phase symmetry in elecromagnetism. The second equivalentand complementary description was in terms of the transformation parameters $(\lambda^2, 2\Delta)$ for the signal reflected in a round trip. The analysis then consisted of two parts.

The first part was to establish the connection formulas between the two descriptive frameworks. This was possible because the case of a signal reflecting from a moving body and returning to the observer admitted a complete description in either framework. The mathematical technique of the Green's function or impulse-response function was employed, the delta-pulse being associated with a mechanical sounding particle with invariant velocity c.

The second part was to work out the complete theory in the new operator framework, and this consisted mainly of enumerating the world of kinematical

operators, and of explicitly showing how to construct various operators from knowledge of other operators.

In three-dimensional kinematics, a new complication arises. From the geometrical point of view, there is no coincidence of bodies, except in the degenerate case of radial motion. Instead, the body has a nearest approach or momentary perigee with the observer at a distance s which we may call the impact parameter. The direction of motion and the direction of a line joining the observer to the perigee point form a pair of orthogonal axes, which we may represent by unit vectors as $\mathbf{x}$ and $\mathbf{z}$, respectively. To this list we may append the unit vector $\mathbf{y} = \mathbf{z} \times \mathbf{x}$ normal to the plane of motion to define a cartesian frame of reference. The approach to geometry via the algebra of 3-dimensional vectors (the Clifford Algebra Cl_3, is ideally suited to describing this geometrical picture).[2]

In that way, there corresponds to every body B, from the viewpoint of an observing body A:

1. A time t_{BA} of closest approach (in exact analogy to the one-dimensional case). As before, this time is defined only up to an arbitrary constant (the zero of the local clock).

2. A distance s of nearest approach

3. A three-dimensional cartesian frame of reference which we may denote by $\mathbf{C}_{BA}$. To give this frame of reference a definite representation (for example with angles (θ, ϕ, ψ)), we may adopt the viewpoint that the observing body A has an *internal* frame of reference. This is analogous to the assumption that body A carries a clock, the zero time of which cannot affect the value of any observable quantity. In exactly the same, way, no observable quantity can depend on the orientation of the internal frame of reference associated with the observer.

We may summarize the description of uniformly moving bodies in the geometrical picture by listing the six quantities which are required to specify the motion of a body B from the point of view of an observing body A. They are the distance s_{BA} of nearest approach, the speed $v_{BA} = v_{AB}$, the time t_{BA} of nearest approach, and a set (θ, ϕ, ψ) of three Euler angles needed to specify the relative coordinate frame in internal coordinates.

[2]Baylis, William E., *Electrodynamics, A Modern Geometrical Approach*, Progress in Physics, Vol. 17, Birkhäuser, Boston MA (1998).

The description in the complementary operator picture is sketchier , because the forms of operators in the frequency representation and the other mathematical details have not been worked out. It appears that, in one dimension, the operator group which emerged naturally was the most general linear mapping of the space of scalar functions onto itself, which preserved the norm of vectors in Hilbert space (that is, signal power). That was the two-parameter affine group. The elements of the new Hilbert space with polarization are complex two-vector or spinor-valued functions of frequency. The most general linear transformation between two such spinor objects is given by a 2×2 complex matrix. If we normalize the power (in analogy to the normalization of the one-way operators), then the determinant of the matrices is 1, and we have operators belonging to the six-parameter linear group $SL(2,C)$, which has a well-known relation to the Lorentz group. The association problem in three dimensions must follow along the same general lines as in the one-dimensional case. The signal reflected from a body in uniform motion is an integral over frequency of a Green's function whose form will result from considering the response of the system to a polarized monochromatic signal. Apparently that will be another monochromatic signal with a modified frequency and phase. The mixing of the polarizations must involve the internal and closest-approach coordinate systems in such a way that the matrix elements of the resulting linear transformation become functions of the Euler angles and the impact parameter. The situation is rather harder to visualize than it was in the time domain, because a single frequency impulse response incorporates reflections from the moving body at all times. Nonetheless, the connection problem is essentially concrete and should be possible to work out along engineering lines, just as it was in the one dimensionional situation.

The logical thread which has been carried through is that the parameters of the final operator theory should be regarded as fundamental physical quantities; Indeed, the observer-centered operator picture is fundamental. So the complexity of the connection formulas, or the relative difficulty of finding them, have no real significance. The formulas for the six parameters form an bridge to a descriptive framework (geometry) which is, in that sense, incidental.

It is necessary to write all these things down and work them out in a careful way, as has been done for the one-dimensional problem. The thing I want to stress is that we are rather strongly guided by analogy in trying to go from the one-dimensional radial case to the three-dimensional case by

extending the signal model.

3 Quantization

The extension of the signal model to a classical polarized signal is motivated by correspondence with classical electrodynamics. A different kind of modification of the signal model is indicated when we consider the origin of photons in the transition of an atomic system between two stationary states of definite energy and angular momentum (that is, the Bohr atom). If the frequencies present in the signal are all of the form

$$f_{mn} = \frac{2\pi}{\hbar}(E_n - E_m) \tag{2}$$

with E_n a set of energies, then the frequency representation of the general signal is not a Fourier integral but a discrete Fourier transform taken at a non-uniform grid of frequencies. If the terms in such a series are arranged in a square array, it can be shown in the usual way (by correspondence) that the operators (here representing the action of bodies on signals) compose like matrices, and so on. Heisenberg's entire development of quantum kinematics appears to carry over without the need of much modification. On the other hand, it is usually suggested that matrix mechanics, being connected to the classical Hamiltonian formulation of mechanics, is not a relativistically covariant theory. On the other hand, the present kinematical operator theory, of which the matrix mechanics of Heisenberg is a special case, is relativistically covariant by construction.

4 Connection to Bit-String Theory

It is possible to define various satisfactory mappings between classical representations of signals as scalar time functions and information-theoretic representations of the same objects as long strings of zeros and ones (bit strings). The theories based on such representations have the property that, in the formal limit where a finite interval of time is represented by an infinite number of bits, they become identical to the corresponding classical continuous theories.

As a concrete example, one may pass real-valued signals through a filter which has enough memory to store two consecutive real signal values, mea-

sured at closely spaced points in time. The filter compares the two signal values and outputs the symbol (+) if the second signal value is larger than the first, and the symbol (-) in the other case. If the signal values are measured and stored to sufficient accuracy, it is clear that the third case (equality of the two signals) almost never occurs, so the representation becomes perfectly faithful in the limit of arbitrarily small time sampling. In another language, that is the ultraviolet limit of frequency. It seems that this discussion of a delicate limit process has some relation to the theory of renormalization, but it will not be necessary to enter into that analogy, since we are pursuing an entirely different path.

Obviously, the signal theory of relativistic kinematics can be expressed as a bit string theory of relativistic kinematics by a mechanical process of transforming the scalar signals to bit strings and the corresponding linear operators to digital filters. This process is not without intrinsic interest. For one thing, it seems to lay bare the root of classical relativistic effects in a framework which should be accessible to students of computer science and information theory. For another, a large body of existing mathematics could be employed to good effect.

The drawback in this construction is that the connection to classical continuous theory is entirely too close. The apparent quantization of the theory is factitious, resting ultimately on the division of time into equal infinitesimal steps, which serves to reintroduce the limit operations of the classical calculus in a different form.

The quantum theory which I am proposing rests on a description of material bodies as Bohr atoms, or in other terms, as discrete Markov processes. To carry it through, the roles of matter and radiation, and indeed the fundamental ideas of space and time must undergo a profound revision.

A DEMONSTRATION OF THE MULTIMEDIA COMPUTER PROGRAM **PlayGROWnd— A Learning Environment: The Evolution and Nature of Cognition**
(A Multidisciplinary Monograph on CD-ROM)

Hale Chatfield

Department of English
Hiram College
Hiram, Ohio 44234 USA
hchatfield@aol.com

Abstract

In exploring the development of human cognition and the nature of mind, the multimedia program *PlayGROWnd* describes and illustrates the important creative roles played by error and delay—and explores the nature and motivation of our own creative play in composing art, building robots, constructing intelligent systems, and fashioning scientific models and theories. In addition to providing a review of relevant literature and scholarship, the software features playing fields consisting of fertile environments in which the user can play with ideas and images; make discoveries; and record notes and observations.

The play of children and artistic adults is a confident and deliberate form of error-making in which pretense is given precedence over reality. Metaphor and mimesis, noteworthy for their usefulness in the processes of invention and discovery, share with art and children's play a willingness to subordinate representational truth, accuracy, or literal-mindedness. Metaphors are lies, forthrightly. Mimesis poses imitation as a valid alternative to actuality. Verbal errors like puns, at which we may wince or groan, nonetheless delight us with their revelations of surprising and unexpected truths.

PlayGROWnd revels in such matters, and is at pains to point out that biology adores errors and delays, too—depending upon them for evolution, in the replicative processes of DNA and in the chemical and electrical delays built into the operations of our nerves and synapses. In our cells and molecules and in our laboratories and studios we thrive on inevitable delays and useful accidents. Perhaps it is fair to say that for many of us a main business of life is the search for fortunate errors.

But fortunate errors are not likely to occur without activity. To stumble upon discoveries in this way, we must be in motion. An increasing recognition of this idea is evident in our ways of doing our science, producing our art, and even conducting our business (consider Tom Peters' "*Ready! Fire! Aim!*"). In all areas we can think of, people seem most effectively creative when they treat their lives as playgrounds: places to do stuff (almost any stuff, sometimes), fields in which to grow, to encounter, to take note of, and to replicate wonderful accidents.

Among the persons whose work and ideas are given specific consideration in *PlayGROWnd* are: Gaston Bachelard, Rodney Brooks, Martin Buber, Grant Gillett, Elizabeth Hart, Stephen Hawking, G. Spencer Brown, Norman O. Brown, Merlin Donald, Rom Harré, Laura Otis, Roger Penrose, Tom Peters, David Porush, Ilya Prigogine, Gertrude Stein, Dylan Thomas, and William Carlos Williams.

I: Lucky Accidents, Fortunate Errors

In mimesis[1] and subsequently in metaphor, the substitution of a fiction for a "true" object or event is a willful mistake. To borrow the language used by John Keats in defining "negative capability," we engage in a "willing suspension of disbelief." The "mistakes" or "errors" we thus make have the benefit of purpose and context, and these fictive purposes and contexts are catalytic in the production of discovery—of new, unanticipated meanings.

Thus when human beings employ mimesis and metaphor, they place themselves at a huge creative advantage—over the mere waiting for "fortunate accidents" to occur in the unmediated environment.

When DNA makes the fortunate errors eventually preferred and selected for survival (effecting evolution), they occur as mere accidents, as far as we can tell, and our providing context or stimulus for them seems difficult, remote, and unlikely. The case is very different with mimesis or metaphor, the occurrence of which are generally subject to our will and control.

Mimesis, and eventually metaphor, are specifically playing fields for the generation, recognition, preference, and selection of useful and creative "fortunate accidents." Does this mean that for creativity it is desirable to increase the number of instances of error? Yes: absolutely. That is: "absolutely" as opposed to "relatively." What is desirable is to enlarge the field of action, the field of discovery. What is desirable is to be situated in a way that is generative of a greater number of events and interactions of all kinds. We do not want more errors in proportion to "hits," but more errors and more successes as well: more everything, if you will.

Perversely, for purposes of creativity we are perhaps more interested in our errors, though. For with the luxury of a little delay we can are able to recognize (to register) those of our errors that are "fortunate accidents" worth hanging on to—and perhaps worth being augmented, replicated, or even preferred over formerly intended successes.

---

[1] In *PlayGROWnd* I "essentially use Merlin Donald's definition of the word *mimesis*: namely, "representation" by "acts [that] include intentionality, generativity, reference, autocueing, and the ability to model an unlimited number of objects" (*Origins of the Modern Mind,* Harvard, 1991—171).

Implicit is the idea of an original phenomenon upon which mimesis performs a replicative task. I would add that an audience (actual or potential) is implied too. Imitation alone won't suffice, because audience does not seem to me a requirement of imitation. Representation alone is also unsatisfying, because of the limitations that require Donald's amplification above, but also because of the ambiguity contained in the prefix "re-": I can present my hand to you, then re-present the identical hand again, without modeling, much less creating, anything; a mimetic act is in itself an original, an addition in the world, a new thing.

Paul Ricoeur offers a similar definition of *mimesis,* much more extensively, describing it as essentially Aristotelian ("The Narrative Function," *Hermeneutics and Human Science,* Cambridge, 1981—274-296, esp. 291-293).

The language I find myself using here is simultaneously suggestive of the work of two rather different thinkers: Rodney Brooks and Tom Peters, whose ideas have more in common than we might expect.

Brooks (in *PlayGROWnd* see "The Brooks Loop") is engaged in robot-building at M.I.T., and espouses a new paradigm of Artificial Intelligence called "behavior-based" systems. It is Brooks' hope that machine intelligence will prove to be "emergent"—that it will grow on its own when appropriately constructed robots are "situated" in real-world settings where they can learn as human beings do, largely by trial and error. (For is it not true that error becomes not only acceptable but even attractive when we specifically identify it as a learning tool?)

Peters (best known, perhaps, for his book *In Search of Excellence*) in recent years has been advocating corporate "dis-organization," chiefly in favor of flattening hierarchies and fragmenting them into "strategic business units" (SBUs) situated in closer proximity to real-world exigencies not unlike those which Brooks covets for his robots. These ideas are featured in Peters' *Liberation Management* (Knopf, 1992), which is reviewed in this paper (see "The Peters Loop"). One of Peters' favorite expressions is "Ready, fire, aim!"—a corollary of another favorite observation that in baseball hits are proportional to at-bats (as strike-outs are as well, of course).

At various points in this discussion I use the expression "error and delay." They are certainly not one and the same, and should not be taken as if intended that way. They are "kissing cousins." Delay is the interval in which some errors can be registered in the consciousness, can be identified as "fortunate" or potentially useful, and can be stored or, perhaps, manipulated or copied.

What might be an application or use of this idea? Well, if I myself were planning to build some robotic critters (or write some computer software) in which I hoped to allow for "emergent" intelligence, I would be cautioned by this idea to seek a way of not making any feedback loops too "tight." It would be important not to have my device or my software correct or compensate for its errors too quickly—before it might discover, register, and make creative use of "fortunate accidents."

One widely held view of AI—and the chief view that operates in *PlayGROWnd*—is that the subject amounts in a sense to "philosophy,"[2] in that it provides a way to speculate and talk about our own thinking. Thus in the present context discussions of AI and robotics are largely reflexive and reflective. See, too, "The Bachelard Loop"—especially Chapter Seven—for insights into the whole issue of seeking "objective knowledge of the subjective." Bachelard's *The Psychoanalysis of Fire* (*La Psychanalyse du Feu*, Librairie Gallimard, 1938; English: Beacon Press, 1964) presents not only a gloss to main ideas here, but some main ideas as well. Bachelard's book is primarily a caution that objectivity is very difficult—that it is inevitably infused and corrupted with desire. And we ask, if intelligence might be emergent, might not desire also be emergent?

[2] Or even theology, for that matter. In a presentation at the 1997 conference of the Society for Literature and Science, Rodney Brooks said in passing that he shared a postdoctoral advisee with Harvey Cox, of Harvard Divinity School.

Indeed, might not a corollary of emergent machine (read "artificial," if you will) intelligence be machine desire—or even machine Libido? As one of my graduate school professors was in the habit of saying, "That's to ponder." If so, we can perhaps ponder those aspects of AI's own "connectionism" that have networks working for approval or reinforcement—which is, in a way, to say "longing for companionship," is it not? Surely it is at least a longing for integration, a longing for unity.

We have somehow come in this discussion to the threshold of love. If that is a surprise, perhaps some insights may be gained from "The N. O. Brown Loop" –a review of *Love's Body* (Random House, 1966).

The sequence seems logical enough: where intelligence arises, what other features arise with it? Is it not possible, or perhaps even likely, that a concomitant of emergent intelligence might be subjectivity itself?[3] Brown's book may press us to consider whether an emergent intelligence would (or "ought" to?) experience ego-separation. Whether it would come to make a distinction (as we have) between itself and the world in which it is situated. Whether, in short, it would eventually require of itself some allegiance to a "reality-principle."

There would be steps on the way, probably. Or at least there must be steps for us in our consideration of such questions. So if there were to emerge something like desire, something libidinal, something tending toward love, would esthetics appear too? Would morality also arise—some ethical sensibility or capacity?

George Santayana has observed that "beauty is love attributed to the object." The esthetic sensibility has not infrequently been suggested as a source or at least an analogue for an esthetic sensibility. (For example, see Francis Hutcheson, *An Inquiry into the Original of Our Ideas of Beauty and Virtue* [1725].)

We must not fail to recognize that some impulse to poesis seems clearly to be at work in all of these matters. The creative desire, the desire to build things, surely participates in our seeking to build robots, to synthesize intelligence, to construct models of anything, including ourselves, to design and manage "organizations," and to fashion theories—as well as our passion to write poems and stories.

In "The Buber Loop" I point out in a footnote a correspondence between Martin Buber's "primal word" *I-Thou* (*I and Thou,* Scribners, 1958) and Gertrude Stein's observation in a 1936 lecture[4] that the mental process involved in the production of a masterpiece involves self-

---

[3] And what kind of subjectivity do we imagine for such an emergent intelligence? Indeed, what kind of intelligence? Would they be like our own? Would we need at least to respect them? Would we need to share values and convictions with them? Could we love them?

We find ourselves asking similar questions about the unified and integrated "body" envisioned by N. O. Brown. Are we one body with even those other human beings we regard as stupid or degraded? Or do "stupidity" and "degradation" limit their access to the common body—and hold them (agreeably?) at bay from us?

[4] "What Are Master-Pieces and Why Are There So Few of Them?"

forgetfulness. In psychoanalytic terms, both Buber and Stein seem to describe (or, for peak artistic creativity, pre-scribe) a reintegration of Ego with the world. In this they are joined, emphatically, by Norman O. Brown. All three share a concept of creative genius that in human terms I call "intimacy," but which in the present context might be regarded as an extreme and profound "situatedness."

Thinking that way, we begin to suspect that mind, in its grandest manifestations anyway, might be thought of as the "space" between subject and object. No wonder, then, that mind is so elusive of observation: it is Heisenberg's problem on steroids! For when we seek to observe the mind (using, of course, the mind) we not only change its properties and location, we cancel it out—and become, in a more than metaphorical way (but metaphorically too), mindless. For we have given ourselves the impossible task of treating the "space between subject and object" as an "object"—and that is absurd.[5]

Fortunately, metaphor[6] enables artists and scientists to do some absurd things like this one. So let us proceed.

Mathematician G. Spencer Brown begins his book *Laws of Form* (George Allen and Unwin Ltd., 1969) with a tiny essay ("A Note on the Mathematical Approach"), which in turn begins:

"The theme of this book is that a universe comes into being when a space is severed off or taken apart. The skin of a living organism cuts off an outside from an inside. So does the circumference of a circle in a plane. By tracing the way we represent such a severance, we can begin to reconstruct, with an accuracy and coverage that appear almost uncanny, the basic forms underlying linguistic, mathematical, physical, and biological science, and can begin to see how the familiar laws of our own experience follow inexorably from the original act of severance. The act is itself already remembered, even if unconsciously,[7] as our first attempt to distinguish different things in a world where, in the first place, the boundaries can be drawn anywhere we please. At this stage the universe cannot be distinguished from how we act upon it, and the world may seem like shifting sand beneath our feet.

"Although all forms, and thus all universes, are possible, and any particular form is mutable, it becomes evident that the laws relating such forms are the same in any universe" (v).

---

[5] In Buber's lexicon, we have tried to convert *I-Thou* to *It.*

[6] Science and art tend to use the words "metaphor" and "model" somewhat differently. In art (and rhetoric) a "metaphor" is a conscious and deliberate lie, employed for delight, discovery, and emphasis. At SLS 98, physicist Sidney Perkowitz observed in a session discussion that for scientists a "model is a metaphor with predictability." Other participants in the discussion warned that this might be intellectually hazardous, for models are in danger of being taken as miniatures of truth, whereas metaphors are forthrightly untrue.

[7] See "The Turner Loop." Is this "severance" what Mark Turner, in *The Literary Mind* (Oxford, 1996), calls an "image schema"?

Thus, in the present context, we are invited to conclude that by virtue of the very process by which it is "formed" (severance of subject from object) the mind may be considered a universe. (Caution, humility, timidity, or all three might urge us to prefer "an analogue to a universe," but I want to proceed boldly.) And naturally that in turn presses us to consider that, therefore, the universe may be a mind. If so, then we (among the *mise en scene* of the universe) occupy, in effect or in fact, the "space" between some "subject" and some "object"—perhaps the very one which many have chosen to propose is "the mind of God."[8]

---

[8] G. Spencer Brown (from the notes to Chapter 12):

"Returning, briefly, to the idea of existential precursors, we see that if we accept their form as endogenous to the less primitive structure identified, in present-day science, with reality, we cannot escape the inference that what is commonly now regarded as real consists, in its very presence, merely of tokens or expressions. And since tokens or expressions are considered to be of some (other) substratum, so the universe itself, as we know it, may be considered to be an expression of a reality other than itself.

"Let us consider, for a moment, the world as described by the physicist. It consists of a number of fundamental particles which, if shot through their own space, appear as waves, and are thus . . . of the same laminated structure as pearls or onions, and other wave forms called electromagnetic which it is convenient, by Occam's razor, to consider as travelling through space with a standard velocity. All of these appear bound by certain natural laws which indicate the form of their relationship.

"Now the physicist himself, who describes all this, is, in his own account, himself constructed of it. He is, in short, made of a conglomeration of the very particulars he describes, no more, no less, bound together by and obeying such general laws as he himself has managed to find and to record.

"Thus we cannot escape the fact that the world we know is constructed in order (and thus in such a way as to be able) to see itself.

"This is indeed amazing.

"Not so much in view of what it sees, although this may appear fantastic enough, but in respect of the fact that it can see at all.

"But in order to do so, evidently it must cut itself up into at least one state which sees, and at least one other state which is seen. In this severed and mutilated condition, whatever it sees is only partially itself. We may take it that the world undoubtedly is itself (i.e. is indistinct from itself), but, in any attempt to see itself as an object, it must, equally undoubtedly, act so as to make itself distinct from, and therefore false to, itself. In this condition it will always partially elude itself.

"It seems hard to find an acceptable answer to the question of how or why the world conceives a desire, and discovers an ability, to see itself, and appears to suffer in the process. That it does so is sometimes called the original mystery. Perhaps, in view of the form in which we presently take ourselves to exist, the mystery arises from our insistence on framing a question where there is, in reality, nothing to question. However it may appear, if such desire, ability, and sufferance be granted, the state or condition that arises as an outcome is, according to the laws here formulated, absolutely unavoidable. In this respect, at least, there is no mystery. We, as universal representatives, can record universal law for enough to say

> and so on, and so on you will eventually construct the
> universe, in every detail and potentiality, as you know
> it now; but then, again, what you construct will not be

That's pretty heady, to risk a pun here where it ought to be welcome. But for us the main issue at this point is a disturbing question: If we revere mind, and if mind is a product of Ego-separation, then must we revere the Reality-principle? Must we revere our "split" or "twofold" nature? Or must we (which is even less appealing) revere our failures of intimacy? My own answer is one that might be predictable, for it invokes a main motif of *PlayGROWnd*: to be grateful for a "fortunate accident" is not necessarily to "revere" it. One thinks of the concept of "the fortunate Fall," and psychoanalytically that well may be what we're considering here. Whatever we are, whatever we have become, we have become as a result of "Original Sin," our psychosexual development, and all those innumerable events which have "caused" us, including our having had monkeys for ancestors. Call such a view contentment or call it joy, this "mind" thing is the object and the tool of much that enriches our lives—and for my part, except for learning and loving ideas and writing poems and computer software: frankly, I don't care. We blossom where we were planted, or we don't blossom at all.

But what about robots? Do they "blossom where they're planted"? And if so, just how much blossoming can they be expected to do? I think a robot will have a "mind" in the instant it undergoes Ego-separation. Is this possible? I think probably it is.

II: "Single Vision"—Issues of Discipline and Person

At several points in *Origins of the Modern Mind* Merlin Donald observes that his book's subject is multidisciplinary. I consider this apparently minor observation to be one of the most significant contributions made by this extraordinarily significant book.

Implicit is an idea that "multidisciplinariness" is real—that it is an authentic quality or position. And that it is a real quality and represents a real position I whole-heartedly agree. The quality and position of being multidisciplinary are in a legitimate sense exclusive, it seems to me. That is, an issue or a subject that is truly multidisciplinary cannot be treated adequately by any single discipline. That, to me, is what the word means—very solidly.

Granting that the specific disciplines collaborating in a useful multidisciplinary study may be various disciplines, variously configured (so that "multidisciplinariness" cannot be considered a single discipline in its own right) —nonetheless with regard to truly "multidisciplinary" issues any single discipline taken by itself (1) cannot compete adequately with the multidisciplinary approach and (2) is as much an alternative or antagonist to the multidisciplinary approach as it is to another single discipline. Many commentators on cognition either fail to recognize cognition as a multidisciplinary subject or choose to treat it as if it were not.

---

all, for by the time you have reached what now is, the
universe will have expanded into a new order to contain
what will then be.

In this sense, in respect of its own information, the universe must expand to escape the telescopes through which we, who are it, are trying to capture it, which is us. The snake eats itself, the dog chases its tail" (104-106).

Roger Penrose's *The Large, the Small and the Human Mind* (Cambridge, 1997) features speculations by Penrose, followed by responses from Abner Shimony, Nancy Cartwright, and Stephen Hawking, these followed in turn by a final word from Penrose himself.

Here are the opening lines of Stephen Hawking's response to Penrose: "To start with, I should say I'm an unashamed reductionist. I believe that the laws of biology can be reduced to those of chemistry. We have already seen this happening with the discovery of the structure of DNA. And I further believe that the laws of chemistry can be reduced to those of physics. I think most chemists would agree with that" (169).

Clearly, Hawking's confession is made in the spirit of a multidisciplinary utterance. That is, he obviously intends it to be construed as a multidisciplinary comment, and it has on its surface the appearance of a multidisciplinary comment. Still, it is not. In its substance it is a statement in behalf of physics, but, more importantly, in its dynamic it follows a single thread—upon which biology, then chemistry, then physics are strung like three beads—with physics largest and strongest, at the bottom of a vertical string, serving as the knot that holds the others up.

It is fair to say that this illusion of multidisciplinariness characterizes most of science's putatively multidisciplinary treatments of cognition. At the bottom they very frequently end up being all physics, or all biology, or all chemistry, sociology, or psychology—all whatever.

Even literary theory can do this. For the longest time I could not enjoy or benefit from Mark Turner's *The Literary Mind* because it angered me on these very grounds. I tried to read it at least seven or eight times, putting it down each time in disgust because it seemed to me fraudulently to pose as multidisciplinary. For it is not multidisciplinary, in any large sense, and it was only after I accepted that recognition that I was able to see it, and accept the book itself, as the fine, essentially disciplinary work that it is.

In the case of Turner's book, I think the problem lay in its own claims in its own preface, which touts the book as multidisciplinary in a grand way: "This book is an attempt to show how wrong the common view is and replace it with a view of the mind that is more scientific, more accurate, more inclusive, and more interesting . . ." (vi). It is a very good book, anyway.

A "secret" truth about physics books that offer to take up topics other than the universe and its parts is that they are almost always books about the adequacy of physics—or the relation between physics and mathematics. They are hardly ever about anything else, no matter what they might say in their prefaces. What Penrose's book is about is somewhat broader in the present instance, though it is not about "the mind." It is about whether physics is the right discipline to study and understand the mind—and, if so, what kind of physics is best suited to the task.

It is to the book's credit that it seems at least partially conscious of its less than overt subordination of its nominal topic to its own disciplinary anxieties. This consciousness is most fully evident in the title of Nancy Cartwright's contribution: "Why Physics?" Yet the degree to which the book fails to treat its announced topic with earnest vigor is dramatic: Penrose's discussions of AI are, as ever, full of references to Turing and to chess problems. It shows no awareness, that I can see, of a new AI paradigm, but instead persists in seeing AI as an enterprise

devoted to the "construction" of intelligence: e. g., ". . . created by deliberate AI (artificial intelligence) construction" (113).

Our tendency to be unable to see beyond academic or scientific disciplines, much less to see how limited we are in this respect, is what I think Blake must have meant in his November, 1802 admonition against "single Vision and Newton's sleep" (letter to Thomas Butts), and perhaps what Francisco Goya had in mind in the engraving "The sleep of reason produces monsters." It seems to me unlikely that these warnings ought necessarily to be viewed as partisan remarks of Romanticism simply because cultural history has tended to define romanticism partially from such testimony of artists later identified as its practitioners. To speak against "single vision" is not to speak against science any more than to speak against "*el sueño de la razon*" is to speak against "*razon.*" It seems to me more likely that Blake and Goya feared single vision or sleep first, and only secondly pointed to Newton or reason as the readiest current examples. It is not a discipline itself that threatens our most fully human life, usually, but a blind or fanatical preoccupation with it.

This concern for clear-sightedness where disciplines are involved is analogous to warnings like Bachelard's concerning the elusiveness of objectivity for us as persons, as individuals—whose judgments, indeed whose selection of problems, issues and experiments, are subject to influence by our desires, our dreams, and our security needs. A truly multidisciplinary approach to ideas is hard, but it is also more than a little scary. It extends us out of our "comfort zones" and it expands our vulnerability. We are tempted—and sometimes tricked—to stay safely at home.[9]

Ilya Prigogine, 1977 Nobel Prize winner in thermodynamics, has observed (in *Order Out of Chaos: Man's New Dialogue with Nature*, Bantam, 1984): "Classical science, the mythical science of a simple, passive world, belongs to the past, killed not by philosophical criticism or empiricist resignation but by the internal development of science itself" (55).

I find Prigogine especially interesting in the manner in which he discusses the inappropriateness of the notion of empiricism (a favorite Prigogine topic). At the 1988 annual conference of the Society for Literature and Science, Prigogine made a plenary presentation in which he observed that scientific "discoveries" not only do not necessarily build on an accumulating body of empirical truth, but often are colored by the culture or even the person of the individual scientist who makes and proclaims them. He said (approximately) that if we cannot say of a symphony composed by, say, Mozart that it would sooner or later have been written by somebody, we also cannot necessarily say of a particular scientific discovery that it would sooner or later have been made by somebody else. Scientific discoveries, says Prigogine, bear the coloration of time, culture, and person very much as works of art do.

[9] Chaos theory is one area in which physics offers rich speculative opportunity for the understanding of cognition, specifically with respect to fractals. (For more on this subject, see "Hunches & Intuitions," below.)

This reminds us that persons, etymologically, are masks. Utterances, art-works, and theories tend to make us forget about the mask even as they help us to see what is on either side of it. One way to think about such a mask is to regard it as a boundary, a locus for potential severance.

III: Narrative

Some systems of expression seek to preserve or even proclaim the mask (i.e., the person). One such "system" is narrative. In recent years, narrative has increasingly come under study, and has enjoyed an elevation in status. (See, for example, the N. Katherine Hayles notes in "The SLS 1997 Loop.")

Though compared to quantitative data, narrative is sprawling, wasteful, and imprecise, it has powerful virtues even for science and technology. For example, it carries with it elements or aspects of content which endure—and preserve information which in more rigorous "data" would have been left by the wayside. That is, narrative can be "revisited" for useful information that might previously have been missed. (We could say that in this manner narrative allows for "fortunate errors" and, considering its need to be interpreted, useful and creative delay.)

Furthermore, because it does indeed require interpretation, narrative stimulates interpretation as a process, and exercises our interpretive skills and capacities. (This seems to be what Shelley meant in saying that poets are the "moral legislators" of the world: poetry exercises the moral imagination.) In addition, if "affect" is significant (as it almost always is, in human matters), narrative allows for it, expresses it, and legitimizes it.

Narrative also is mnemonic, helping us to remember ideas and events. It is implicitly social—and continually reaffirms our membership in community. (Incidentally, just what is the "basic unit" of humanity: the individual person—the word means "mask"—or the culture?) And narrative is "user-friendly"; it is acceptable and effective in human contexts where clinical data is less welcome. (See Kachur, in "The SLS 97 Loop.")

But narrative may have special importance because of its intimate role in the workings, or indeed the very shaping, of the human mind. In their book *The Discursive Mind* (Sage, 1994), Rom Harré and Grant Gillett propose that narrative processes play a role in the physical structuring of the brain itself, and a similar set of propositions are featured by Mark Turner in *The Literary Mind.*

If narrative is the source and image of what we mean by "mind" an interest in narrative surely needs neither defense nor explanation.

IV: Nerves and Neurons

The subject of nerves and neurons is far too extensive for *PlayGROWnd,* but some very useful observations (including issues of time and delay) are provided by Laura Otis and David Porush in their portions of "The SLS 97 Loop."

Any discussion of delay in respect to cognition presses us to consider the speed of human thought. It seems to us very fast. Yet Rodney Brooks proposes that in conventional computational terms it is astonishingly slow: "The usual estimates for the computational speed of neuronal systems are no more than about 1 kHz" (Brooks, 52). Still, our impression of the rapidity of our thought processes persists in spite of any such information. We are persuaded that our thinking is very fast, and we are ready to insist that this rapidity is something that, though we can't prove it, we know—even if we must resort to such unscientific conceptions as visceral (as opposed to logical or rational) knowledge. We conclude that the speed of thinking must be a different kind of speed from "computational" speed. We may propose that "computation" involves sorting (through layers or through "filters") or locating, whereas "thinking" may involve selection of objects or patterns already, constantly, present.

Time has become a darling subject of mathematics and physics (we could almost say they are preoccupied with it). Yet time as it is treated in those disciplines seems in many respects a very different subject from time as we live in it. (Both: *Live. In.*)

For mathematics and physics time is an "arrow," which is to say a line with a point on it. But in our lives it is also (perhaps more) a matter of quantity—and virtually has volume. This seems to me so true and so important that I find myself wishing we had an array of words to use in place of or alongside the word *time.* One close approximation of an alternative word may be *duration.*

Take, for example, such urgent human issues as pain and loss. Time participates vitally in both of these human experiences—so much so that we can say pain or loss have no existence without time. For pain, imagine, as an example, being hit with a hammer. Except for time, the impact of the hammer and the cessation of impact would be a single "point," and (leaving aside "permanent" damage) there would be no space (no location) for pain to happen in. Pain, then, is "the duration of hurting."

"Loss" is similar (though less precise a word than pain; I mean here loss that carries a sense of itself, a cousin to loneliness or grief—indeed, in some ways a form that pain takes). Loss is "the duration of no longer having."

But there is more to this issue than a difference between vectors and volumes.

In *The Large, the Small and the Human Mind* (Cambridge, 1997), Roger Penrose engages in a discussion of time (similar printed discussions are of course ubiquitous in physics and math) in which he describes (and draws a typically delightful Penrose cartoon of) a glass of wine falling from the edge of a table:

> "Imagine a glass of wine perched on the edge of a table. It might fall off the table, smash to pieces and the wine spill all over the carpet . . . . There is nothing in Newtonian physics which tells us that the reverse process cannot happen. . . . [W]e need the Second Law of

Thermodynamics which tells us that the entropy of the
of the system increases with time" (40).

Is this time of Penrose's the same time that we thrive or suffer in, do you think?

Try this: Is it true that we can't reverse the process of telling or hearing a story? If so—by Occam's razor, is entropy the most satisfying explanation?

V: Hunches and Intuitions

Are hunches and intuitions "protonarratives"? Perhaps by so regarding them we can, without dismissing them or impetuously indulging them, treat them as resources.

PlayGROWnd Cases in Point
—Is cognition a fractal, or "fractal-like"?
—Are we too timid when we say that sometimes photons behave "as if they know " which gates are open or closed or that electrons "seem to know" that they are being watched?
—Dare we propose that they "know "?
—Is our understanding of cognition impaired by ego, by arrogance?
—Do our hierarchical models of existence (the "Great Chain") seduce us into supposing that the cognitive process by which we study such matters as cognition is the only form that cognition can take, or that we ourselves occupy only "level" at which it can occur?[10]

---

[10] A richly interesting article on this subject is: Mark E. Wildermuth, "'And Anarchy Without Confusion Know': The Dynamics of Chaos in Pope's Essay on Man," *The Eighteenth Century: Theory and Interpretation,* 39:1 (Spring, 1998), 85-103.

Wildermuth's thesis statement reads: "In this essay I will show that theories on chaos and its relation to self-organizing systems in Pope's time were much more sophisticated than has often been reckoned" (86).

Understandably comparing Pope's ideas to those of Leibniz, Wildermuth writes: "Pope, like Leibniz, anticipates [the] conceptualization of the global and the local in modern physics. As Prigogine and Stengers say in Hermes ["Postface: Dynamics from Leibniz to Lucretius," Hermes: Literature, Science, Philosophy, Josue V. Harari and David Bell, eds. (Baltimore, 1982)], the language of today's dynamics has become Leibnizian: 'The world of trajectories determined by forces can henceforth be thought of as being identical to the Leibnizian system of the world in which every point locally expresses the global law' [(140)]" (87).

Wildemuth's point is that this "Leibnizian" conception sees the small instance as fractal kin to the large, rather than as a monad, atom, or molecule of the large. (An example is the homunculus, which as an intuitional "protonarrative" turns out to be "wrong"—but no tragedy; it might as well have turned out to be "right.")

He quotes David Castillejo (The Expanding Forces in Newton's Cosmos, as Shown in His Unpublished Papers {Madrid, 1981}), who "indicates that such forces operate 'in the radiation of light, in chemical composition, in biological growth and [in] [C's brackets] . . . the mind and behavior of human beings' (15)" (89).

—Is that notion itself more reliable than a hunch or intuition?[11]

Wildermuth continues: ". . . Pope seems entirely confident that the same forces shaping nature, both chaotic and orderly, shape the mind and what it constructs through language and metaphor—and thus a harmonics of nature and human artifice is possible" (93).

[11] See Wildermuth: "Ultimately, the monadic structure forms the basis of [*An Essay on Man*'s] ethics: pride is the denial of a paradigm which relates the global and the local while maintaining a hierarchy in a self-similar structure" (96).

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 Executive Council is the governing body of the Association. It consists of:

   (a) All past presidents of the Association.

   (b) Officers (acting president, vice president, treasurer, secretary and co-ordinator if one is appointed).

   (c) Ordinary members nominated by classes (a) and (b), who serve for three years, with the possibility of re-nomination.

2. Members of the Association are (a) members of the Executive Council and (b) others nominated by the members and approved by the Executive Council.

3. The membership and the Executive Council nominate vice-presidential candidates during the first year of the President's term of office. Any nomination must be accompanied by a statement from the nominee that he will serve a full term if elected. If there is more than one nominee, selection will be made by mail ballot to the Membership decided by plurality of votes. The Vice-President is elected to serve concurrently with the President during his last year of office. He will then serve as President for at most five years and cannot run for re-election until three years after his initial term has elapsed. If the President decides to stand down before his five year term has elapsed, he should give the Executive Council one year's notice of his intention, so that a Vice-President may be elected.

4. The President is the official representative of the Association in external affairs, and has the responsibility for calling meetings of the Membership, at least annually, for the determination of overall policy.

5. The Treasurer is the responsible financial officer of the Association for the receipt and disbursement of funds and shall maintain appropriate records of the Association Activities, membership, mailing-lists, etc.

6. The Secretary is responsible for keeping minutes of the Membership and Executive Council meetings, production of a newsletter to keep members of the Association informed of its activities, and such other duties as may be assigned.

7. President, Secretary and Treasurer will not be paid for their services but may, as appropriate, receive funds for travel expenses, secretarial help, etc.

8, The Co-ordinator, if one is appointed, may be paid an appropriate salary for his services, funds permitting. These 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.

9. 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.

10. Dues are currently £20.00 per annum.

President: Dr. Keith Bowden, 139 Sandringham Road, Barking, Essex, IG11 9AH, UK. [Tel: 0208 594 5064, Email: k.bowden@physics.bbk.ac.uk].

Co-ordinator and Secretary: At present no co-ordinator or secretary is appointed.

Treasurer: Tony Deakin, 75, Clatterford Road, Carisbrooke, Isle of Wight, PO30 1NZ, UK. [Tel.: 01983 524140]

Newsletter Editor: Arleta Griffor, 1 Venetia Rd, London N4 1EJ. [Tel: 0208 340 7985, Email: a.griffor@physics.bbk.ac.uk]

Proceedings Editor: Keith Bowden.