Implications

Scientific Aspects of ANPA 22

Keith G. Bowden, *Editor*

Proceedings of the 22nd Annual International Meeting of the

Alternative Natural Philosophy Association

Wesley House, Jesus Lane, Cambridge,
September 2000

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

July 2001

Implications: Scientific Aspects of ANPA 22/Keith G. Bowden, *Editor*

ISBN 0 9526215 6 8

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

Contents

Editorial

Owen Maroney recently showed me the following which he thought seemed relevant to the interests of ANPA. I haven't got the reference but I am sure he would point interested parties towards it.

"Blackett, at the time, was particularly interested in Bethe's work on the energy loss of charged particles in their passage through matter, because Bethe's calculations could be used to identify cosmic particles by the tracks they left in Blackett's cloud chambers. And Bethe responded by making detailed calculations that proved very helpful to Blackett.

Evidently, the Cambridge surroundings allowed the rigidity that a German education bestowed on scholars to be shed, for the year 1931 opened with the appearance of a startling *kurze Originalmitteilung* in *Naturwissenschaften* entitled "Concerning the Quantum Theory of the Absolute Zero of Temperature." The note was signed by G. Beck, H. Bethe, and W. Riezler, three postdoctoral fellows at the Cavendish Lab. Coming on the heels of Eddington's attempt to explain the numerical value of the fine structure constant, the note read as follows:

> Let us consider a hexagonal lattice. The absolute zero of the lattice is characterized by the fact that all degrees of freedom of the system are frozen out, ie. all inner movements of the lattice have ceased, with the exception, of course, of the motion of an electron in its Bohr orbit. According to Eddington every electron has $1/a$ degrees of freedom where a is the fine structure constant of Sommerfeld. Besides electrons our crystal contains only protons and for these the number of degrees of freedom is obviously the same since, according to Dirac, a proton is considered to be a hole in a gas of electrons. Therefore to get to the absolute zero we have to remove from the substance per neutron (= 1 electron plus 1 proton; our crystal is to carry no net charge) $2/a$ -1 degrees of freedom since one degree of freedom has to remain for the orbital motion. We thus obtain for the zero point temperature $T0=-(2/a-1)$ degrees. Putting $T0 = -273°$, we obtain for $1/a$ the value 137 in perfect agreement within the limits of accuracy with the value obtained by totally independent methods. It can be seen very easily that our result is independent of the particular crystal lattice chosen.

Since in those days papers in respected scientific journals were read with absolute trust in the honorable intentions of the authors and the editors, it took a while for the community to realize that the *Naturwissenschaften* had been fooled and that the paper was a prank. Arnold Berliner, the editor of the *Naturwissenschaften,* was not amused. Nor was Sommerfeld. Berliner demanded an apology and on March 6 there appeared a "Correction" in *Naturwissenschaften.*"

I trust that this calculation could prove to be of the utmost importance to ANPA.

This year we have split Proc. ANPA into two parts, ***Philosophical Aspects***, replacing both the Newsletter and the old Philosophical Appendix and edited by Arleta Griffor and ***Scientific Aspects***, edited by myself. We hope to continue this format in the future. This year's Scientific Aspects is named **Implications** in honour of Basil Hiley's contributions. Once again I publish comments on papers and offer the right to reply in future issues of the Proceedings.

I have continued my policy of reprinting old ANPA material by reproducing four articles from ANPA West, two by Vaughan Pratt, one by Fritz Lehmann and one by Lou Kauffman from the ANPA West Journal. The subject of posets comes up in two of these. This has been a topic of some interest at Birkbeck recently. ANPA may hear more about it soon. I also include some beautiful notes from ANPA West 7. If anyone can enlighten me as to the artist I would appreciate it. I have chosen articles by ANPA West members that have not been printed in these Proceedings (except for Lou, whose beautiful article on cats has not been seen here before). I have been blasé about copyright, assuming that ANPA has the right to non commercially reproduce material that it has printed before. If this upsets anyone then they have my apologies, but there was simply not time to obtain permission if the Proceedings was to be ready on time.

ANPA needs some new members. This would be easy to achieve. We could advertise various places, eg. NETWORK (the magazine of the Scientific and Medical Network), or PERT or HPS or Sigma Club mailing lists, for example. What do people think? Any other suggestions? This issue comes up time and again, but it is appropriate to revisit it. To some extent ANPA also needs to reposition itself. It has achieved its initial purpose most successfully. It has a healthy infrastructure and history and many possible interesting futures! Feedback on these issues would be much appreciated.

Finally my thanks are due once again to David and Patrick in the Print Unit at SOAS for producing this volume. And also to everyone who helped with the The Undivided Universe, a seminar held at Birkbeck College in June in honor of Basil Hiley, particularly Chris Clarke, and many other ANPA members.

KEITH BOWDEN, TPRU, BIRKBECK COLLEGE, JULY 2001

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 THE LAST 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 top centre. *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 ANPA2001 Proceedings is January 1st 2002. The issue will go to print on April 1st 2002. 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

Some ANPA PreHistory v1.0

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

This (part) chapter (of an intended book) is cobbled together from interviews with Pierre Noyes, in Septembers, 1991 and 1992, and with Clive Kilmister in February 1992 and Ted Bastin in November 1992 and comments by Chris Clarke at Southampton University and others. It chronicles the events leading up to the formation of ANPA (records don't agree!) and starts with Kilmister's version of the story in 1961/62 when Ted Bastin, who at that time had a fellowship at King's College, Cambridge University and Frederick Parker-Rhodes, a highly intuitive colleague of "independent means", were thinking about quantum phenomena and relativity and asking what a system would be like, which necessarily involved limitations on the parameters of the system. In particular they were trying to understand the Scale Constants of Physics by means of a machine analogy, and managed to get time on one of the last large analogue machines (PACE), in Brussels. This was at the time when people used to have computers and not know what to do with them. They had several sessions on the machine but soon the Belgians were finding their own things to do with the computer and saying, "Well, aren't you getting any results? Perhaps you ought to going back to England..."

On the day of the last session Parker-Rhodes had the 'flu and Ted Bastin went to work on the machine by himself. Frederick stayed at home and spent his time thinking about the Scale Constants, in particular how a mathematical series could be generated that started with two small numbers, then a medium sized number, and lastly a very large number. When Ted returned Parker-Rhodes waved a piece of paper at him, saying "I wonder if you have been wasting your time because I think that this captures what you're trying to do" and produced the algebra now known as the Combinatorial Hierarchy. Fred felt that as a mathematical series it must be wrong as (due to its combinatorial nature) it stopped after four numbers. Noyes emphasises that at this point Fred did not know the significance of the numbers (and particularly their precision). Ted was, (perhaps?), elated. Not only was this a finite series of four numbers, but the last two agreed with the values known for the fine structure constant (137) and the gravitational constant ($1.7\text{x}10^{38}$) to a completely unexpected degree. The work was eventually published as "The Scale Constants of Physics" in Studia Philosophia in 1966. This work was very much in the spirit of Eddington, who had attempted to derive the value of the electromagnetic scale constant from first principles, in the 1930's in his much maligned book ***Fundamental Theory*** and elsewhere.

Frederick's reasoning behind his relating this structure to physics, in such a way that the scale constants appear in the right places, has to do with what eventually became his major work, which was published as the ***Theory of Indistinguishables*** (1981). This was described to me by Chris Clarke at Southampton University, as possibly the most difficult piece of Mathematics ever written. It has four levels of syntax, each of which describes the level below it. The idea behind the theory is that two elementary particles, such as a pair of electrons, are indistinguishable from each other. No other physical objects have this property. Thus to describe these beasts a new set theory is required that allows sets containing identical elements. In fact, as is well known, conventional set theory can only avoid paradoxes, such as Cantor's, by means of some rather unsatisfactory cludges. To make things worse, the whole edifice of conventional mathematics, and hence physics, is built on conventional set theory. It turns out that the problem is caused by the set theorist's refusal to allow sets with duplicated entries. The multiset theory proposed by Wayne Blizzard and others avoids this at the expense of generating a much more complex theory. It turns out that multiset theory is just a much simpler version of Frederick Parker-Rhodes' Theory of Indistinguishables. Parker-Rhodes died in 1982.

Ted, Fred, Clive Kilmister and John Amson formed the original core of ANPA in that they had known each other over a decade before these occurrences. Ted and Clive were both undergraduates at Queen Mary College, London. In fact Ted had been aware of Clive some years earlier at school. Ted went to the Sir George Moneux Grammar School in Walthamstow. Clive was at Leyton County High School in Essex Rd, Leyton. After the first wave of war evacuations it was found that not everyone wanted to go to live in the country so a number of teachers were brought back and the two schools were put into one building. Ted is reported to have been in awe of Clive, being two years his junior. (Clive is one month younger than Pierre.) Clive however did not get to know Ted until they were together at QMC. Clive spent three years working on radar then returned to QMC after the war (1947) to submit his thesis on the work of Eddington. On trying to find a book on Eddington in the library he discovered that it was booked out to Ted and their acquaintance was renewed. Ted was doing his National Service as a Research Assistant working on semiconductors. At the same time he worked on Eddington and produced an MSc thesis on an application of Eddington's theory to liquid Helium. Ted left London to get a job at the new University at Ibadan, 100 miles North of Lagos, in Nigeria. He spent around three years there, the first year in the old University building (with no air conditioning) then later in the magnificent new University building, and returned to England with a lot of money in his pocket. Clive comments that ex-patriates were paid a lot but spent it on cars, women and booze. Ted (according to Clive) saved the money!

On returning Ted went to Cambridge to work on a PhD in Philosophy, later changing to Mathematics. Braithwaite and then Bondi were his tutors at King's. Later, Ted joined in a sort of cooperative at the home of Braithwaite and his wife, at 11 Millington Rd, as the nucleus of a group working on the philosophy of religion (the Epiphany Philosophers, a group which still exists). He stayed there for some years. Later, Braithwaite's wife, Margaret Masterman (who was the daughter of the then Home Secretary) started the Cambridge Language Research Unit (CLRU) at 20 Millington Rd where her group were attempting to computerise Roget's Thesaurus, with the aim of language translation. Everyone else, they reasoned, was in effect computerising a dictionary. This was alright if you were translating from English to French but once you got onto Mandarin Chinese (which Clive refers to as Masterman's secret weapon) the Thesaurus method would be better. Ted got an administrative home at CLRU for research grants. Later 20 Millington Rd became too small (and noisy with a newly installed telephone). The philosophy of religion group built new premises at 9 Marion Close. Clive picturesquely recalls arranging to meet Ted there. At this time the building was at the back of a convent. Well, said Ted, there isn't a lavatory there, but the nuns have one we can share, otherwise we can go to Millington Rd first if you wish! During my first ever ANPA conference in 1986 Mike Manthey and I stayed in this same peaceful building at Marion Close, by this time complete with lavatory and shower.

When Ted's fellowship came to an end he decided to stay in Cambridge even though he knew that the Cambridge Physicists would never employ him to do his kind of physics. When he married Suzanne Padfield (the parapsychologist), and left Millington Rd, he needed more money and worked as a bricklayer and later still, having received a little in an inheritance, invested in a car repair company. Unfortunately his partner turned out to be a crook and absconded to Los Angeles with some of the money. All this time he continued to develop the framework of the Combinatorial Hierarchy with the Cambridge based group now including John Amson, Gordon Pask (later of Brunel University), Fred and Clive. Pierre recalls Gordon Pask building a physical computer based on the mathematical ideas developed at CLRU, and Gordon and Ted taking it around the lecture circuit, encountering some very suspicious customs officers on entering the USA with it! This work was carried out with US Air Force funding on the understanding that they were developing machines for Natural Language.

In 1969 Ted organised the conference (with delegates including Bohm, Hiley, Penrose, Chew, etc) that led to the book "Quantum Theory and Beyond". Subsequently he travelled to Stanford University to give a talk on the Hierarchy under the sponsorship of Pat Suppes, Chairman of the Department of Philosophy at Stanford University. Pierre Noyes who at this time had become interested in the foundations of Physics met him at a Faculty Club lunch no later than 1971. Ted told Pierre about Fred's sequence of numbers and its correspondence with the Scale Constants. Pierre responded "You're a dangerous man!" This was mystical nonsense!

By 1972 Ted was back in Stanford again to give another seminar. Pierre was ready with guns on both hips. At the start of the seminar Ted made the remark that he thought the basic quantisation was that of mass. This was new to Pierre and sounded right but he still didn't understand and so paid closer attention. Ted moved on to talk about the scale constants of Physics and Pierre thought about how to phrase his objections to Ted's ideas. Suddenly Pierre realised that there was a place in Physics where combinatorics led to a number related to the scale constants. In 1952 Freeman Dyson had put forward an argument, based on counting, showing that not more than 137 charged particle pairs could be brought together within their Compton wavelength and remain stable. This could be extended to gravitation. Suddenly it all fitted together! After the seminar Pierre told Ted of his idea. Ted had not heard of the Dyson argument before but calmly accepted the revelation. For some years Ted had been looking for a high energy physicist to join the research group based in Cambridge and immediately suggested Pierre who took him up on it. After all this time the exoskeleton of a Discrete Physics seemed within reach!

Pierre already had strong contacts with England. He had been Fulbright scholar in the Maths and Physics Department at Birmingham University in 1950 and '51 and Leverhulme lecturer in Physics at Liverpool University in 1957 and '58. Pierre and the Marion Close group met at Burnham Overy Staithe Mill near Kings Lynn. He agreed to write an article (which was never published) putting the Dyson argument into the new context. The date on the SLAC preprint is 1974 and the title "Nonlocality in Particle Physics". This is largely the Combinatorial Hierarchy as it exists for ANPA today.

In 1976 Ted and Pierre attended a conference in Tutzing. Their paper was not published. A promise was received that it would appear in the 1978 proceedings. About this time John Amson said "Hold everything. We cannot know that the hierarchy actually exists without a proof." Nobody knew that the third level actually existed! Earlier an acquaintance of Ted's (Gefwert?) had travelled over to Finland. Pierre followed and spent a month or so over there by himself. He proved that the third level of the hierarchy exists by constructing 16x16 examples. Later Clive constructed a general proof. More recently Clive was asking *how many* possible ways of building the hierarchy there are. Pierre comments that it "blew his mind" when he found that the structure of the hierarchy was not unique. Ted and Clive knew all along.

Also in 1976 Ted received a National Research Council Senior Visiting Fellowship to make several visits to SLAC to work with Pierre. Why associate bit strings with particles? It felt natural at the time. Particles and bit strings were the most fundamental elements in Physics and the Combinatorial Hierarchy respectively. The paper presented at the Tutzing conference had already associated bit strings with the Feynman diagram. The discrimination of two strings into a third gives a diagram that looks like an F vertex. (Pierre emphasises that Feynman objects to the reification of the Feynman diagram with collisions. The arrows on the diagrams in the 1978 paper are considered to be a mistake.) Further a hierarchy of particles already existed and the Dyson argument had associated the third level with electromagnetism. The Hierarchy has two basic states at level 1. That suggests an association with the chiral neutrino which has only two states. The argument was published in full in Noyes' Physics essays paper. Ted asked awkward questions. Like what is isospin? Nowadays the argument is based in S-matrix theory.

About the same time Parker-Rhodes had come up with his controversial but remarkably accurate calculation of m_p/m_e and things were beginning to move. Frederick sent the m_p/m_e calculation to Nature without telling anybody in the group. It wasn't published. By this time it was becoming clear that the promise to publish the Amson/Bastin paper was not going to be honoured and that something funny was going on. In 1978 Pierre once again attended a conference in London. Fred and Pierre were authorised to report at a later conference in Tutzing. Finkelstein was there. Pierre asked if he would consider a paper in the International Journal of Theoretical Physics and received a backup commitment. As it became clear that the original Tutzing paper was not going to materialise Pierre sent a copy of "The Physical Interpretation of the Combinatorial Hierarchy" (PICH) to Finkelstein. John Amson had now come up with a new appendix in which the uniqueness of the combinatorial hierarchy itself was made clear. Thus "the" in the title was changed to "The" and PICH became PITCH. Finkelstein asked that the paper be shortened somewhat. The group agreed and the combinatorial hierarchy received its first publication since 1966.

Pierre's story now returns to 1979 and Ted is in financial trouble. Pierre received a letter from Ted saying "Help!" and describing his "cash flow" problems. By this time Ted was working nearly full time on the Hierarchy. Pierre had been asked to give a lecture on Galileo for the Western Culture program. This was followed by a dinner for eight to ten people. One of the women delegate's husband was an investment councillor called Dougal Thomas and he and Pierre hit it off at once. Pierre invited him to lunch at the Faculty Club at Stanford. Half way through the meal Pierre let drop the story about his contacts in foundational physics. Thomas clearly did not know what to do with all this institutional largesse. This put a bee in Pierre's bonnet about forming a sponsored organisation administered by Ted to coordinate the investigation into the Hierarchy. That night he put together a $300,000 proposal for what was to become ANPA. The next day he sent this to Thomas who was intrigued. However shortly after these events Thomas switched jobs and moved from the West coast to the East. During his last week at Stanford, Thomas and Pierre got together for a talk. Thomas agreed to try to raise some funds.

Three or four months later Pierre independently received a US Senior Scientist Award from the Alexander von Humbolt Organisation which enabled him to spend some time in Germany. He also discovered that he could simultaneously take sabbatical pay. Pierre travelled to Germany for six weeks. He was feeling "flush". Thomas was clearly not coming up with the financial goods and Pierre thought "What the hell" and wrote a letter to Clive and Ted. He volunteered to put up 5000 marks towards running the new organisation. Frederick (who enjoyed a "modest sufficiency") donated £1000. They had enough to get off the ground.

In 1979 ANPA was formed and held its first meeting (ANPA 1) in Clive Kilmister's "Red Tiles" cottage in Sussex. John couldn't come. The meeting consisted of Ted, Fred, Clive and Pierre. It lasted three days and the minutes still exist. It was a scientific meeting but also discussed the organisation of the new Association. All subsequent meetings were held in Cambridge. The organisation was set up so that the coordinator received a salary for the first few months.

In 1980 or 1981 (after ANPA2 or ANPA3) during a meeting in Clive's office at King's College in London, it was realised that in addition to discrimination a generation operator was needed. It could not be assumed that the bit strings existed. Further the bits in the strings were separated into label and content bits. The label was associated with the quantum numbers of a particle and the content with the space-time component. The association of the hierarchy with the elementary particles had already been published in PITCH. Program Universe (with Mike Manthey) and the Counter Paradigm were developed over the period 1980 to 1983. The bit strings were generated as Bernoulli sequences in order that they should have least structure. About this time interest started in the Stein random walk model. This gave a new way to interpret the bit strings. The content part could represent energy-momentum rather than space-time.

So ANPA in its current form was off the ground. In 1985 with Tom Etter, and the help of Pat Suppes, Pierre started a Western Chapter of ANPA, based at SLAC (Stanford Linear Accelerator Centre). Clive Kilmister and Ted Bastin recently published a book on their work, entitled Combinatorial Physics. Clive described their relationship to me like this, "Ted has the ideas and I put them into mathematical form. When he sees what I am trying to do he says either 'That's right, Yes', or 'No, That's wrong!' and I have to start again". Clive and Ted's approach to the Combinatorial Hierarchy is now very different to that of the early days. They are talking about the process of Scientific Discovery. All things are taken to be a member of one of two sets, the Known and the Unknown. The process of discovery involves taking an element out of the Unknown and putting it in the Known. To do this we must compare the potentially new element with those already in the set of Known things to see if we already knew about it. Clive likens this to the actual practice of Physics when looking for new particles. However it must be emphasised that we are most definitely *not* talking about particles at this stage and that this is only a loose analogy (albeit with hints). The process of comparison is known as discrimination.

Non-commutative Geometry, the Bohm Interpretation and the Mind-Matter Relationship[*].

B. J. Hiley

Theoretical Physics Research Unit, Birkbeck, University of London, Malet Street,
London WC1E 7HX.
b.hiley@bbk.ac.uk www.bbk.ac.uk/tpru

Abstract.

It is argued that in order to address the mind/matter relationship, we will have to radically change the conceptual structure normally assumed in physics. Rather than fields and/or particles-in-interaction described in the traditional Cartesian order based a local evolution in spacetime, we need to introduce a more general notion of process described by a non-commutative algebra. This will have radical implications for both for physical processes and for geometry. By showing how the Bohm interpretation of quantum mechanics can be understood within a non-commutative structure, we can give a much clearer meaning to the implicate order introduced by Bohm. It is through this implicate order that mind and matter can be seen as different aspects of the same general process.

1. Introduction.

The aim of this talk is provide a general framework in which the relation of ordinary matter to ordinary mind can be discussed[1]. I will not address any details concerning the structure of the complex of neurons or of the electro-chemical processes occurring in the brain, vital though these details are. Rather I will try to provide a general framework in which we can eventually explain how the physical-chemical-electrical properties of the brain can give rise to thoughts, feelings and ultimately consciousness.

I do not believe that today's physics is rich enough to handle these questions and it will be necessary to develop new concepts before we can really begin to explore this relationship adequately. It has been argued that classical physics will provide all the answers we need. I do not share this position. Nor does Stapp (1993) who writes "Classical physics strives to exclude the observer from physics and succeeds. On the other hand quantum mechanics strives to exclude the observer and fails". The first part of this quotation is undoubtedly correct and therefore classical physics excludes the very thing that we are hoping to understand.

On the other hand I do not share Stapp's belief that quantum mechanics already contains sufficient structure to answer the deep questions. My position here does not stem only from my study of the Bohm interpretation (Bohm and Hiley 1987 and 1993). It also

---

[*] To appear in Proc. CASYS'2000, Liege, Belgium, Aug. 7-12, 2000.

[1] By 'ordinary mind' I specifically exclude the so-called 'paranormal'.

follows from my study of the Copenhagen view. Unfortunately the Bohm interpretation is misunderstood and misrepresented in the literature. It does not stand diametrically opposite Bohr's views. It actually shares some of its radical conclusions. For example, it clearly shows that there is an essential element of participation involved, a notion that has already been shown by Wheeler (1991) to be a feature of the standard interpretation. Bohr saw it slightly more radically when he remarked that basic to quantum theory is the impossibility of making a sharp distinction between the observed system and the means of observation. At this stage participation does not necessarily involves the human observer, but it certainly involves the observing apparatus. The extension to the brain/mind interface seems very inviting.

However this brings us to one of the main difficulties in trying to discuss whether quantum theory will help us to understand the mind/matter relationship. Quantum mechanics itself is plagued by problems of interpretation. Why else do we have all these other interpretations, viz. the standard interpretation, the Copenhagen interpretation, the statistical interpretation, the many-worlds interpretation, the Bohm-de Broglie interpretation, the Bohm ontological interpretation, the consistent histories interpretation, the transactional interpretation, the many minds interpretation, the modular interpretation.......

This list is by no means exhaustive! Clearly such a profusion of interpretations can only lead to the conclusion that there is something very wrong somewhere. I feel that there is a deep and fundamental problem in the framework into which we can try to fit quantum processes.

Notice that amongst the main properties demanded of an explanation in classical physics is that it is deterministic, continuous and local in space-time. In quantum theory at the particle level we find indeterminism, quantum jumps and non-locality in space-time and we are perplexed. However at the level of the wave function we restore determinism, continuity and locality in space-time, through the Schrödinger equation, the Dirac equation, and the Klein-Gordon equation (Dirac 1973). This means we have implicitly given the wave function ontological status by considering it to be the most complete description of the *state* of the system.

We as physicists are happy with this even though it leaves us with all the problems of interpretation at the particle level. We are happy that we have found a way of describing quantum phenomena without being forced to give up the classical paradigm. But by keeping this paradigm, which I will for convenience call the Cartesian order, we have continued to separate the observed from the means of observation. Thus we must necessarily maintain the sharp separation between mind and matter, between res cogitans and res extensia (Hiley 1997). By retaining this classical order, we have made it very difficult to see how physics is ever going to explain what I call the 'ouch factor'.

Before I go on to discuss how we can change this classical paradigm, I must say a little about the way I came to this position through my collaboration with David Bohm. Bohm started effectively with the question "In quantum mechanics, can we keep the notion of a particle with its simultaneously well-defined position and momentum and always talk about particles following trajectories?" This of course was denied by the conventional wisdom of the day. Indeed as far back as 1936 Norbet Weiner (1936) wrote,

> One might suppose that it is still possible to maintain that a particle such as an electron has a definite momentum and a definite position, whether we can measure them simultaneously or not, and that there are precise laws of motion into which this position and momentum can enter. *Von Neumann has shown that this is not the case, and the indeterminacy of the world is genuine and fundamental* [my italics]. There are no clear-cut laws of motion which enable us to predict the momentum and position of the world at future times in any precise way in terms of any observable data whatever at the present time.

We had to wait until the sixties for Bell (1987a) to point out exactly why the von Neumann theorem was limited as were the subsequent improvements of Gleason (1957) and Kochen and Specker (1967). These limitations do not apply to the Bohm approach. They were based on attributing to the particle eigenvalues of all possible operators simultaneous, together with the assumption of non-invasive measurement. The Bohm approach does not make such an assumption and treats measurement as non-invasive.

Let us now show how all this comes about. Following Bohm (1952), we substitute $\Psi = R\exp[iS/\hbar]$ into the Schrödinger equation. By splitting this equation into its real and imaginary parts, we obtained two equations, one showed that probability is conserved, while the other could be interpreted as providing, as Weiner (1936) put it, a "precise law of motion into which this position and momentum can enter". Solving this second equation we can calculate sets of trajectories for each quantum situation. All of the relevant details appear in the literature and references can be found in Bohm and Hiley (1993) and in Holland (1993).

This second equation can be written in the form

$$m\left(\frac{d^2x}{dt^2}\right) = \nabla[V+Q] \qquad \text{where} \quad Q = -\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R} \tag{1}$$

These equations give the impression that we have returned to a classical account of quantum phenomena. Nothing could be further from the truth. Not only is there no return to Newtonian mechanics, but no form of mechanics can be sustained (Bohm 1957). Central to understanding how the Bohm interpretation is the appearance of the quantum potential, Q. It is not *ad hoc* as suggested by Heisenberg (1959) but emerges directly from the Schrödinger equation and without it, energy would not be conserved.

The quantum potential does not have the usual properties expected from a classical potential. It does not arise from an external source; it does not fall off with distance. It seems to indicate a new quality of internal energy and more importantly from our point of view, it give rise to the notion of participation, non-separation and non-locality. At the deeper level it arises because, as Bohr often stressed, it is not possible to make a sharp separation between the observing instrument and the quantum process while the interaction is taking place. The Bohm approach makes a *logical* distinction between the two but then the quantum potential links them together again so that they are *actually* not separate. It is this factor that gives rise to context dependence, and to the irreducible feature of participation between relevant features of the environment in the evolution of the system itself. It was this factor was not incorporated into by the no-go theorems discussed above.

The appearance non-separability and non-locality in the Bohm approach led Bell (1987b) indirectly to his famous inequalities. Of course, non-locality is not a feature that fits comfortably within the mechanical paradigm, but it was this feature that led Bohm to the conclusion that his approach was NOT mechanical. More details can be found in Bohm and Hiley (1993).

Our conclusion from our detailed investigations into these questions in the Bohm approach and in our review of other approaches to quantum mechanics led us to the conclusion that the Cartesian order could no longer be used to explain quantum processes. Indeed even in a model that re-writes the Schrödinger equation in a form that apparently brings it closer to the equations of classical mechanics contains non-separability, participation and non-locality. What is needed is a radically new order in which to understand quantum phenomena. Bohm (1980) suggested that this new order would be based on process. He called this new order the implicate order. I will not discuss Bohm's own justification for his proposals, but would rather approach the subject from a different point of view, which suggests that novel structure actually lies beyond space-time. This structure is rich enough to discuss the relation between mind and matter.

2. Is spacetime primary?

I first came across this possibility from a lecture by Geoffrey Chew (1960). He pointed out that there is no necessity to start an explanation of quantum processes in space-time. Complementarity shows that we could either start in space-time, or we could have started in the energy-momentum plane, but we can never start with *both together*. This is actually an old idea stressed by Bohr (1925) in the early days of quantum mechanics. He writes

> I am quite prepared that the view we proposed (Bohr-Kramers-Slater theory) on the independence of the quantum process in widely-separated atoms should turn out to be incorrect..... the Ramsauer's results on the penetration of slow electrons through atoms, presents difficulties for our ordinary space-time description of nature similar to those presented by a simultaneous understanding of interference phenomena and a coupling through radiation of the changes of state of widely-separated atoms. I believe that these difficulties so thoroughly rule out the retention of the ordinary space-time description of phenomena...

Chew (1960) brought this idea out in a new and striking way by drawing attention to the S-matrix approach to high-energy processes. Here the energy-momentum plane is taken as basic so that we can exploit strict energy and momentum conservation. But then the role of the spacetime manifold has to be *derived* since it can no longer be regarded as basic. This brings us to the question of the role of space-time itself. Why is it regarded as primary and basic?

When we come to consider the problems of quantising gravity while retaining general relativity, we face the following dilemma. As is well known in general relativity the gravitational potential is identified with the metric tensor. Now in any quantum field theory, the fields themselves are subject to quantum fluctuations. Thus the quantised gravitational field would imply fluctuations in the field and since the gravitational potential reflects the metric properties of the space, the space-time itself must be fluctuating. But what then is meant by a fluctuating space-time?

The third problem in assuming that space-time is fundamental arises from the appearance of quantum non-locality. If space-time is taken as primary, then, *ipso facto*, locality is absolute. Indeed the space-time manifold dominates classical physics because it has locality built into it right at the beginning. If we retain the space-time manifold, then quantum non-locality sits very uncomfortable in such a structure.

Could it be that our insistence on taking a given space-time as basic is at fault? Could space-time merely be an appearance, a feature that has to be abstracted from some deeper structure, a structure where space-time itself is not taken as basic? If this were the case, then it would be establishing locality that would present the problem. Could it be that locality itself is merely a relationship? This relationship dominates the macroscopic world, but it would not be universally valid at the quantum level. Yes there is relativity, but does that theory apply to the level of a single photon or only to a statistical ensemble of photons?

The first suggestive example of showing how locality could be a relationship appears in the hologram. Here a picture of an object is recorded as an interference pattern. The image of the original object can be re-created by using an appropriate light source. If the hologram is now torn in half and the light passed through this half, we again see the *whole* object, albeit with some loss of overall definition. Clearly the local regions of the original object are mapped into the whole of the photograph, so that locality is being carried in a non-local way. Thus locality here is clearly carried as a relationship. Can idea be generalised?

Suppose locality is a relationship, could it be that quantum phenomena are in some sense beyond space-time and are merely projected into space-time by our macroscopic instruments? In other words, could quantum processes be evolving in some more general space, which for convenience we call simply 'pre-space'. This pre-space (Hiley 1991, Hiley and Monk 1993) would then give rise to Wheeler's (1980) pre-geometry. In this view, the space-time of the classical world would be some statistical approximation and not all quantum processes can be projected into this space without producing the familiar paradoxes, including non-separability and non-locality. In classical physics everything is local so that a single space-time can provide a contradiction free description.

If we adopt this radical view, we can see that it is not necessary to insist on the Cartesian division between res extensa and res cogitans. Matter actually has its origins in a deeper structure, a structure where space-time and hence extension is not primary. If such an approach were viable then matter and mind need no longer be separated by space-time constraints as illustrated in the picture below.

Cartesian cut

Res extensa	Res cogitans
Locality, Continuity & Determinism	Nonlocality, Jumps & Indeterministic!!
IN SPACE-TIME	NOT IN SPACE-TIME

This is fundamentally the wrong view. Something new is needed, and this new order must not take space-time as basic and fundamental.

3. Non-commutative algebras.

The question I now want to turn to is how are we going to implement this general programme mathematically? I want to suggest the answers lie in the non-commutatity of the quantum formalism. To bring this point out, consider a particle in motion. In classical mechanics both the position, x, and the momentum, p, are known, so that we can construct a phase space in which we can track the particle as it follows a specific trajectory (see figure 1)

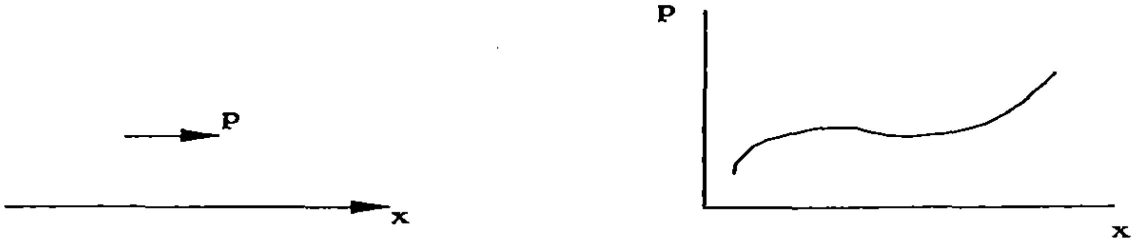


Figure 1. Classical phase space.

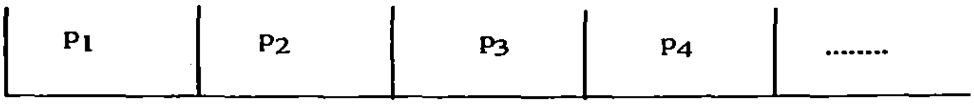
In quantum mechanics, non-commutativity implies that we can either have exact information about x through the eigenvalues of the position operator, X, or we can have exact information about the momentum eigenvalues, p, of the momentum operator, P. We can not have both together because $[X, P] = i\hbar$. Thus we have either

$$S_1 X S_1^{-1} = X_{dia} = \begin{pmatrix} x_1 & & & \\ & x_2 & & \\ & & \cdot & \\ & & & \cdot \end{pmatrix}$$

x_1	x_2	x_3	x_4	....

Or

$$S_2 P S_2^{-1} = P_{dia} = \begin{pmatrix} p_1 & & & \\ & p_2 & & \\ & & \cdot & \\ & & & \cdot \end{pmatrix}$$



Thus we cannot construct an exact phase space as we can in classical mechanics. All of this is, of course, well known, but I want to take it further. To do this I have to take you back to the end of last century, to the work of Hamilton (1967), Grassmann (1995) and Clifford (1878). In this period in the development of mathematics, the study of the properties of algebras was in a fairly primitive stage and there was an energetic discussion as to the metaphysical significance of algebra in general.

To give a flavour of the attitudes of the time, consider the title of Hamilton's (1837) lecture "The Metaphysics of Mathematics-Algebra as Pure Time", a title that one would expect from someone on the fringe, and not someone at the centre of things! In that lecture he wrote:

> In algebra relations are between successive states of some changing thing or thought. In other words algebra is not about material process but something more general that could be applied to both matter and mind.

Grassmann (1995) takes this further. He argues that mathematics is about thought, not material reality. It is about relationships of form, not relationships of content. Mathematics is to do with ordering forms created in thought. Thus since thoughts are not located in space-time, mathematics is not necessarily about material things in space-time.

Now thought is about becoming, how one thought evolves into an other. Is a new thought independent of the old thought; is the old thought independent of the new thought? Surely the old thought contains the potentiality of the new thought and the new thought contains a trace of the old thought. There is no separation. Thinking is about becoming not being. Being is a relative invariant in the overall process of becoming. The basic ingredient must therefore be activity or process and this process is described by the elements of an algebra.

4. The algebra of process.

The main novel feature of the mathematics of quantum theory lies in the non-commutative structure of its algebra of operators. If we regard the eigenvalues as labelling the properties of things, non-commutativity does not seem to make any sense. For example, objects should have position and momentum simultaneously. Objects should have an x-component and a y-component of angular momentum simultaneously, but in quantum theory they do not.

There no room for non-commutativity in the classical world of objects. Yet the classical world does actually contain lots of non-commutativity. Try taking a cup from a cupboard before opening the door! Try rotating an object through a 90° rotation about the x-axis and then 90° about the y-axis. Repeat by first rotating about the y-axis and then about the x-axis. You end up with different configurations. In other words the classical world contains plenty of examples of non-commutativity, but it is always associated with activity or process.

With this in mind let us look more closely at what the algebra of quantum theory implies when we apply it to simple situations. To highlight the problem let me illustrate the difficulty if something like colour and shape were described by non-commuting operators[2]. To make things even simpler let us suppose there are only two colours red and blue, and there are just two shapes, spheres and cubes. Furthermore we cannot view these properties directly but need some instrument to determine the colour and another to determine the shape.

Suppose further that in this example our observables are represented by the non-commuting operators C and S, with $[C, S] \neq 0$. Our objects must be described by wave functions, Ψ_R for red, Ψ_B for blue, Φ_S for sphere and Φ_C for cube.

Now let us try to collect together a set of red spheres. First we measure the colour and collect all the reds together in one group, separating them from the blues. Take the red set and find out which of these red objects are spheres. Thus we can collect a set of objects that were red according to the first measurement and spheres according to the second measurement[3].

We might be tempted to conclude that we now have a collection of red spheres, but we had better check that they are all still red! When we check this, we find half of our spheres have changed colour and are now blue! This result follows from the fact that the 'observables' do not commute. If $[C, S] = 0$ then re-measuring the colour of the objects would still all be red. It is only in this case that we can divide our objects into four unique sets; sets of red spheres, sets of red cubes, sets of red cubes and sets of blue cubes. In the world of non-commutativity this is something you cannot do. You cannot display every property in one 'picture'.

This example shows very clearly what we are up against in quantum theory. The central question is how do we understand this situation. We can follow Bohr (1961) and argue that it is simply a fact that must be understood in terms of the principle of complementarity.

As an alternative you could try to maintain the assumption that colour and shape are still properties but that the measurement process itself changes the complementary variables in some new way. Measurement simply makes manifest one particular partial view of nature and it is not possible to make manifest *all* aspects of reality in one single universal 'picture'. This is the view expressed by Bohm's implicate order (Bohm 1980).

At the deeper level, the order is not explicit, its is implicit and the structure of this implicate order is captured by the algebra. Our measurement merely displays one particular aspect, which we call the explicate order. Different measurements produce different explicate orders. Thus in the example above, colour would be one explicate order and shape another. At the deepest level the process has neither colour nor shape. These features arise only in relation to the process of manifestation. Each process forms a totality and our attempts to describe this requires us to divide the process into text and context to make it meaningful.

---

[2] Here we will effectively replace spin by colour and direction by shape.

[3] Remember we have two different pieces of apparatus, one to measure colour and the other to measure shape.

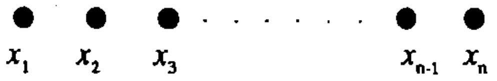
5. Consequences for phase space.

Let us now move on to consider how it might be possible to discuss the general evolution of such a process. I have already discussed a number of these ideas in some detail elsewhere (Hiley 1991, Hiley and Fernandes 1997). Here I simply want to illustrate what I have in mind without going into detail again, much of which can be found in the above papers.

All process starts with a law of succession, how one process evolves globally into another. This law is expressed through a rule of multiplication. Succession must be complemented with coexistence, giving rise to the law of coexistence. This is represented by addition thus defining an algebraic structure. We can add one further feature, namely, the process must be made manifest relative to a context. Different contexts are then determined by different representations.

In order to show how a quantum phase space can be constructed we need the symplectic Clifford algebra (Hiley and Monk 1998, Crumeyrolle 1990). This algebra contains the necessary structure to build a quantum phase space. The points of this space can be constructed from the idempotents of the algebra. An idempotent is defined through the relation $P^2 = P$, that is a point is a process, which under the law of succession, transforms into itself. Thus points themselves are not static concepts, but part of the underlying process.

The algebra itself contains elements that enable one idempotent to be translated into another, so that the algebra contains within itself, its own translation operator. This whole structure enables us to define a space, which we will call the x-space. This is illustrated in the figure below.



x-space

But this is not the only space that is contained in the algebra. An appropriate inner automorphism produces the p-space.

p-space

However the p-space can only be 'constructed' at the expense of 'destroying' the x-space. Thus the basic underlying process itself is such that it is not possible to construct simultaneously a phase space in which both x and p are sharply defined. This is of course exactly what the theorems of von Neumann (1955) and Gleason (1957) are about.

Notice we are not regarding the lack of precision as being due to an 'uncertainty' as if everything is actually certain, but that we, as observers, are uncertain as to the precise values because of some 'ham-fisted' use of apparatus. We are arguing that the process itself is such that it is not possible in principle to define x and p together because simultaneous x and p does not have a meaning.

The basic underlying assumption of this general approach is that the ontology is based on process, a process that cannot be described explicitly. It can only be described implicitly, hence the terminology 'implicate' order. This implicate order is a structure of relationships, and this order of structures is described by an algebra, the algebra of process (Hiley 1995). Here the implicate order is not some woolly metaphysical construction, it is a precise description of the underlying process, mathematically expressed in terms of a non-commuting algebra. This process only allows partial views because nature is basically participatory.

To put it more strongly, it is not that we as observers who participate in nature, *but that nature participates in nature.* Thus the observer is not something special. The cosmos does not need observers to function and evolve. Observation is simply a particular example of general notion of transformation in which the observed and observing processes fuse in an indivisible and irreducible way. Bohr (1961) talked about it as "the indivisibility of the quantum of action". It is not that we can never separate objects from the observing process, we can once the interaction has ceased. But during the interaction the individual becomes an intrinsic part of the whole process, and becomes transformed in the process. This is how the example of the colours and shapes outlined above can be understood. The colour measuring process can transform the shape and the shape measuring process can transform the colour.

6. The evolution of process.

The notions discussed in the previous section are very different from what we are used to but fortunately Bohm found a very illuminating metaphor through which to illustrate some of the key features. Indeed the metaphor has the advantage of suggesting how we may describe the evolution of process mathematically without the need of a space-time manifold.

Consider a hollow outer cylinder containing an inner cylinder that can be rotated relative to the outer cylinder. Glycerine is poured between the cylinders as shown in figure 2.

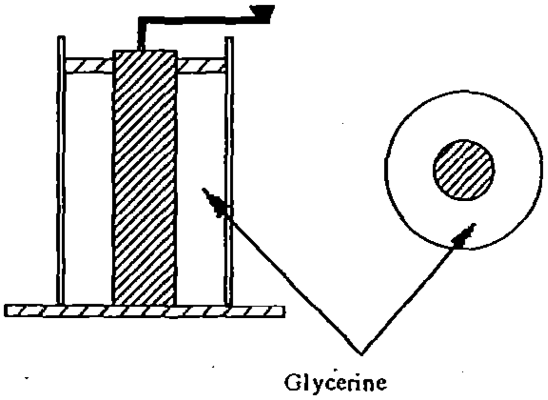


Figure 2. The Unmixing Experiment.

Then a spot of dye is introduced into the glycerine at some suitable point. If the inner cylinder is rotated, the dye disappears. There is nothing remarkable about that, but if the inner cylinder is rotated in the opposite direction, the spot re-appears. (There is some diffusion if this is carried out in a real experiment, but the diffusion is actually small and can be ignored for the purposes of the metaphor.)

In the spirit of the implicate order, we can regard the 'order' of the spot to be enfolded in the glycerine, so that it becomes 'implicit' in the glycerine. It can be made manifest again by rotating the inner cylinder in the opposite direction. We could imagine a series of dots enfolded at different times and at different neighbouring positions. As the process is 'unfolded' a succession of spots are re-manifested giving the appearance of something moving through the glycerine. This gives the appearance of a particle following a trajectory although there is no particle, there is just a process of enfolding and unfolding.

The physical process lying behind the glycerine illustration has a clear classical explanation at the atomic level. That is not the point. What the metaphor is intended to do is to bring out the fact that if the basic process was activity *per se*, then the 'track' left in, say, a bubble chamber could be explained by such an unfolding and enfolding process. Thus rather than the track being seen as the continuous movement of some material object, it can be regarded as the continuity of a quasi-stable form, evolving within the unfolding process.

We can think about describing such a process in the following way. Suppose we consider two successive moments described by the explicit orders $e(\tau_1)$ and $e'(\tau_2)$, where τ is some parameter[4]. Let the unfolding process be described by M_1 and the enfolding process be described by M_2. Then if we use the law of succession we have the two processes $e(\tau_1)M_1$ and $M_2e'(\tau_2)$. But the continuity of form demands that

$$e(\tau_1)M_1 = M_2e'(\tau_2) \qquad \tau_2 > \tau_1.$$

Thus the enfolding and unfolding movement is an automorphism of the algebra since

$$e'(\tau_2) = M_2^{-1}e(\tau_1)M_1 \tag{2}$$

If we now assume for simplicity that

$$M_1 = M_2 = M; \qquad M = \exp[-iH(\tau_2 - \tau_1)] \tag{3}$$

then for small τ we find

$$e' = (1 + iH\Delta\tau)e(1 - iH\Delta\tau)$$

so that

---

[4] Here we can regard τ as an unfolding parameter.

$$i\,\frac{\Delta e}{\Delta \tau} = [H, e], \tag{4}$$

which is the same form as the Heisenberg equation of motion if we identify the unfolding parameter as time. Thus our general ideas have led us to an equation of motion that is identical to the one basic to quantum mechanics.

If we now write $e = AB$, then equation (4) becomes

$$i\left(\frac{\Delta A}{\Delta \tau}\right)B + iA\left(\frac{\Delta B}{\Delta \tau}\right) = (HA)B - A(BH).$$

The form of this equation suggests the possibility that the whole process is can be described by a pair of equations

$$i\left(\frac{\Delta A}{\Delta \tau}\right) = HA \qquad \text{and} \qquad -\,i\left(\frac{\Delta B}{\Delta \tau}\right) = BH$$

If we identify A with Ψ and B with Ψ^* we see that these equations have the same form as the Schrödinger equation and its conjugate. There is an important difference however, A and B are *operators*, not wave functions, so that the equations are in the algebra itself.

7. Schrödinger equation in operator form.

In a recent paper Brown and Hiley (2000) have shown formally how to obtain the Schrödinger equation in terms of operators from within the quantum algebra itself. We will not repeat the details here but will merely note the results.

We first considered the wave operator $\Psi(A)$, where Ψ is some function of the operator A. If we again use the polar form for the wave operator, we can arrive at the two equations

$$i\frac{d\rho}{dt} + [\rho, H]_- = 0 \tag{5}$$

and

$$\rho\frac{dS}{dt} + \tfrac{1}{2}[\rho, H]_+ = 0 \tag{6}$$

where $\rho = \Psi^*(A))(\Psi(A)$, is the density operator[5]. Equation (5) is an expression for the conservation of probability, while equation (6) is an expression of the conservation of energy when the energy is well defined in the quantum system.

In passing I wish to point out the close relation of our results to those exploited in the Bohm approach to quantum theory. If we express (5) and (6) in the x-representation we arrive at the original Bohm theory. It is straight forward to show that they become

---

[5] It is essential to include the projector onto the standard ket)(. For details see Brown and Hiley (2000).

$$\frac{\partial \rho}{\partial t} + \nabla_r . j = 0 \qquad \text{(Conservation of probability) (7)}$$

and

$$\frac{\partial S}{\partial t} + \frac{(\nabla_r S)^2}{2m} + Q(r,t) + V(r,t) = 0 \qquad \text{(Conservation of energy)} \qquad (8)$$

It can be shown that equation (8) is equivalent to equation (1). Now we see that this extra energy, *Q*(*r*, *t*) is needed to conserve energy because what we have called *p* is only the real part of *p* since

$$\nabla_r S = \Re[\Psi^*(r,t) P \Psi(r,t)] = p_{Bohm} \qquad (9)$$

It is through this identification that the streamlines of the probability current given by equation (7) can be regarded as particle trajectories with p_{Bohm} being the momentum of the particles at any point on the trajectory. These are the usual Bohm trajectories discussed in section 1.

Before going on to explain the meaning of these ideas in the present context, we should also note that (5) and (6) in the *p*-representation become

$$\frac{\partial \rho}{\partial t} + \nabla_p j = 0 \qquad (10)$$

$$\frac{\partial S}{\partial t} + \frac{p^2}{2m} + Q(p,t) + V(\nabla_p S, t) = 0 \qquad (11)$$

Once again we have a quantum potential, but it no longer is a simple expression since it now depends on the form of the classical potential. In the momentum representation the momentum is the observable momentum, but the 'beable' position is now given by

$$\nabla_p S = \Re[\Phi^*(p,t) X \Phi(p,t)] = x_B \qquad (12)$$

Once again we see that we need the quantum potential energy because the classical potential energy is determined by $V(x_B)$ and not by $V(x)$ where x is the observable position. Again, the streamlines of the probability current appearing in equation (10) give the particle trajectories in this representation.

The appearance of two sets of trajectories might at first sight be surprising. The Bohm approach, particularly under the guise of "Bohmian mechanics" has insisted on attributing absolute reality to the particle evolving in a phase space. Bohm himself long abandoned that position. In Bohm and Hiley (1993) we were very careful to present the trajectories in the sense that if we assumed these were particle trajectories, then no inconsistency would arise. In fact not only was the approach consistent, but it was also free of many of the quantum puzzles such as the cat paradox, the measurement problem and so on. At no point did we insist that the particle view corresponded to what actually did take place. We did not insist that our description provided a mechanical approach to quantum processes. I thought we made that clear throughout the book, but it seems that we failed to get the message across. Again I thought the last chapter, which was a very brief summary of the ideas that I am developing in this paper, would have given a strong message that we did not believe the simplistic mechanical approach was viable.

For us the ontology is the notion of activity or process that was described by the algebraic structure of quantum formalism. This can be understood in terms of the implicate order, which in turn, finds it observable consequences in explicate orders.

As we have remarked earlier it is not possible to describe quantum processes in terms of a classical phase space because x and p cannot be defined simultaneously. This is a consequence of the non-commutative structure of the formalism. All we are able to do is to construct shadow phase spaces, each one being an explicate order defined by the context in which it is displayed.

Thus in the example we have discussed above we have two shadow phase spaces, one based on the x-representation, the other based on the p-representation. These are shown in figure below. Each, although different from the classical point of view, is necessary for a full representation of the quantum process. The non-commutativity of the underlying process produces an *ontological complementarity*. This must be contrasted to Bohr's epistemological complementarity.

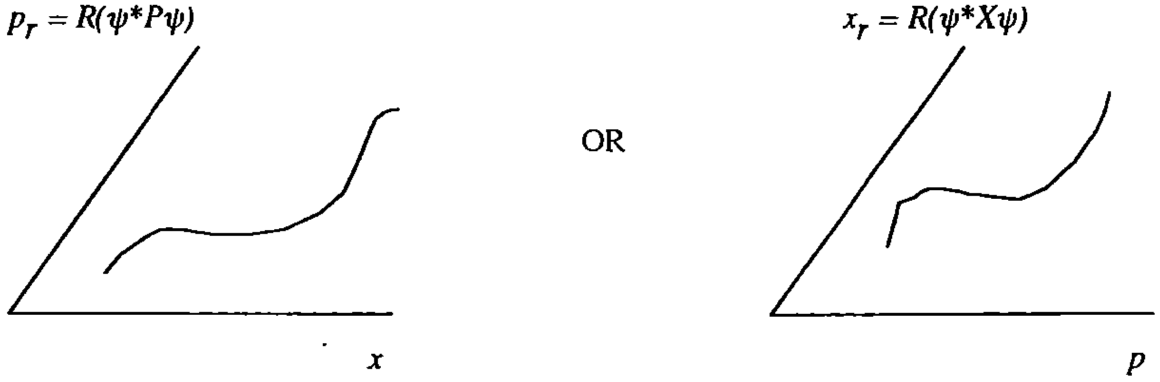


Figure 3. Shadow Manifolds.

Our shadow manifolds are not mutually exclusive, they are complementary and are a consequence of the participatory process of manifestation, or, more conventionally, of observation. Within this description the quantum potential plays the role of an internal energy necessary because of the way we are constructing our shadow manifolds. This potential is totally unlike any classical potential. It has features more akin to a self-organising potential. Indeed this self-organisation occurs in response to the environment in which the quantum process finds itself. In fact we have argued elsewhere that expressing the process in a shadow manifold determines an information dynamics (Hiley 1999)

8. The mind-matter relationship.

What I have tried to argue above is that for quantum processes space-time is not the basic manifold in which quantum processes evolve. The basic process unfolds in this pre-space, which is not subject to the Cartesian division, res extensa-res cogitans. What I want to suggest that it is in this pre-space that mind and matter appear as different aspects of the same underlying process.

Thus mind and matter are united through mutual participation in which separation is not possible. They are two aspects of an indivisible totality, the implicate order. Aspects of this whole activity involve the process of thinking, feeling, desire etc. The dance of the neurons is only the outward material manifestation of these processes. These physical processes are merely an explicate order in which one aspect of the overall process is projected. By restricting our discussions to the electrochemical process of the brain, we miss the deeper implicate order which contains our experience of the physical and mental worlds. But even in using those words, it must not thought that there are two sides, mind and matter. Mind and matter are but different projections from this deeper implicate order where such a division does not exist.

We experience this implicate order directly when we try to explain to others how we feel or think. The words we use are only signifiers that seem to float on a sea of inner energy. We struggle for words to try to capture what is implicit in our thinking. But the meaning is not merely in the words, it is in the context in which those words are used. The context is often implicit and as we try to clarify this context, another, yet deeper context is assumed. But we can never make any complex set of ideas totally explicit. What we do is to try to create in the reader the implicate structure that we feel within ourselves.

Perhaps the clearest example of the role of this implicate order comes from listening to music. Listening is an active experience where we participate in the movement itself. We do not perceive a series of isolated notes. We hear new notes reverberating within the memory of the previous notes. This together with the anticipation of future notes constitutes an unbroken movement. What is apprehended, then, is an undivided state of flowing movement. We can argue that we directly perceive the implicate order because we become part of the total movement. We comprehend movement in terms of a series of inter-penetrating, intermingling elements of different degrees of enfoldment all present together.

To summarise then we have on the one hand mind where the content of thought is displayed in explicate orders, while the process of thinking, feeling, etc occurs through the activity of unfolding and enfolding in the implicate order. At the same time, the display of matter occurs in the explicate order, but its deeper quantum movement occurs through unfoldment and enfoldment in the implicate order. Thus the ground of both thought and matter is in the implicate order. Our task is to find an algebraic description of those aspects of this implicate order where mind and matter have their origins.

9. References.

J. S. Bell (1987a). *Speakable and Unspeakable in Quantum Mechanics*, pp. 1-13, Cambridge University Press, Cambridge.

J. S. Bell (1987b). *Speakable and Unspeakable in Quantum Mechanics*, pp.14-21, Cambridge University Press, Cambridge.

D. Bohm (1952). *Phys. Rev.*, 85, 166-179; 85, 180-193.

D. Bohm (1953). *Phys. Rev.* 89, 319-20.

D. Bohm (1957). *Causality and Chance in Modern Physics*, Routledge & Kegan Paul, London.

D. Bohm (1980). *Wholeness and the Implicate Order*, Routledge, London.

D. Bohm and B. J. Hiley (1987). *Phys. Reports*, **144**, 323-348.

D. Bohm and B. J. Hiley (1993). *The Undivided Universe: an Ontological Interpretation of Quantum Theory*, Routledge, London,

N. Bohr (1925). Letter to Geiger, 21 April 1925, Coll. Works, V, p. 353

N. Bohr (1961). *Atomic Physics and Human Knowledge*, Science Editions, New York,

M. R. Brown and B. J. Hiley (2000). quant-ph/0005026.

G. Chew (1963). *Science Progress*, **51**, 529-539.

W. K. Clifford (1878). *Am. J. Math.*, **1**, 350-358.

A. Crumeyrolle (1990). *Orthogonal and Symplectic Clifford Algebras: Spinor Structures*, Kluwer, Dordrecht.

P. A. M. Dirac (1973), in *The Physicist's Conception of Nature*, p. 10, ed., J. Mehra, Reidel, Dordrecht-Holland.

A. M. Gleason (1957). *J. Math. Mech.* **6**, 885-893.

H. Grassmann (1995). *A New Branch of Mathematics: the Ausdehnungslehre of 1844, and other works*, trans. by L. C. Kannenberg, Open Court, Chicago.

W. R. Hamilton (1837). *Trans. Roy. Irish Acad.*, 17, 293-422.

W. R. Hamilton (1967). *Mathematical Papers*, Vol. 3, *Algebra*, Cambridge University Press, Cambridge,.

W. Heisenberg (1959). *Physics and Philosophy: the revolution in modern science*, p.116, George Allen &Unwin, London.

B. J. Hiley (1991). Vacuum or Holomovement, in *The Philosophy of Vacuum*, ed., S. Saunders and H. R. Brown, pp. 217-249, Clarendon Press, Oxford.

B. J. Hiley and N. A. M. Monk (1993). Mod. Phys. Lett., **A8**, 3225-33.

B. J. Hiley, The Algebra of Process (1995), in *Consciousness at the Crossroads of Cognative Science and Philosophy*, Maribor, Aug. 1994, pp. 52-67.

B. J. Hiley (1997). Quantum mechanics and the relationship between mind and matter, in *Brain, Mind and Physics*, 37-53, ed P. Pylkkänen et al, IOS Press, Helsinki.

B. J. Hiley and M. Fernandes (1997). Process and Time, in *Time, Temporality and Now*, ed., H. Atmanspacher and E. Ruhnau, 365-383, Springer-Verlag.

B. J. Hiley (1999). *Quantum theory, information dynamics and the relationship between mind and matter*, Proc. Int. School of Biophysics "Neuronal Coding of Perceptual Systems, Naples, Italy.

P. Holland (1993). *The Quantum Theory of Motion*, Cambridge University Press, Cambridge.

S. Kochen and E. P. Specker (1967). *J. Math. Mech.* **17**, 59-87.

H. P. Stapp (1993). *Mind, Matter and Quantum Mechanics*, Springer, Berlin.

J. von Neumann (1955). *Mathematical Foundations of Quantum Mechanics*, (Princeton University Press, Princeton.

J. A. Wheeler (1991). *At Home in the Universe*, p .286, AIP Press, New York.

J. A. Wheeler (1980). *Pregeometry: Motivations and Prospects*, in Proc. Symp. Loyola Uni., New Orleans, ed., A. R. Marlow, pp. 1-11, Academic Press, New York.

N. Wiener (1936). *Phil. of Sci.*, **3**, 307.

FROM STRINGS TO CLIFFORD ALGEBRAS

Arleta Griffor

TPRU Birkbeck College
University of London, Mallet Street
London WC1E 7HX
a.griffor@physics.bbk.ac.uk

1 INTRODUCTION

This paper is part of a larger work investigating the mathematical foundations of Clifford and related algebras.[1] It covers most of the material given in my talk *The Algebra of Movement* at last year's ANPA meeting.[2] The paper introduces a new mathematical language in terms of which the algebras can be derived, expressed, and their properties investigated.

The new mathematical language is based on the concept of string defined in the next section. Although the string concept is well known in the context of formal language theory and computing science, the role it plays in the foundations of Clifford algebras has not been noticed before. As the title suggests, the main motivation of the paper is the explication of this role.

Strictly speaking, I don't consider individual strings as elements of algebraic structures introduced in this paper.[3] Rather, I take as the elements equivalence classes of strings defined in sections 3 and 4. In other words, an operation is defined on a setoid (i.e. a set with an equivalence relation) not a set of strings.

The operation, as it is defined in section 5, is none of the standard operations defined on strings (or their sets) in the context of formal language theory. It is a new operation defined in the string context, and I call it a *string product.* In sections 6 and 7 I show that the setoid of strings with the string product is a group isomorphic to a known group (often referred to as the Dirac group).

In section 8 the groups are generalised in such a way that their algebra extensions (sec. 10) are isomorphic to Clifford algebras. Decomposition, classification, and periodicity of Clifford algebras follow directly from decomposition, classification, and periodicity of the groups (sec. 9).[4]

In the last section of the paper I illustrate how to represent real numbers, complex numbers, and quaternions in terms of 'string language'. It is interesting that the quaternion group can be represented in terms of *binary* strings with the defined string product. Going to larger groups, and consequently algebras, would involve consideration of strings over larger alphabets. Of these the alphabet of cardinality 4 might be of particular significance because of its intriguing ubiquity: quantum error correction code, DNA computing, the *E*-number algebra of Eddington, Dirac algebra, the crucial role of the group $G_{4,0}$ in the periodicity law for the groups and their decomposition...

2 STRINGS

The concept of *string* is understood here in the way it is defined in the context of formal language theory. That is, strings are finite ordered multisets of symbols from a finite set called an alphabet. Multisets rather than sets are considered because a string may contain more than one occurrence of the same symbol. More formally, let us introduce the following definitions of strings and related concepts:

Definition 2.1 A finite, nonempty, and ordered set, U, is called an *alphabet*. The elements of U are called *symbols* or *characters*.

Definition 2.2 A finite sequence of symbols from a given alphabet U is called a *string* over the alphabet U. A string over U is denoted by a juxtaposition of symbols from U.

Definition 2.3 The finite sequence that contains no symbols is called the *empty string* and denoted by λ.

The set of all finite strings over U, including the empty string, λ, is denoted U^*.

For example, if $U = \{a, b\}$ then *a*, λ, *aa*, *abbbb*, *ba*, *babababa*, *b* are strings over U, or in other words, they are elements of U^*. An alphabet of cardinality 2 is called a *binary alphabet*, and strings over a binary alphabet are called *binary strings*.

An alphabet of cardinality 1 is called a *unary alphabet*, and strings over it are called *unary strings*. For example, $\{0\}$ is a unary alphabet, and unary strings λ, *0, 00, 000, etc.*, are elements of $\{0\}^*$.

The set of nonempty strings over U, that is, $U^* \setminus \{\lambda\}$, is denoted by U^+.

Definition 2.4 The *length* of a string is defined as the number of symbols in it.

Definition 2.5 Z is called a *substring* of X if $X = YZV$ for some $Y, Z, V \in U^*$.

For example, substrings of the string *abb* are: λ, *b, a ab, bb, abb*.

3 EQUIVALENCE RELATION ON STRINGS

Let U be an alphabet of n symbols, and let U^* denote the set of all strings, including the empty string λ, over U. I define a similarity relation on the set of strings U^*. As the first step, I consider a subset D of U^* consisting of strings with no substrings of repeated symbols of length two. I define a mapping F on U^* that erases such a substring from a string. Any string from U^* can be thus transformed into a string from D by a certain finite, including zero, number of steps, each step being an application of the mapping F. Two strings are regarded as similar if they can be transformed into *the same string* from D by *the same number of steps*. It is easy to show that such a relation of similarity is an equivalence relation on the set of strings U^*.

Let $U = \{x_1, x_2, \dots, x_n\}$, we define a subset D of U^* in the following way

$$D := \{\lambda\} \cup \{ x_i \mid i = 1, 2, \dots n\} \cup \{X = x_{i_1} x_{i_2} \dots x_{i_m} \mid 1 \le k < m \quad i_k \ne i_{k+1}\}$$

Elements of D are: the empty string, strings of length 1, and strings with different adjacent symbols.

Let $X = x_{i_1} x_{i_2} \dots x_{i_m}$, we define a mapping $F\colon U^* \setminus D \rightarrow U^*$ in the following way

$$F(X) = x_{i_1} x_{i_2} \dots x_{i_{k-1}} x_{i_{k+2}} \dots x_{i_m} \qquad k = \min\{1 \le j < m \mid i_j = i_{j+1}\}$$

The mapping F *erases* from the string X the first (from the left) substring of two repeated symbols. Note that F is defined for $X \notin D$ only.

For example, for $U = \{a, b, c, \ldots\}$

$$F(aa) = \lambda$$
$$F(abbbdda) = abdda$$
$$F(abdda) = aba$$

Let us define

$$F^0(X) = X$$
$$F^i(X) = F(F^{i-1}(X)) \quad \text{for } i \geq 1$$

It is easy to show that $\forall X \in U^*$ there is only one $k \geq 0$ such that $F^k(X) \in D$.

We define the following relation of *similarity* between strings

$$\forall X, Y \in U^* \quad X \approx Y \quad \textit{iff}$$

$$\exists k \quad F^k(X), F^k(Y) \in D \quad \text{and} \quad F^k(X) = F^k(Y)$$

The relation of similarity is an *equivalence relation* on U^*.[5]

4 EQUIVALENCE CLASSES OF STRINGS

The equivalence relation on U^* gives rise to a partition of the set U^* into equivalence classes. Each class consists of a set of equivalent strings. Let us denote the set of equivalence classes by G.

$$G := U^*/_{\approx} = \{[X] \mid X \in U^*\}$$

Note that the elements of G are *sets* of strings.

The following two lemmas state that if we replace a substring of identical symbols of length two with another substring of identical symbols of length two in a given string we obtain a string which is equivalent to the original string.

Lemma 4.1

$$[XaaY] = [XbbY] \qquad \forall a, b \in U \text{ and } \forall X, Y \in U^*$$

The next lemma follows immediately as a particular case when the substrings X and Y are empty strings.

Lemma 4.2

$$[aa] = [bb] \qquad \forall a, b \in U$$

The lemmas 4.1 and 4.2 allow us to define the following notation:

$$[e] := [aa] = \{aa \mid a \in U\}$$
$$[XeY] := [XaaY] = \{XaaY \mid a \in U \text{ and } X, Y \in U^*\}$$

The next two lemmas state that any two strings which differ only in the position of their substring e (i.e., a substring of repeated symbols of length 2) are equivalent.

Lemma 4.3

$$[eXY] = [XeY] = [XYe] \qquad \forall X, Y \in U^*$$

More generally

Lemma 4.4

$$[eX_1X_2 \ldots X_k] = [X_1 eX_2 \ldots X_k] = \ldots = [X_1X_2 \ldots X_k\, e] \qquad \forall X_i \in U^*$$

The lemmas 4.1 – 4.4 easily generalise to the case when the number of substrings of repeated symbols of length two in a given string X is greater than one. We can therefore define the following notation

$$[ke] := [eeeee\ldots] := \{a_1a_1a_2a_2a_3a_3\ldots a_ka_k \mid a_i \in U\}$$
$$[XkeY] := \{Xa_1a_1a_2a_2a_3a_3\ldots a_ka_kY \mid a_i \in U \text{ and } X, Y \in U^*\}$$

We have 'movability' of e

$$[XkeY] = [keXY] = [(k\text{-}l)eXYle] = [(k\text{-}l\text{-}m)eXmeYle] = [XYke]$$

that can be easily extended to a generalisation of lemma 4.4 for larger than two number of substrings X_i of a given string.

Using the notation introduced here note that if two strings X and Y are equivalent then

$$\exists\, k \text{ and } \exists\, Z \in D \qquad [X] = [keZ] \text{ and } [Y] = [keZ]$$

5 STRING PRODUCT

Let us define the following *string product* on the elements of *G*

$$[x_1x_2 \dots x_{k-1}x_k][y_1y_2 \dots y_m] = \begin{cases} [x_1x_2 \dots x_{k-1}y_2 \dots y_m] & \text{for } x_k = y_1 \\ [x_1x_2 \dots x_{k-1}y_1x_ky_2 \dots y_m] & \text{for } x_k \neq y_1 \end{cases}$$

$\forall x_i, y_i \in U$

For example, $\forall a, b \in U$ such that $a \neq b$

$[a][a] = [\lambda]$
$[a][b] = [ba]$
$[b][a] = [ab]$
$[aa][a] = [a]$
$[ba][ab] = [bb]$
$[ab][ab] = [aabb]$

Note that strictly speaking the string product is not everywhere defined on *G*. It is defined on the equivalence classes that can be represented by nonempty strings. Since the equivalence class of the empty string $[\lambda]$ contains only one string λ, it cannot be represented by a nonempty string. The product is therefore defined on $G \setminus \{[\lambda]\}$. We can, however, extend the definition of the string product to all the elements of *G* by using lemma 5.1.

Lemma 5.1

$$[\lambda] = [ee]$$

Proof. From lemma 4.2 we have that for every $a,\ b \in \boldsymbol{U}$, $[aa] = [bb]$. Let us assume that $a \neq b$, and multiply the equation $[aa] = [bb]$ on the right by $[a]$. We obtain $[aa][a] = [bb][a]$, and therefore $[a] = [bab]$.[6]

From the product definition we have that $\forall\ a \in \boldsymbol{U}$, $[\lambda] = [a][a]$. We have shown that $[a] = [bab]$, by substituting $[bab]$ for $[a]$, we therefore get

$$[\lambda] = [a][a] = [a][bab] = [baab] \qquad \forall\, a,\ b \in \boldsymbol{U} \text{ such that } a \neq b$$

Using lemma 4.3 and the notation introduced in the previous section, we may therefore write

$$[baab] = [beb] = [ebb] = [ee] \qquad \ast$$

Now we are in the position to extend the definition of the string product to all the elements of $\boldsymbol{G}$ by defining

$$\begin{aligned} &[\lambda][X] := [ee][X] \\ &[X][\lambda] := [X][ee] \\ &[\lambda][\lambda] := [ee][ee] \qquad\qquad \forall\, [X] \in \boldsymbol{G} \setminus \{[\lambda]\} \end{aligned}$$

6 GROUP STRUCTURE

In this section I will show that the set G with the string product everywhere defined on G is a group.

Lemma 6.1

The element $[e] = [aa]$ $\forall\, a \in U$ is the two sided identity in G.

Proof. Each element of G, including $[\lambda]$, can be expressed as $[X] = [x_1 x_2 \dots x_{k-1} x_k]$, where $k \geq 1$, and $x_i \in U$. Because $\forall\, a \in U$, $[e] = [aa]$, it follows that $[e] = [x_1 x_1] = [x_k x_k]$. We thus have

$$[e][X] = [x_1 x_1][x_1 x_2 \dots x_{k-1} x_k] = [x_1 x_2 \dots x_{k-1} x_k] = [X]$$
$$[X][e] = [x_1 x_2 \dots x_{k-1} x_k][x_k x_k] = [x_1 x_2 \dots x_{k-1} x_k] = [X] \qquad \ast$$

To show that the string product is associative, we need the next three lemmas.

Lemma 6.2

$$[e] = [\lambda][\lambda]$$

Proof. From the definition of $[e]$, and the definition of the product, we have

(1) $[e] = [ab][ba]$

$[ab] = [ab][bb] = [ab][aa] \qquad \forall\, a, b \in U$

therefore $\forall\, a,\ b \in U$, such that $a \neq b$

$$[ab] = [ab][aa] = [aaba]$$

If we substitute the calculated value of $[ab]$, which is $[aaba]$, into the equation (1), from the product definition we get

$$[e] = [ab][ba] = [aaba][ba] = [aabbaa] = [eee] = [ee][ee] = [\lambda][\lambda] \quad ✱$$

Lemma 6.3

$$[X] = [eeX] \qquad \forall\ [X] = [x_1x_2 \ldots x_{k-1}x_k] \in G$$

Proof. From the previous lemma, and $[e]$ being the identity we have

$$[e] = [\lambda][\lambda] = [ee][ee] = [eee]$$
$$[X] = [e][X] = [eee][X] = [eex_1x_1][x_1x_2 \ldots x_{k-1}x_k] =$$
$$[eex_1x_2 \ldots x_{k-1}x_k] = [eeX] \quad ✱$$

Lemma 6.4

$$[X][Y] = [\lambda][XY] = [eXY] \qquad \forall\ [X], [Y] \in G$$

Proof. Let $[X] = [x_1x_2 \ldots x_{k-1}x_k]$ and $[Y] = [y_1y_2 \ldots y_{m-1}y_m]$. From lemmas 5.1 and 6.3, and from lemma 4.3 it follows that

$$[X][Y] = [Xee][Y] = [Xey_1y_1][y_1y_2 \ldots y_{m-1}y_m] = [XeY] = [eXY]$$

and

$$[\lambda][XY] = [ee][XY] = [ex_1x_1][x_1x_2 \ldots x_{k-1}x_k Y] = [eXY] \quad ✱$$

Now, we are able to prove that the induced string product is associative.

Lemma 6.5

$$([X][Y])[Z] = [X]([Y][Z]) \qquad \forall\ [X], [Y], [Z] \in G$$

Proof. Let $[X] = [x_1 x_2 \ldots x_{k-1} x_k]$ and $[Z] = [z_1 z_2 \ldots z_{m-1} z_m]$, from lemma 6.4 and lemma 4.3, it follows that

$$([X][Y])[Z] = [eXY][Z] = [XYe][Z] = [XYz_1z_1][z_1z_2 \ldots z_{m-1}z_m] = [XYZ]$$
$$[X]([Y][Z]) = [X][eYZ] = [x_1x_2 \ldots x_{k-1}x_k][x_kx_kYZ] = [XYZ] \qquad ✱$$

Lemma 6.6

$$[X_1X_2X_3 \ldots X_k] = [\lambda]^{k-1}[X_1][X_2][X_3] \ldots [X_k] \qquad \forall\ [X_i] \in G \text{ and } k > 1$$

Proof. From lemma 6.4, it follows that

$$[X_1][X_2][X_3] \ldots [X_k] = [\lambda]^{k-1}[X_1X_2X_3 \ldots X_k]$$

Multiplying left sides of the equation by $[\lambda]^{k-1}$, we get, using the lemma 6.2

$$[\lambda]^{k-1}[X_1][X_2] \ldots [X_k] = [\lambda]^{2(k-1)}[X_1X_2 \ldots X_k] = [X_1X_2 \ldots X_k] \qquad ✱$$

Corollary 6.6.1

$$[ke] = [\lambda]^{k-1}$$

Proof. If we take $\forall i\ [X_i] = [e]$ then lemma 6.6 implies

$$[ke] = [\lambda]^{k-1}[e]^k = [\lambda]^{k-1} \qquad ✱$$

Corollary 6.6.2

$$[keX] = [\lambda]^{k}[X] \qquad \forall\, [X] \in \boldsymbol{G}$$

Proof. From 6.6 and 6.6.1 it follows that

$$[keX] = [\lambda][ke][X] = [\lambda][\lambda]^{k-1}[X] = [\lambda]^{k}[X] \qquad ✱$$

Corollary 6.6.3

$$[X] = [\lambda]^{k-1}[x_1][x_2]\ \dots [x_k] \qquad \forall\, [X] = [x_1x_2 \dots x_k] \in \boldsymbol{G} \text{ such that } x_i \in \boldsymbol{U}$$

Proof. If we take $\forall i \quad [X_i] = [x_i]$ then lemma 5.6 implies

$$[X] = [x_1x_2 \dots x_k] = [\lambda]^{k-1}[x_1][x_2]\ \dots [x_k] \qquad ✱$$

Lemma 6.7

$\forall\, [X] \in \boldsymbol{G}$ there is an inverse $[X]^{-1} \in \boldsymbol{G}$, such that $[X][X]^{-1} = [X]^{-1}[X] = [e]$

Proof. Let $[X] = [x_1x_2 \dots x_{k-1}x_k]$ and $[X]^{R} = [x_kx_{k-1} \dots x_2x_1]$. From lemmas 6.6 and 6.2 it follows that

$$[X][X]^{R} = [\lambda]^{k-1}[x_1][x_2][x_3]\ \dots [x_k][\lambda]^{k-1}[x_k][x_{k-1}][x_{k-2}]\ \dots [x_1] = [\lambda]^{k}$$
$$[X]^{R}[X] = [\lambda]^{k-1}[x_k][x_{k-1}][x_{k-2}]\ \dots [x_1][\lambda]^{k-1}[x_1][x_2][x_3]\ \dots [x_k] = [\lambda]^{k}$$

Therefore for k even $[X]^{-1} = [X]^{R}$, and for k odd $[X]^{-1} = [\lambda][X]^{R}$ ✱

Lemma 6.7 completes the demonstration that the set $\boldsymbol{G}$ of equivalence classes of strings with the defined string product is a group.

7 GROUP PROPERTIES

From now on lemmas will be given without proofs. These will appear elsewhere.

Lemma 7.1

Each element of G, different from $[e]$ and $[\lambda]$, can be uniquely expressed as

$$[X] = [\lambda]^k[x_{i_1}x_{i_2}\dots x_{i_m}] \qquad \text{for } k = 1,\ 2 \text{ and } 1 \leq i_1 < i_2 < \dots i_m \leq n$$

Corollary 7.1.1

From lemma 7.1 it follows that each element of G, different from $[e]$ and $[\lambda]$, can be expressed as

$$[X] = [x_{i_1}x_{i_2}\dots x_{i_m}] \qquad \text{where } i_k \neq i_l \text{ for } 1 \leq k, l \leq m$$

Note that such an expression is in general not unique with regard to the *order* of the symbols. For example, $[X] = [x_1x_2x_3] = [x_2x_3x_1]$. However, the *set* of symbols is the same in each such an expression of a given element of G.

Let us thus define the symbol Δ_m

$$\Delta_m := [x_{i_1}x_{i_2}\dots x_{i_m}] \qquad \text{where } i_k \neq i_l \text{ for } 0 \leq k, l \leq m$$

That is, Δ_m denotes an element of G of length m with no repeated symbols in it.

The following lemma about the order of the group G follows directly from lemma 6.4 and 7.1.

Lemma 7.2
Let $U = \{x_1, x_2, \dots, x_n\}$ be an alphabet of n symbols. Then the order of the group G, defined as the set of equivalence classes of strings over U with the string product, is 2^{n+1}.

Thus, if we begin with an alphabet of n symbols, the number of the elements in the group G will be 2^{n+1}.

The following lemma allows us to determine the order of the elements in the group G.

Lemma 7.3

$$\Delta_k\Delta_k = [e] \qquad \text{for } k = 0, 3 \ mod\ 4$$
$$\Delta_k\Delta_k = [\lambda] \qquad \text{for } k = 1, 2 \ mod\ 4$$

From lemma 7.3 and corollary 7.1.1 it follows that the elements of G are of order 1, 2 or 4.

The next lemma express the law of *commutativity* for the elements of G.

Lemma 7.4
Let c be the number of *common symbols* in Δ_k and Δ_m, then

$$\Delta_k\Delta_m = [\lambda]\Delta_m\Delta_k \qquad \text{for } km - c^2 \text{ odd}$$
$$\Delta_k\Delta_m = \Delta_m\Delta_k \qquad \text{for } km - c^2 \text{ even}$$

In particular, lemma 7.4 implies that the centre of the group G is $\{[e], [\lambda]\}$ for n even, and $\{[e], [\lambda], [x_1x_2 \dots x_{n-1}x_n], [ex_1x_2 \dots x_{n-1}x_n]\}$ for n odd.[7]

For an alphabet $U = \{x_1, x_2, \dots , x_n\}$ of n symbols, the group G can be considered as generated by the set of elements: $\{[x_1], [x_2], \dots [x_n]\}$.

From now on we shall use the notation $G_{n,0}$ for the group G whose minimal set of generators is a set of n elements of order 4. For example

$$G_{n,0} = \langle [x_1],[x_2], \dots [x_n]\rangle \qquad [x_i][x_i] = [ee] = [\lambda] \qquad \text{for } 1 \leq i \leq n$$

The group $G_{n,0}$ is isomorphic to a known group Γ, defined as follows[8]

$$\Gamma = \langle \gamma_1, \gamma_2, \dots \gamma_n\rangle$$

$$\gamma_i\gamma_i = -1 \qquad \text{for } 1 \leq i \leq n$$
$$\gamma_i\gamma_k = - \gamma_k\gamma_i \qquad \text{for } i \neq k$$

The isomorphism of $G_{n,0}$ and Γ is of the form

$$[x_i] \rightarrow \gamma_i \qquad \text{for } 1 \leq i \leq n$$

8 GROUP GENERALIZATION

In this section I introduce a group denoted $G_{p,q}$, whose minimal set of generators consists of p elements of order 4, and q elements of order 2, such that $p + q = n$.

Let $G_{n,0} = \langle [x_1],[x_2], \ldots [x_n] \rangle$, and $\Delta_2 = [x_\alpha x_\beta]$, where the string $x_\alpha x_\beta$ is not a string over $U = \{x_1, x_2, \ldots, x_n\}$. Multiplying q generators of $G_{n,0}$ by Δ_2 we get a new set of generators $\{[x_1],[x_2], \ldots [x_p], [x_{p+1}]\Delta_2, \ldots [x_n]\Delta_2\}$.

We define the group $G_{p,q}$ as a group generated by the new set of generators, that is

$$G_{p,q} := \langle [x_1],[x_2], \ldots [x_p], [x_{p+1}]\Delta_2, \ldots [x_n]\Delta_2 \rangle$$

The group $G_{p,q}$ is a generalization of the group $G_{n,0}$ in the sense that in a particular case when $q = 0$, the group $G_{p,q}$ becomes the group $G_{n,0}$.

From lemma 7.4 it follows that the generators of $G_{p,q}$ 'anticommute' with each other. From lemma 7.3 it follows that p generators of the form $[x_i]$ are of order 4, and q generators of the form $[x_i]\Delta_2$ are of order 2.[9]

Lemma 8.1
For $p + q = n$ the order of the group $G_{p,q}$ is 2^{n+1}.

Groups isomorphic to the groups $G_{p,q}$ were studied by N. Salingaros[10], H. W. Branden[11], G. Bergdolt[12], and T. L. Smith[13]. The authors consider them as finite groups associated with (or derivable from) Clifford algebras generated by the algebras' basis units.

9 GROUP CLASSIFICATION AND DECOMPOSITION

To construct a classification[14] table for *all* the groups $G_{p,q}$ with their (horizontal) isomorphisms and (vertical and diagonal) 4-periodicity explicitly displayed, we need lemmas 9.1, 9.2, 9.3, and the initial (4 × 4) square block of groups, that is, the groups $G_{p,q}$ for $p,\ q < 4$.

Lemma 9.1

$$G_{p,q} \cong G_{q-1,p+1}$$

Lemma 9.2

$$G_{p+4,q} \cong G_{p,q+4}$$

Corollary 9.2.1

$$G_{p,q} \cong G_{p-4,q+4}$$
$$G_{p,q} \cong G_{p+4,q-4}$$

Lemma 9.3

$$G_{p+4,q} \cong G_{4,0} \copyright G_{p,q}$$

where © is a *central* (associative and commutative) product of groups.

Because of lemma 9.2 we have 'diagonal' and 'vertical' 4-periodicity expressed in following lemma

Corollary 9.3.1

$$G_{p+4,q} \cong G_{p,q+4} \cong G_{4,0} \copyright G_{p,q} \qquad \text{(diagonal 4-periodicity)}$$
$$G_{p+8,q} \cong G_{p+4,q+4} \cong G_{4,0} \copyright G_{4,0} \copyright G_{p,q} \qquad \text{(vertical 4-periodicity)}$$

In more general form, for any multiplicity of 4 we have corollary 9.3.2

Corollary 9.3.2

$$G_{p+4k,q+4l} \cong G_{p+4(k+l),q} \cong G_{p,q+4(k+l)} \cong (G_{4,0})^{k+l} \copyright G_{p,q} \qquad \text{for } k,\ l \geq 0$$

where $(G_{4,0})^{k+l}$ denotes a central product of $k + l$ copies of $G_{4,0}$.

The corollary 9.3.2 implies that any group $G_{p,q}$ is a central product of groups $G_{4,0}$ and $G_{p,q}$ where $p,\ q < 4$. Knowing the groups $G_{p,q}$ for $p,\ q < 4$ would thus let us know the explicit form of decomposition of any group $G_{p,q}$ into a central product of smaller groups.

For $G_{p,q}$ where $p,\ q < 4$ we have the following isomorphisms of groups:

(1)	$G_{0,0} \cong C_2$	the cyclic group of order 2
(2)	$G_{1,0} \cong C_4$	the cyclic group of order 2
(3)	$G_{0,1} \cong C_2 \times C_2$	the Klein 4-group
(4)	$G_{2,0} \cong Q$	the quaternion group
	$G_{1,1} \cong G_{0,2}$	from 9.1
(5)	$G_{0,2} \cong D$	the dihedral group

$G_{3,0} \cong G_{2,0} \copyright G_{0,1}$

$G_{2,1} \cong G_{2,0} \copyright G_{1,0}$

$G_{1,2} \cong G_{0,2} \copyright G_{0,1}$

$G_{0,3} \cong G_{2,1}$ from 9.1

$G_{3,1} \cong G_{2,0} \copyright G_{0,2}$

$G_{2,2} \cong G_{2,0} \copyright G_{2,0} \cong G_{0,2} \copyright G_{0,2}$

$G_{1,3} \cong G_{2,2}$ from 9.1

$$G_{3,2} \cong G_{2,0} \copyright G_{0,2} \copyright G_{1,0}$$
$$G_{2,3} \cong G_{2,0} \copyright G_{2,0} \copyright G_{0,1}$$

$$G_{3,3} \cong G_{2,0} \copyright G_{2,0} \copyright G_{0,2}$$

From 9.1 it follows that $G_{4,0} \cong G_{3,1}$, and thus $G_{4,0} \cong G_{2,0} \copyright G_{0,2}$. Therefore for any p and q the group $G_{p,q}$ is a central product of groups (1) – (5). For any p and q the explicit form of the product is given in the lemma 9.4.

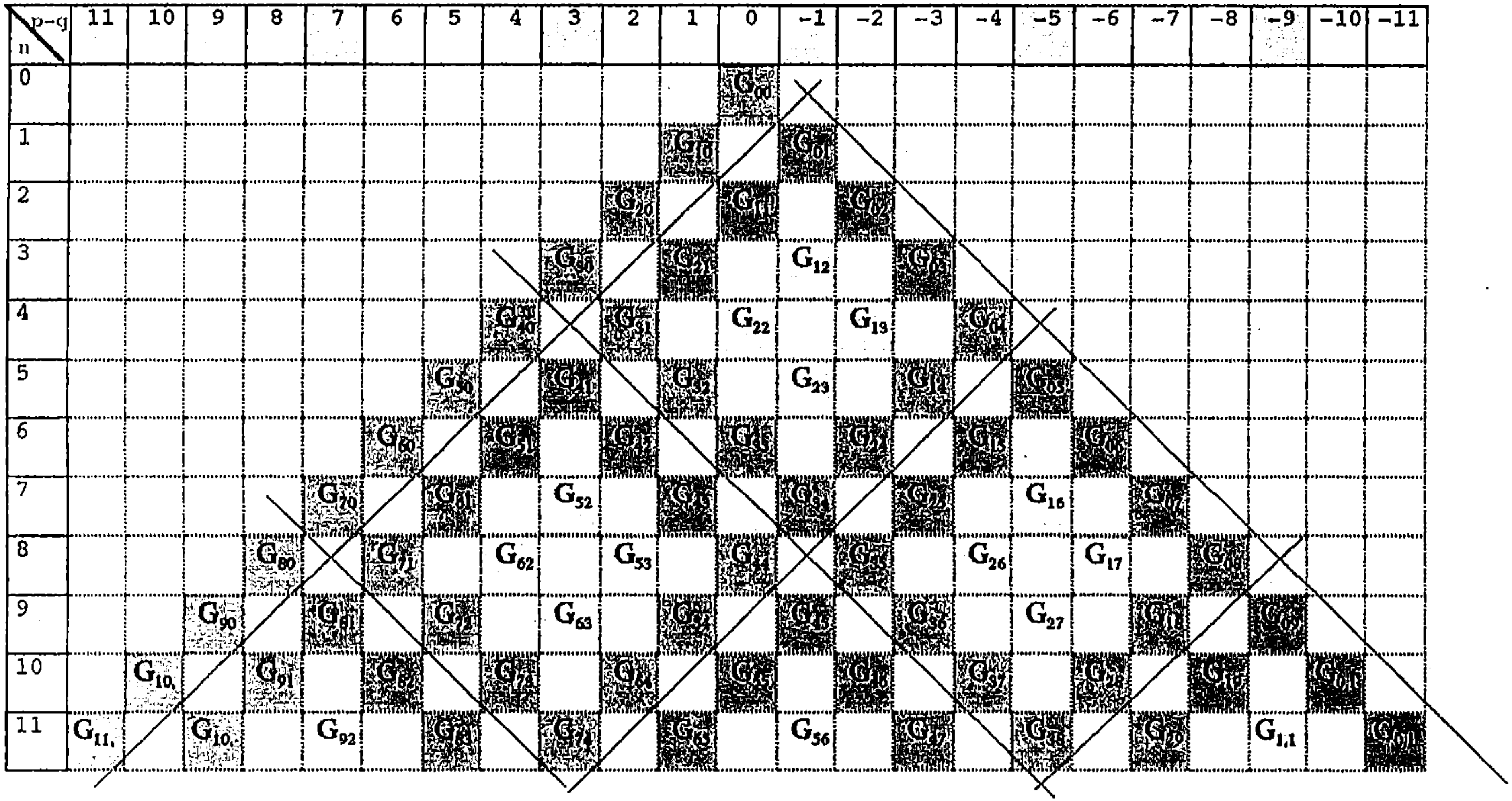
Lemma 9.4

$$G_{p,q} \cong (G_{4,0})^{r+s} \copyright G_{p',q'} \cong (G_{2,0} \copyright G_{0,2})^{r+s} \copyright G_{p',q'}$$

where

$$p = 4r + p'$$
$$q = 4s + q' \qquad r, s \geq 0 \text{ and } p', q' < 4$$

CLASSIFICATION OF GROUPS $G_{p,q}$



p-q / n	11	10	9	8	7	6	5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
0												G_{00}											
1											G_{10}		G_{01}										
2										G_{20}		G_{11}		G_{02}									
3									G_{30}		G_{21}		G_{12}		G_{03}								
4								G_{40}		G_{31}		G_{22}		G_{13}		G_{04}							
5							G_{50}		G_{41}		G_{32}		G_{23}		G_{14}		G_{05}						
6						G_{60}		G_{51}		G_{42}		G_{33}		G_{24}		G_{15}		G_{06}					
7					G_{70}		G_{61}		G_{52}		G_{43}		G_{34}		G_{25}		G_{16}		G_{07}				
8				G_{80}		G_{71}		G_{62}		G_{53}		G_{44}		G_{35}		G_{26}		G_{17}		G_{08}			
9			G_{90}		G_{81}		G_{72}		G_{63}		G_{54}		G_{45}		G_{36}		G_{27}		G_{18}		G_{09}		
10		$G_{10,}$		G_{91}		G_{82}		G_{73}		G_{64}		G_{55}		G_{46}		G_{37}		G_{28}		G_{19}		G_{01}	
11	$G_{11,}$		$G_{10,}$		G_{92}		G_{83}		G_{74}		G_{65}		G_{56}		G_{47}		G_{38}		G_{29}		$G_{1,1}$		G_{01}

10 FROM GROUPS TO ALGEBRAS

Beginning with the group $G_{p,q}$ we construct an algebra $A_{p,q}$ in two steps; as the first step we extend the group into a ring, and as the second step we extend the ring into an algebra.

The group $G_{p,q}$ has one operation, the string product, defined. We introduce a second (commutative and associative) operation on the elements of the group. The second operation we call *addition*, and symbolise by '+'. We assume that the product is distributive over addition, and we define a new element, denoted by [ ]

$$[\ ] := [e] + [ee] = [e] + [\lambda]$$

The element [ ] is assumed to be the 'zero' or the identity of the addition.

The resulting structure is a *ring*. The elements {[ ], [e], [ee]} lie in the centre of the ring. The set of elements, {[ ], [e], [ee]} with the addition and the string product is a finite field, F_3, isomorphic to the field Z_3.

We define an *algebra*, $A_{p,q}$, as the ring over the field F_3. The algebra is isomorphic to the Clifford algebra $Cl_{p,q}$ over the finite field Z_3.

Note that the field Z_3 is of characteristic 3.

We can now extend the field of the algebra into a larger field which contains as a subfield a field isomorphic to F_3. None of such larger fields will be therefore of characteristic 2.[15] The resulting algebras are then isomorphic to Clifford algebras $Cl_{p,q}$ over these fields. The dimension of $Cl_{p,q}$ will be 2^{p+q}, and the signature of its vector space will be (p, q).

In general, the classification table for Clifford algebras, including their isomorphisms and periodicities derives from the classification of the groups $G_{p,q}$, including their isomorphisms and periodicities indicated in this paper.[16]

11 EXAMPLES

Real Numbers

The group $G_{0,0}$ is associated with the Clifford algebra $G_{0,0}$. To construct the group, we begin with the alphabet of cardinality 0, that is, $U = \emptyset$. The set of strings over the empty alphabet U consists of one element only, the empty string λ, that is, $U^* = \{\lambda\}$. The group $G_{0,0}$ is generated by $[\lambda]$, $G_{0,0} = \langle[\lambda]\rangle$, and its elements are $\{[\lambda], [e]\}$.

The group $G_{0,0}$ is an untypical or *virtual* case in the sense that it is derived from an empty alphabet. There are no symbols available to define the string product and the elements $[\lambda]$ and $[e]$. However, we may always consider the group as a subgroup of a larger group derived from a non-empty alphabet.

For example, we may begin with an unary alphabet of one symbol $U = \{0\}$. The group $G_{0,0}$ is then isomorphic to a subgroup of $G_{1,0}$ generated by $[0000]$. That is, $G_{0,0} \cong \langle[0000]\rangle$ and its elements are $\{[0000], [00]\}$. We then understand $[\lambda]$ as $[0000]$, and $[e]$ as $[00]$, as they are generally defined in the context of groups derived from nonempty alphabets.

Complex Numbers

The group $G_{1,0}$ is associated with the algebra of complex numbers $Cl_{1,0}$. It is derived, as mentioned above, from a unary or one element alphabet, for example, $U = \{0\}$. Strings over U are called *unary* strings. [17]

The group $G_{1,0}$ may be considered as generated by $[0]$, that is, $G_{1,0} = \langle[0]\rangle$. The elements of the group are $\{[0], [e], [e0], [\lambda]\}$. We can write the elements as $\{[0], [00], [000], [0000]\}$ in terms of symbols from the alphabet $U = \{0\}$.

There are the following isomorphisms between the elements of $G_{1,0}$ and the complex numbers in standard notation

$$[0] \cong i$$
$$[00] \cong 1$$
$$[000] \cong -i$$
$$[0000] \cong -1$$

Examples of calculations on the elements of $G_{1,0}$

$$[0][0] = [\lambda] = [0000] = [ee]$$
$$[0000][0] = [\lambda][0] = [000] = [e0]$$
$$[\lambda][0][0] = [\lambda][\lambda] = [e]$$
$$[000][0] = [00] = [e]$$
$$[000][000] = [0000]$$

Quaternions

The group $G_{2,0}$ is associated with the Clifford algebra $Cl_{2,0}$. To construct the group, we begin with an alphabet of cardinality 2, for example, $U = \{0, 1\}$.

Strings over U are called *binary strings*. The group $G_{2,0}$ is generated by $[0]$ and $[1]$, that is, $G_{2,0} = \langle [0], [1] \rangle$.

The elements of the group are $\{[0], [e], [e0], [\lambda], [1], [01], [e1], [e01]\}$ which can be expressed as $\{[0], [00], [000], [0000], [1], [01], [001], [0001]\}$ for example, using only the symbols from the alphabet $U = \{0, 1\}$.

We have the following example of an isomorphism of the group's elements and the quaternion group elements in standard notation

$$[0] \cong \boldsymbol{i}$$
$$[1] \cong \boldsymbol{j}$$
$$[0001] \cong \boldsymbol{k}$$

Examples of calculations on the elements of $G_{2,0}$

$$[0][1] = [e01] = [0001]$$
$$[0][1] = [10]$$
$$[1][0001] = [1][10] = [0]$$
$$[0][0001] = [001] = [e1] = [\lambda][1]$$

---

[1] By related algebras I mean tensor products of Clifford algebras, Grassmann and Muses algebras, symplectic Clifford algebras, and possibly some other algebras expressible in terms of 'string' language.

[2] The initial motivation in developing what eventually became the present structure was to modify certain algebraic aspects of B. Hiley's *Algebra of Process*. (See, B. Hiley and M. Fernandes, "Process and Time" in *Time, Temporatily, Now*, ed. H. Atmanspacher and E. Ruhnau, Springer 1997, and references therein for the details of proposed interpretation, also an interview with B. Hiley in *Aspects II, Proceedings of ANPA 20*, May 1999). Since it was not the interpretation I was primarily concerned with, the tacit assumption was that the modified algebraic structure may well have the interpretation as originally proposed. By changing the title I want to indicate that the original interpretation of the mathematical structure is not necessarily unique. Rather, it represents one of possible interpretations.

---

[3] There is another version of the paper to be published elsewhere, where I take individual strings as the set of elements, and define an associative version of the same (string) product on it. The groups obtained in this way are isomorphic to the groups obtained in this paper.

[4] Relations between groups isomorphic to those introduced in this paper and Clifford algebras were studied before (Refs 10-12). The results I am discussing here were obtained independently and in a somewhat different way. Details will be published elsewhere.

[5] The proof is quite easy and based essentially on the fact that for each string there is *only one* k such that $\boldsymbol{F^k(X) \in D}$

[6] We write $[aa] = [bb]$ because these are different forms of the same element of $\boldsymbol{G}$. For the product definition to be consistent, (i.e., the product is a *function* or a *mapping*), it is necessary that the same arguments are mapped into the same images, that is if $[aa] = [bb]$ and $[a] = [a]$, then $[aa][a] = [bb][a]$, and therefore $[a] = [bab]$.

[7] The laws of commutativity for the group $\boldsymbol{G}$ are identical (as I found out later) to the laws that were for the first time introduced by WK Clifford in *Applications of Grassmann's Extensive Algebra* in *American Journal of Mathematics*, 1878, Vol. I, 350-58.

[8] See, for example J.S. Lomont, *Applications of Finite Groups*, Dover 1959. For n = 4 the group is often called the Dirac group, or the group of gamma matrices. C. W. Clifford (ref. 7) considers its properties calling it the *algebra of units*.

[9] Note that $[x_i]A_2 = [x_i][x_\alpha x_\beta] = [\lambda][x_i x_\alpha x_\beta]$.

[10] N. Salingaros, "Realization, extension, and classification of certain physically important groups and algebras", J.Math. Phys. 22, 226-232 (1981), "On the classification of Clifford algebras and their relation to spinors in *n* dimensions", J.Math. Phys. 23, 1-7 (1982), "The relationship between finite groups and Clifford algebras", J.Math. Phys. 25, 738-742 (1985).

[11] H. W. Braden, "*N*-dimensional spinors: Their properties in terms of finite groups", J.Math. Phys. 26(4), 613-620 (1985).

[11] T. Y. Lam, T. Smith, "On the Clifford-Littlewood-Eckmann groups: A new look at periodicity mod 8", Rocky Mountain Journal of Mathematics, 19(3), 749-786 (1989).

[12] G. Bergdolt, "The complete reduction of Clifford algebras", J.Math. Phys. 34(12), 5924-5934 (1993).

[14] For another approach to the classification and decomposition of the groups, based on properties of extraspecial groups, see N. Salingaros and H. W. Braden.

[15] We thus come in a natural way to the fact always assumed in the context of defining Clifford algebra that the field cannot be of characteristics 2.

[16] For details or relationship between the groups and Clifford algebras see refs. 10-12.

[17] See section 1 of this paper for examples of strings over $\{0\}$.

Extended Causality as a Model of Discrete Gravity

Brian D. Koberlein

Department of Physics, SUNY Empire State College,
1475 Winton Road N., Rochester, NY 14609

August 2000

Applying the principle of causal symmetry to the relativistic form of causality, the concept of extended causality is derived. It is shown this concept may be used to generate a discrete classical field with many quantum-like qualities. Applying extended causality to a classical scalar field, a simple solution is obtained. This solution is demonstrated to have the properties of a "classical graviton."

I. Introduction

At ANPA 20, an extension of general relativity was presented [1]. The generalization was necessary to allow for the possibility of matter-antimatter gravitational repulsion, as predicted by Noyes and Starson [5]. Central to the Noyes-Starson model was a generalized concept of symmetry. In this work another form of symmetry, known as causal symmetry, is applied. In Section II, the concept of causal symmetry is introduced and used to derive the principle of extended causality. Section III applies this principle to the derivation of a discrete scalar field. In Section IV, a simple solution of this field is obtained. The field is shown to represent a "classical graviton." Finally, Section V outlines areas for further study.

II. Extended Causality

Typically when one thinks of symmetry one imagines spatial symmetry, such as reflection or rotation. But at its most basic level, symmetry is simply the lack of change within a system where change (at least in theory) occurs [2]. If, for example, one performs the Michelson-Morley experiment at a particular place and time, and then again at another place and time, the measured speed of light will be the same. This is a form of causal symmetry, in which equivalent causes (the experiments) yield equivalent effects (the results). By applying this symmetry to the hypercone structure of free objects, it is possible to extend the concept of cauality, creating a finite hypercone. This in turn generates a discrete finite field.

Consider a pair of points, A and B, in a Minkowski spacetime. In general these points define a 4-vector Δx, however in order for A and B to represent two observations of a free object (particle or field), Δx must satisfy the causality constraint

$$\Delta\tau^2 + \Delta x^2 = 0, \tag{2.1}$$

where τ is a real-valued parameter, known in relativity as the proper time. Equation (2.1) defines a local hypercone for the point A, which is the free object support. The object must lie on the hypercone, and B is assumed to define its path. The angle θ, given by $\tan\theta = |\Delta x|/|\Delta\tau|$, is a measure of the free object's speed. A change of θ corresponds to a change of speed and is thus an indication of interaction. For the motion of a null (massless) object, $\tan\theta = 1$ [see Fig. 1].

We then apply causal symmetry to the hypercone, by assuming the hypercone is unchanged when shifted by an infinitesimal but discrete step, *dx*. This *extended causality* (EC) generates the finite nature of the object, introducing a second causal constraint,

$$(\Delta\tau + \mathrm{d}\tau)^2 + (\Delta x + \mathrm{d}x)^2 = 0. \tag{2.2}$$

By combining this with Eq. (2.1), this becomes

$$\Delta\tau + f \cdot \Delta x = 0, \tag{2.3}$$

where

$$f^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau^\mu}, \qquad f_\mu = -\frac{\partial\tau}{\partial x^\mu}. \tag{2.4}$$

Here the f represents a fibre in the spacetime. A line tangent to this forms the f-generator of the local hypercone. In applying extended causality to an object, one defines both its location and its immediate future [see Fig. 2].

III. The Discrete Scalar Field

To apply extended causality to a massless scalar field ϕ, it is convenient to introduce a 3 + 2 flat spacetime,

$$(x^\mu, x^5) \in \mathbb{R}^5, \quad (x^5)^2 + x^2 = 0, \quad \phi(x^\mu, x^5)\,. \tag{3.5}$$

It is clear that in setting $x^5 \equiv \tau$, one imposes local causality. The EC constraint then becomes

$$(\tau - \tau_o) + f_\mu (x - x_o)^\mu = 0. \tag{3.6}$$

Applying this to the scalar field $\phi(x, \tau)$ constrains it to the hypercone generator, and it is designated by

$$\phi_f(x, \tau) \equiv \phi(x, \tau)|_{\mathrm{EC}}\,. \tag{3.7}$$

Equation (3.6) also induces a direction to field derivatives, thus

$$\partial_\mu \phi_f = (\partial_\mu - f_\mu\, \partial_\tau)\, \phi_f \equiv \nabla_\mu \phi_f\,. \tag{3.8}$$

With Eqs. (3.7) and (3.8) it is possible to derive a discrete field equation, following the usual method [3]. The action of the field is given by

$$S_f = \int d^5 x \{ \frac{1}{2} \eta^{\mu\nu} \nabla_\mu \phi_f \nabla_\nu \phi_f - \chi \phi_f \rho(x, \tau) \}, \tag{3.9}$$

where $d^5 x = d^4 x d\tau$, χ is the coupling constant, and $\rho(x,\tau)$ is the (discrete) field source. The field equation thus becomes

$$\eta^{\mu\nu} \nabla_\mu \nabla_\nu \phi_f(x, \tau) = \rho(x, \tau), \tag{3.10}$$

with an energy tensor

$$T_f^{\mu\nu} = \nabla^\mu \phi_f \nabla^\nu \phi_f - \frac{1}{2} \eta^{\mu\nu} \nabla^\alpha \phi_f \nabla_\alpha \phi_f . \tag{3.11}$$

In analyzing the properties of this solution, it is useful to express the solution as a Green's function. Thus

$$\phi_f(x, \tau) = \int d^5 y \, G_f(x-y, \tau_x - \tau_y) \, \rho(y, \tau_y), \tag{3.12}$$

and

$$\eta^{\mu\nu} \nabla_\mu \nabla_\nu G(x, \tau) = \delta^{(5)}(x) . \tag{3.13}$$

The Green's function is then

$$G_f(x, \tau) = \frac{1}{2} \Theta(bf^4 \, t) \, \Theta(b\tau) \, \delta(\tau + f \cdot x), \tag{3.14}$$

where $b = \pm 1$ and Θ is the Heaviside step function. Here the $b = +1$ or $\tau > 0$ solution represents an emission at the retarded time, while the $b = -1$ or $\tau < 0$ solution is an absorption at the advanced time [see Fig. 3]. Unlike the standard Green's function, Eq. (3.14) contains no singularity, therefore the discrete field propogates without changing its amplitude.

Equation (3.14) has the additional property of being independent of any transverse components of x. Specifically,

$$f \cdot x_T = 0, \quad \frac{\partial}{\partial x_T} G_f = 0. \tag{3.15}$$

Since the transverse dimensions are not affected by the field and do not contribute to it, the general 3 + 1 field is reduced to a 1 + 1 manifold.

In applying extended causality, we have created a model in which there are three general types of objects:

1. discrete, finite, point sources,

2. discrete, finite, point fields,
3. discrete, finite, point interactions.

In this way, we have introduced a unifying symmetry between fundamental objects (fermions and bosons).

IV. A Simple Solution

As a simple example, consider the field source to be a single scalar charge,

$$\rho(x, t_x = t_z) = q(\tau_z)\, \delta^{(3)}(x - z(\tau_z))\, \delta(\tau_x - \tau_z), \tag{4.16}$$

where $q(\tau)$ is the scalar charge and $z(\tau)$ is its world line. The solution for the emitted field is then

$$\phi_f(x, \tau_x) = \chi \int d\tau_y\, \Theta(t_x - t_y)\, \Theta(\tau_x - \tau_y)\, \delta[\tau_x - \tau_y + f \cdot (x - y)]\, q(\tau_z), \tag{4.17}$$

which can be reduced to

$$\phi_f(x, \tau) = \chi q(\tau)\, \Theta(t)\, \Theta(\tau)|_f, \tag{4.18}$$

or simply

$$\phi_f(x, \tau) = \chi q(\tau)|_f, \tag{4.19}$$

where $\tau \geq 0$, and $t > 0$.

Since this is a discrete model the emission or absorption of a field gives a discrete change for q [see Fig. 4]. In general

$$q(t) = \begin{cases} q_{\tau_j} & \text{if} \quad \tau_j < \tau < \tau_{j+1}, \\ \dfrac{q_{\tau_{j-1}} + q_{\tau_j}}{2} & \text{if} \quad \tau \equiv \tau_j, \\ q_{\tau_{j-1}} & \text{if} \quad \tau_{j-1} < \tau < \tau_j. \end{cases} \tag{4.20}$$

Thus, with the finite difference as the derivative,

$$\dot{q}(t) = \begin{cases} \Delta \mathrm{q}_{\tau_j} & \text{if } \tau = \tau_j, \\ 0 & \text{if } \tau \neq \tau_j. \end{cases} \tag{4.21}$$

The directed derivative of this solution becomes

$$\nabla_\nu \phi_f = -f_\nu\, \dot{q}|_f, \tag{4.22}$$

from which one can obtain the energy-momentum tensor for the field,

$$T_f^{\mu\nu}(x, \tau_x) = f^\mu\, f^\nu\, \dot{q}^2|_f. \tag{4.23}$$

The 4-momentum of the field can then be defined as

$$p_f^{\mu} = T_f^{\mu\nu}\,\eta_\nu = f^{\mu}\,\dot{q}^2\big|_f, \tag{4.24}$$

where $f^{\mu}\,\eta_\mu = 1$. Since f is a null field, $f_\mu\,f^\mu = 0$ and

$$\nabla_\mu T_f^{\mu\nu} = -2\,f_\mu\,f^\mu\,f^\nu\,\dot{q}\,\ddot{q}\big|_f = 0, \tag{4.25}$$

thus the energy-momentum is trivially conserved. This also ensures the scalar charge is conserved. Thus, the discrete field is a finite, point-like spacetime deformation projected on the null generator, with a well defined and conserved energy-momentum and charge.

The nature of the discrete scalar field can be summarized as follows:

1. it is a null field,
2. it carries a scalar charge Δq,
3. it is, in fact, a charged abelian field,
4. its source is any object with energy,
5. all objects interact with it via the energy tensor.

Therefore, one may conclude that ϕ is *the gravitational field.* As such, it can be said to represent a "classical graviton."

It is possible, in fact, to derive a metric from such a discrete gravitational field [4]

$$g^{f}_{\mu\nu}(x) = \eta_{\mu\nu} - f_\mu\,f_\nu\,\chi\phi_f(x,\tau), \tag{4.26}$$

where Einstein's field equation is reduced to

$$f_\mu\,f_\nu\,\eta^{\alpha\beta}\,\nabla_\alpha\nabla_\beta\,\phi_f(x,\tau) = \chi T_{\mu\nu}\,. \tag{4.27}$$

V. Conclusions

Comparing the discrete and continuous general forms, a shift from continuum to discrete can be summarized as

$$\{x\} \Rightarrow \{x, x^5\}, \tag{5.28}$$

$$\phi(x) \Rightarrow \phi(x,\tau), \tag{5.29}$$

$$\partial_\mu \Rightarrow \nabla_\mu, \tag{5.30}$$

while the continuum can be obtained from the discrete via

$$\phi(x,\tau) = \frac{1}{2\pi}\int d^4 x\, \delta(f^2)\, \phi_f(x,\tau). \tag{5.31}$$

It is clear extended causality forms the basis of a discrete and finite gravitational model, which reduces to the Einstein model when averaged across the continuum. It has been argued [6] that given its fundamental nature, gravity cannot be formally quantized. Given the strong analogy to quantum models, it is therefore interesting to speculate if the EC model could be developed into a complete model of gravity. Clearly this an area for much further study.

REFERENCES

1. B. D. Koberlein, in *Aspects II: Proceedings of ANPA 20*, ed. K. G. Bowden (Cambridge University, 2000), p. 249.
2. J. Rosen, *Symmetry in Science: An Introduction to the General Theory*, (Springer, New York, 1995).
3. see, e.g., G. Arfken, *Mathematical Methods for Physicists*, (Academic, San Diego, 1985).
4. M. M. Souza and R. N. Silveira, "Discrete and Finite General Relativity," (gr-qc/9801040, 1998).
5. H. P. Noyes and S. Starson, "Discrete Anti-gravity," in *Proc. ANPA WEST 7*, (Stanford University, 1991).
6. see, e.g., J. A. Wheeler, *Geometrodynamics*, (Academic Press, New York, 1962).

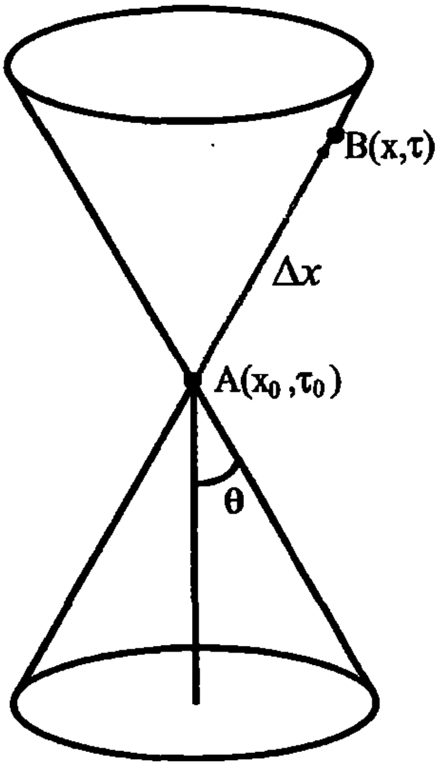


Figure 1: The Hypercone

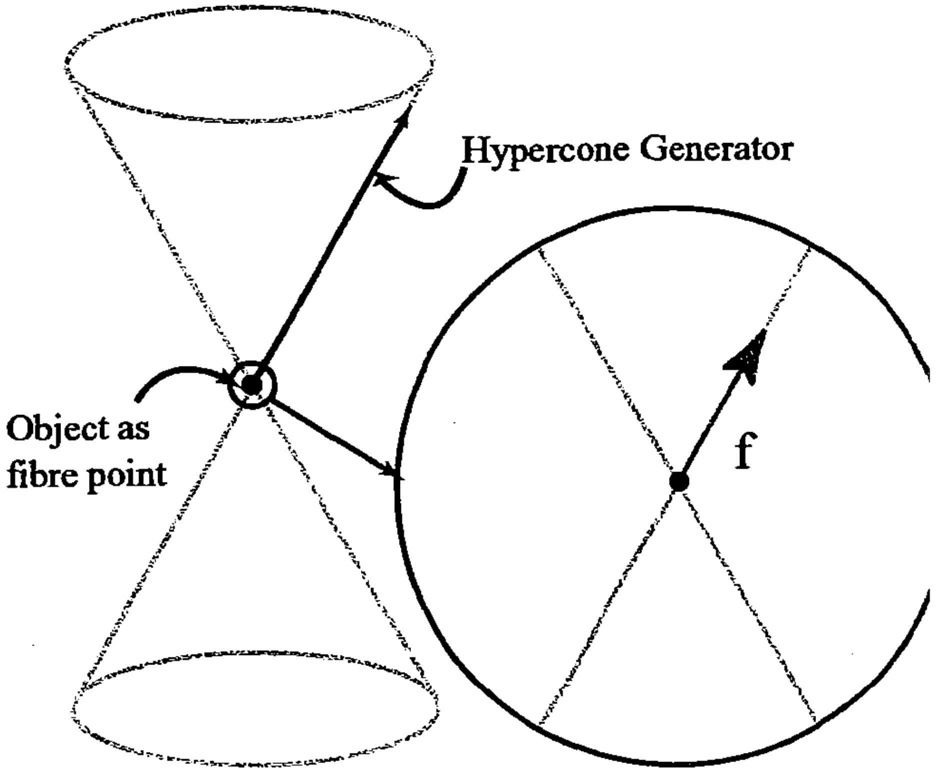


Figure 2: The f-generator

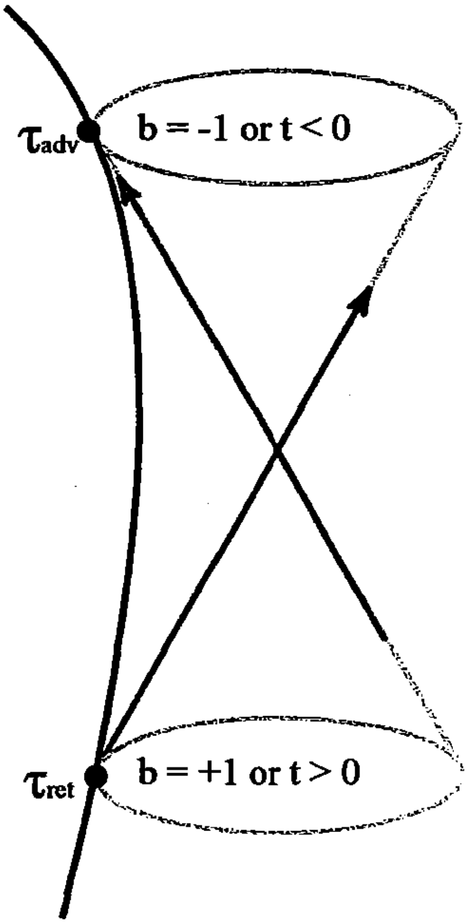


Figure 3: The Particle Solutions

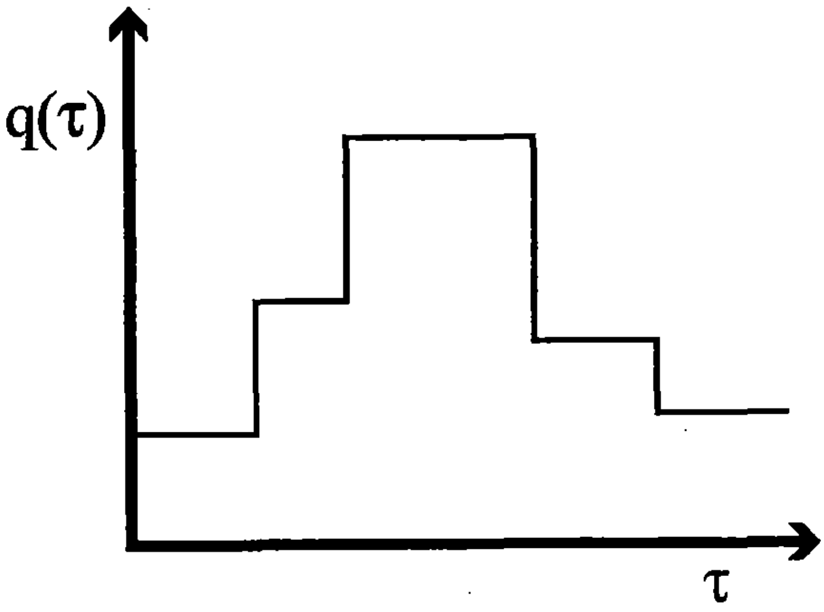


Figure 4: The Discrete Charge Function

PROSPERO AND CALIBAN

CLIVE KILMISTER

Red Tiles Cottage
High Street
Barcombe
Lewes
BN8 5DH

1. INTRODUCTION

Two remarks first about the title: (i) the meaning will be revealed towards the end of the talk, so if you want to know, you will have to wait till then. (ii) As to titles, let me refer you to the rhyme about that great music critic of the Sunday Times, Earnest Newman. Referring to an actual occurrence the rhyme says: Mr Earnest Newman

Said:'Next week it will be Schumann'
But when next week came
It was Wagner just the same.

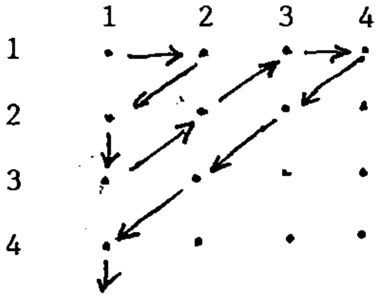
So, unsurprisingly, I will be considering the Combinatorial Hierarchy.

Frederick Parker-Rhodes gave a construction which leads to some significant numbers and in particular to 137. I argue (and here and later I am echoing Ted Bastin but I had better not assume his agreement with absolutely everything I say) that the only way to understand this construction is to see it as arising from a process construction. Just what this means has been questioned. Well, what Arleta did in her paper [these Proceedings] exemplified a process theory. Strings arose: it is true that she cut them down to size pretty soon by fancy equivalence relations but there was a continual generation process. Another question has been what connexion such an algebraic approach can have to physics. My general idea is that it provides a framework from which space, time and the rest of physics can be constructed.

Entities come into consideration in the process and are labelled in order that they may enter the theoretical description ("the mathematics"). All that can arise about two entities is whether they are the same or not. The process must check this - I call it discrimination - because if two are the same, they must be given the same label. The process nature means that:

I: The labelling has to be carried out in an iterative way.

We are comparing two elements and the result is a single one, so we have a binary operation. We need a systematic way of doing this iteratively. I have used one particular way: I use the ordinals, 1, 2, 3, ... as labels for entities and then I make the comparisons along this recursively-defined path:



I don't know if another path could be devised which gave significantly different results; I don't think so.

2. CONWAY'S CONSTRUCTION

We took over a construction given by John Conway [1976] though modified in an important way. Essentially it depends on the rule:

C: pq = the least r different from any p'q, pq', p or q

Here p' is any label earlier than p and so for q'. Frederick used addition for this binary operation which was very convenient but for reasons which will become clear I will write it as juxtaposition (and think of it as multiplication). As a matter of history, Ted and I had already used juxtaposition in discovering something of this in our Concept of Order papers

E W Bastin & C W Kilmister [1954, 1955] but we had not been able to make Frederick Parker-Rhodes' next step. Anyway, in my notation the system must have the discrimination property:

D1. $pp = p^2 = qr \rightarrow q = r$.

D1 is not quite enough for my purpose. For by forming qr I can deduce $q = r$ only if qr belongs to the set of squares. I might recognise a p^2 equal to it but in general I need to check in turn whether qr is any of the p^2 - by a further discrimination! An infinite regress beckons; so as well as D1 I need:

D2. There are signals, which are not entities, and p^2 is a signal for any p.

The idea of a signal is that the process has an inbuilt recognition of it without the need of further discrimination. At this stage there is no loss of generality in taking just a single signal and I call it z (historically because Frederick's signal was zero).

If with Frederick one starts from 2 generators, a, b say, the system [z, a, b, ab] would be the quadratic group S so long as C is augmented by the rules

R1. $z^2 = z$, $zp = pz = p$.

These cannot arise from C alone because it defines discrimination only between entities not between entities and signals or between two signals. I say that the system is weakly associative and weakly commutative because adding R1 exhibits it as embedded in a larger system, including z, which is commutative and associative. Is associativity a useful bonus or an evil temptation? I am not sure but, as I was discussing with Peter Rowlands, it is very difficult to see any physical reason for associativity to appear.

3. THE NEXT STEP

Frederick Parker-Rhodes' next step splits into two. Firstly, he generates from entities already labelled "discriminately closed subsets" : that is, sets of entities with the property that the result of discriminating any two different entities lies in the set. And he represents such a dcss, U, by an automorphism A_U defined by

PR. $A_U(p) = p \Leftrightarrow p$ in U.

Secondly, and this point has been less appreciated, he argued that these automorphisms were "entities just like the original ones" but 3 in number instead of 2. It is worth teasing out, in order to carry out Frederick's subsequent calculations, just what sort of similarity there has to be between the new entities and the old. One thing that is needed is for the new entities to satisfy D (ie. be a discrimination system). The way Frederick actually argued was this: the original entities were bit-strings of length 2 or members of a vector space over $\mathbb{Z}_2$ of dimension 2. The automorphisms were then 2 x 2 square matrices over $\mathbb{Z}_2$ and these were members of a vector space over $\mathbb{Z}_2$ of dimension 4. Frederick rearranged the matrix elements to emphasise this; this caused some raised eyebrows in the early days. When the process view came along we could see that the essential feature here was that a discrimination operation was induced in the standard way by the rule

Induced product. $AB(p) = A(p)B(p)$ for any p.

If the new entities did not satisfy D then the dcss at the next level would include signals as well as entities and, moreover, it would raise further obstacles to the already difficult task of understanding just what is going on in Frederick's construction.

It is not necessary for Frederick's argument that such sets of 3 things should be able to arise for physical reasons. But it is comforting to note how they do. One way is in the emergence of physical time from the process (considered atemporally). As I said, the process is one of labelling entities. The first term of the sequence is simply the labelling of an entity, say with the label a. The second term labels an entity with the label b. The third term, c, is a discriminator; that is, it determines whether the entities are the same or not. If they are, they must have the same label. In this case, if the next term, d, of the sequence is the replacement of b by a wherever it occurs, I say that a is earlier than b. If it is the other way round, then b is earlier than a. If a, b are not the same, no temporal statement can be made at this stage. This introduces a temporal ordering but only locally. Another entity is labelled, e. There are now two ways of proceeding. If the next terms of the sequence are the comparison successively with a, b then a further time-ordering is established and this time-ordering will be transitive as one would expect. But there is an alternative: the next term is a compound comparison which determines whether e is in the set [a, b, c] or not. This involves coalescing a set of labels into a single entity and this alternative corresponds to spatial rather than temporal arrangement.

In any case, if the new entities are like the originals, the process can be repeated. Because 2 entities produce 3 dcss and r produce $2^r - 1$, one gets the Combinatorial Hierarchy:

CH. $3 \rightarrow 7 \rightarrow 127 \rightarrow 10^{39}$ **OR** $G_{12}^{(1)} \simeq G_{13}$, $G_{12}^{(2)} \simeq G_{13}^{(1)} \simeq G_{17}$, ...

I will explain the notation in more detail below. Frederick's result was that only 4 stages were possible. The numbers are suggestive but the Combinatorial Hierarchy is not just numbers. We have a hierarchy of groups. I am using here a notation I got from Arleta some years ago. She wrote G_{SD} for a group arising from, as she would have said, S similarities and D differences but I would say S signals and D generators. (When S is bigger than 1 it is best to note a lack of independence in the signals and to define S appropriately. I come to that later on.) Then she writes $G_{SD}^{(1)}$ for the structure going up a level and this gives the groups tabulated above. Strictly I ought not to use these groups because the actual hierarchy structures are only discriminately closed. They leave out the signal z. I denote the actual structures by underlining: $\underline{G}_{SD}$, so $|\underline{G}_{SD}| = 2^D - 1$.

And this process analysis of Frederick's construction was found to improve the predicted value of $1/\alpha$ from 137 to 137·0351.. (latest observed value 137·036..). Making my first reference back to the title of this talk, this happy situtation was like that in Milan before the Duke was driven out by his usurping brother.

4. ASPECT.

Everything changed a few years ago by the discovery of what I called aspect. Frederick - and my process construction mimicked this by means of C - had taken a weakly commutative and weakly associative discrimination operation. But this is evidently wrong, because pq means that q is compared with an already labelled p. So if $pq = z$, q is to be replaced by p wherever it occurs, whereas qp represents something different. So C needs modifying, but this is not too hard. John Fawn gave the clue in his paper this year where he distinguished between the objective aspect of a comparison and the human part. I would not use those words but just say that an alternative simpler process, in which the difference between pq and qp is ignored, would also be possible. To make that explicit:

H. The system with D generators has a homomorphism, H, down to G_{1D}.

This homomorphism then defines an equivalence relation, $\sim$, in the usual way:

$$p \sim q \leftrightarrow Hp = Hq \leftrightarrow (Hp)(Hq) = z.$$

For example, $rs \sim sr$. This equivalence relation has to be determined in the same way as before, so at least two signals are now needed:

$$pq = qp = z \leftrightarrow p = q, \quad pq = qp = y \leftrightarrow p \sim q \text{ but } p \neq q.$$

The modified Conway construction is then:

C*. pq = least r different from $p, q, p'q, pq', qp$, subject to H.

As there are now two signals, the lowest level will be G_{22}. This turns out to be weakly associative and the extra rules to produce an associative system in which it is embedded are:

R2. $z^2 = y^2 = y, \quad zy = yz = z, \quad yp = py = p.$

With these extra rules adjoined G_{22} is embedded in the quaternion group Q as one can see because the table becomes a semi-direct product:

$$C_2 \rtimes S \simeq Q$$

I owe to Arleta my introduction to the semi-direct product. $G \simeq H \rtimes K$ means that H is normal in G and that every element g in G is a product g = hk with h in H and k in K. Actually in this case S happens to be the quotient group Q/C_2. At this point we have survived the tempest and are upon the island, and here the subtlety and the magic start.

5. ON THE ISLAND.

Let me start with the magic but it is rather the subtlety to which I shall come later that interests me. The semi-direct product $Q \simeq C_2 \rtimes S$ is just mathematician's obscurantism for saying that Q is got by taking S and putting in plusses and minusses, because a realisation of C_2 is [1, -1] under multiplication.So the process pairs off elements; but it does it in a particular way. This causes a notion of "anti-ness" (in the physicist's sense) to emerge. It is here I make contact with Rowlands and Cullherne. But like all magic, this has its limitations. -1 is just one representation of the non-identity element of C_2. We'll see later that more C_2's arise and they cannot all be [1, -1]. As far as the algebra is concerned, I think Rowlands and Cullherne would go along with that. But it is a bit more tricky if you want to use the -1 to give you, say, e and -e for charge. The trouble is that bits of physics as sophisticated as charge are a long way down the line. I must get there in the end but I don't like to think how many years it will be.

The hierarchy construction can now be carried out but the first subtle question is whether Frederick's (ii) still holds. That is, is $G^{(1)}_{22} \simeq G_{23}$? To find out I introduce the idea of a property P being hereditary:

PH. $P[G^{(r)}_{SD}] \rightarrow P[G^{(r+1)}_{SD}]$.

It is easy to see that associativity and commutativity are hereditary and so is H. Now begin with G_{23}, that is, just sit down, use C* and work out the first closed system when there are three generators. It turns out to have 16 members and to be (strongly) non-associative. There is a pairing of elements here just as there was for Q, so you could say that you had 8, ±. But it is not octonions (Cayley numbers) as one might guess, the reason being that any subset generated by any two elements is, for octonions, quaternions. Here, by contrast, of the seven such sets, three are weakly associative and so can be quaternions but four are strongly non-associative though "quaternion-like". (In fact, it is what you get from quaternions by attaching a $\sqrt{-1}$ to two generators, as far as products of different elements are concerned but not changing the diagonal elements in the way that would imply). The notation G_{23} was meant to suggest "group" and so here becomes inappropriate. I write instead L_{23} (L for loop). Evidently L_{23} cannot be isomorphic to $G_{22}^{(1)}$ because $G_{22}^{(1)}$ is associative (since associativity is hereditary). Yet L_{23} turns up as the uniquely determined (by C*) discrimination system, so what do we make of $G_{22}^{(1)}$?

6.PROSPERO & CALIBAN.

Before I get on to that let me stand back a bit and see where we are. Frederick's original construction, our original process version and now the aspect complication with only two generators all provide an island of associativity - a haven from the tempest of non-associative systems which inhabit the world outside. I get now to the title. Caliban inhabited the island - indeed should have inherited it from his mother but Prospero usurped him. Caliban knew nothing of the world outside, supposed the island was the whole world. Prospero learnt about the island as well as the world he had lost. I am going to investigate the topography of the island with the idea of supporting an outrageous claim:- that the orthodox physics community is a set of Calibans, their attention fixed on the associative systems which the mathematicians have given them and oblivious to the bigger picture. The fact that G_{22} is a Clifford algebra is a first straw in the wind; they and I agree on that.

Now what about $G_{22}^{(1)}$? Again, as with the original construction, there are three dcss, two with one element and the remaining one with six (not three, as before, because of the doubling). Of course we have to change the definition of a dcs appropriately. Instead of the qualification "any two different entities" we need "any two non-equivalent entities", since otherwise H would be infringed. It is easy to work out the commutation relations and the squares of the three generating automorphisms corresponding to these three dcss. It then turns out that there is a whole set of signals, eight in number, which form a group (as they must) $C_2 \times C_2 \times C_2$ because the eight come from three generating signals. There is thus an "eight-fold way" of being equivalent. Using Arleta's notation, this would be G_{43}. (That is, she counts in S the number of generating symbols + the identity, so as to cover the previous two cases.)

To set $G_{22}^{(1)}$ in some sort of context I consider other things on the island: that is, systems artificially constrained to be associative, satisfying D1, D2, H and also I to the extent that the associative condition allows. If the signals form $(C_2)^r$, I call this a $G_{(r+1)D}$. The strategy for looking at G_{SD}'s is to start with $S = S^*$, the largest possible S, which is easily seen to be $S^* = \frac{1}{2}(D^2 + D + 2)$. I just quote results:

$$\begin{tikzcd}
G_{42} \simeq (C_4 \times C_2) \rtimes C_4 \arrow[d] \\
G_{32} \simeq C_4 \rtimes C_4 \arrow[d] \\
G_{22} \simeq C_2 \rtimes C_4 \simeq Q
\end{tikzcd}
\qquad
\begin{tikzcd}
& G_{73} \simeq (C_4 \times C_2 \times C_2) \rtimes G_{42} \arrow[d] & \\
& G_{63} \simeq (C_4 \times C_2 \times C_2) \rtimes G_{32} \arrow[dl] \arrow[dr] & \\
G_{53}^{(\alpha)} \simeq (C_4 \times C_2 \times C_2) \rtimes Q & & G_{53}^{(\beta)} \simeq (C_2 \times C_2 \times C_2 \times C_2) \rtimes (C_2 \times C_2 \times C_2)
\end{tikzcd}$$

Notice that there are two G_{53}'s, a situation which turns up frequently when the number of generators gets larger.

But when you look for another homomorphism to go down further, there aren't any! So no G_{43} can exist (nor any G_{S3} if S is less than 4); there must be a lower as well as an upper limit on the number of signals. So what now of $G_{22}^{(1)}$? The sad answer is: it does not exist. A closer look at the structure soon shows why: it fails to fulfil D. At this point Caliban is reduced to a state ofpuzzlement, because if $G_{22}^{(1)}$ is not a discrimination system, what are the implications for understanding Frederick's construction on the island? But Prospero can continue with his supercilious attitude: he knows that L_{43} was the real starting point, not G_{43}. Not that that seems to put him into any much better position. For he has now to understand the Parker-Rhodes construction without the clause (ii). The first question to arise is: why does $G_{22}^{(1)}$ not satisfy D? This is not too difficult; the trouble arises far back over the rule I and the way in which it is used. The trouble is this: at the second level the rule for Induced Product constructs the discrimination as a triad of discriminations between elements at level 1. These elements include both the original three and their discriminators and these discriminators - you will find - will include level 1 signals as well as genuine non-signal elements. Now 2 elements are the same if they discriminate to z but two signals are the same if they discriminate to y. This is what causes the trouble. Actually Prospero should have been alerted to this earlier, once he saw the eight signals coming up in $G_{22}^{(1)}$. For of these eight signals (the direct product of three groups of order 2) there is a unique special one (made by drawing the identity from each group) and there is also one special equivalence (the identity). So the hint to Prospero was that at level 1 z should be chosen as the identity signal and y as the other member of C_2. And so, although G_{22} is weakly associative, it should not be made associative by the rule R2 but rather made non-associative by the rule

R3. $z^2 = y^2 = z,\ zy = yz = y,\ \ zp = pz = p.$

Then G_{22} is not Q. Indeed as it is non-associative it should according to my earlier rule be re-named L_{22}. It is in fact what you get by taking the table for quaternions and putting 1's down the diagonal. Now it again makes sense to ask whether $L_{22}^{(1)} \simeq L_{43}$. Or rather it would if one knew how to generate L_{43} but it is not clear why, using I, one should bring in so many signals.

A better question is: is there a homomorphism $L_{22}^{(1)} \rightarrow L_{23}$? There is not, but does it matter? It is only a matter of jettisoning I for the higher levels. And, of course, with z as the identity element, D becomes hereditary as well.

What then of Prospero? His magic has restored his Dukedom to him, but at some cost. He may well regret leaving the security of the Island of Associativity.

7. ADDENDUM

There was a degree of shock amongst some of my audience at my enthusiastic embrace of non-associativity. To them I would just say: much of this was already adumbrated in my joint paper with Ted Bastin last year. It should not be taken as a sign that constructing physics will now become impossible. Rather, at the second level with aspect, a clear choice has to be made:- either to require associativity and so make continuing contact with orthodox physics or to insist on discrimination systems at each level and so construct something with the same homomorph (the Combinatorial Hierarchy) and free of confusions. Then the important work will be to establish the relations between the two systems so as to get the best of both worlds.

REFERENCES

Conway, J.H. 1976 "On numbers and Games" LMS publications.

Bastin E.W.& Kilmister C.W. 1954 "The Concept of Order I. The Space-time structure". Proc.Camb.Phil.Soc.50, 278.

Bastin E.W.& Kilmister C.W. 1955 "The Concept of Order II. Measurements." Proc.Camb.Phil.Soc.51, 454.

Bastin Ted & Kilmister C.W. 2000 "The Participant Observer" in "Participations", Proceedings of ANPA 21, pp 7 - 21.

WHY I FIND ROWLANDS AND CULLERNE EXCITING

TED BASTIN
Maesllwyn, Tan y Groes, Cardigan SA43 2JF
Tel. 01239 810 404 ted@tor.co.uk

1. INTRODUCTION

In this note I rely entirely on work conducted with Clive Kilmister but do not assume his agreement with everything.

I shall concentrate upon one aspect of the theory of Rowlands and Cullerne:- their putting some algebraic, or, as I would prefer to call it, *combinatorial* structure into the basis of high energy theory. In carrying through this program new definitions are provided for some of the fundamental physical concepts, although these authors usually pass lightly from these new constructions to the corresponding conventional ones, and back again. I find it a great advantage to emphasize that this cannot be performed lightly. Lack of definiteness at this point seems to me fundamentally to vitiate the whole of quantum physics. A bit of history may make my point.

Recently Pierre invited the original authors of PITCH to comment on his assessment of that paper and it consequences. That reminded me that I made several extended trips as Pierre's guest at SLAC. Part of the agenda was for Pierre to teach me the principles of high energy physics and his approach to it. I learned a certain amount of course, but the further we went the more I found it impossible to see the whole subject as a consistent series of arguments. It wasn't just that there were obscurities and gaps in the understanding: you would expect that. There seemed to be a built-in confusion at work. As near as I could pin this down, it was due to switches from classical realism to the combinatorial principles appropriate to the high energy situation without any principles to guide which set of principles one was to follow or when it was legitimate to switch. To cope with the new ideas one would expect the classical concepts to have been replaced by combinatorially based ideas with careful prescription of the steps to take to get to the continuum thinking. In particular the notions of mass, charge and spin were spoken of with, *faut de mieux*, the assumption that their definition in classical mechanics would always be legitimate and applicable. Of course this criticism can be made of the foundations of quantum theory itself. However there the theology which forces the two irreconcilables into a marriage which never was, and whose most colourful sobriquet is the 'collapse of the wave-function', is done in one step. In high energy work it is disseminated throughout. Moreover, discretizing the mechanics gets you nowhere if you still rely on the intuition of unrevised classical concepts.

One centre of my lack of comprehension was the use of momentum space with such things as Feynman diagrams. I was assured that use of configuration space led to disaster, which I appreciated, but then I couldn't follow the argument because I couldn't find how they introduced the new kind of momentum space which would replace the appeal to configuration space. I found later, that however useful the techniques (analytic continuation and the rest) for establishing mechanics may be in quantum field theory they actually were in effect practically never used in the study of particle processes.

All the algebraic structures in use for particles come from the symmetry groups whose interpretation comes from spatial rotation and special relativity. There is usually a paucity of usable algebra for this reason. People are inhibited from exploring what mathematics might be appropriate to the new situations. One or two cases where the old interpretation is forced into action with ludicrous results are hackneyed and to speak of them embarrassingly naive. Parity non-conservation and time reversal are algebraic possibilities which are used a lot, but they make no sense with the spatio-temporal interpretation of the symmetry groups. If we reverse the time locally can we leave it nice and comfortable for us outside the locality. Is there a critical radius where this uneasy boundary is crossed? People may say I ought to have more sense than to talk like that, but, please, how are we to talk?

If I am right about this enormous conceptual mismatch then the consequences are far-reaching. The theorists actually seem to revel in the mystery of it all instead of confessing that it is a disaster at least as awful as the ultra-violet catastrophe. If they would only say that this group symmetry investigation is purely formal and must not be confused with the classical meaning which is separate, and admit that it has so far no space-time meaning, we could manage. Of course they don't. That each piece of argument relates importantly to experimental fact I do not dispute, but none of the relationships that are assumed in default can be taken at face value

I think what I have to say fits in with R&C's central placing of algebras. [Of the papers by Rowlands and Cullerne I shall use "The Dirac Algebra and SU(5) Symmetry" for reference.] Their work seems to offer the people who have worked on the combinatorial hierarchy the hope of giving them a bridge to the standard model and so to bring the work on the coupling constants into current thinking. This paper consists of a series of comments on the relationship between the two approaches but at the time of ANPA 2000 these remained a bit disconnected. My longer aim is to use the mathematical apparatus of the combinatorial hierarchy to provide a minimal list of the basic concepts that are needed for R&C's work without depending on interpretation in continuum physics. It would be going too far to say that this attempt would start with uninterpreted mathematical symbolism. Most importantly, both sides have built into their arguments the interpretation of the three basic interaction strengths with their corresponding 'charges' s, e and w.

2. SOME R&C PRINCIPLES

R&C have an outer product of two quaternion groups one of which is referred to as a 'vector', which suggests some spatial significance. The workings of a physical system are represented by 0 and 1 symbols which prescribe the presence or absence of one of a set of properties which the system either has or does not have in that place at a given time or stage of development. The changes in these 0/1 patterns are transformations of the quaternion groups. Though R&C do not emphasize this, we suppose that we are to imagine a flux of activity represented by these transformations which goes on all the time and is subject to some statistical control, so that exchange forces of a generalized kind are at the basis of the theory.

The 'properties' just mentioned are three 'charges', labelled s, e, and w respectively, one of which, e, relates to the electromagnetic interaction, with the other two, s and w having an analogous relation to the strong and weak interactions respectively.

Some time ago, R&C suggested that it would be a good idea to compare their and our uses of 0, 1. For R&C, 0 and 1 are 'existence symbols'; that is to say they assert the existence or non-existence of something. However they do not assert the existence or non-existence of objects such as particles. They assert the existence of properties -in the first place *charges* in their sense of that word. It is convenient to call such a property which something either has or does not have, a *descriptor*. It would be natural to speak of particles or perhaps quarks as the carriers of a set of such descriptors. Our position is not quite the same. We need the 'something there /something not there' of the 0, 1, but we can only begin physical interpretation from a population of strings of elements. To give them individuality so as to be descriptors requires further steps.

I was once involved with classification problems such as the specification of diseases by descriptors in order to have a method for distinguishing real diseases from arbitrary associations of symptoms. The symptoms were the descriptors, and the patient was assumed either to have a given one or not to have it, and it was assumed that the medical practice would decide this unambiguously. Statistical concepts were invented to describe the degree of reality of the *clumpings* of data. (A 'clump' was a set of descriptors which corresponded, in a way to be specified, with a real disease. The use of quantised properties of particles is a bit like the use of descriptors in important ways. The logical situation is simpler when the attributes can be descriptors (0/1). R&C's preference for the Han Nambu integral as opposed to fractional charges is clearly partly due to the need for this simplicity.

Language forces us to talk of a 'particle' for the properties to be 'of', but there is a big jump when we identify these with the particles of the experimentalist. Particles defined as groupings of properties are more like the diseases of which I spoke. (I also remind you that we have yet to say what we mean by these groupings: the idea is simple enough, but finding the mathematics to nail it down, given our situation, is a major task). To the usual theorist a charge can equally well be integral or fractional, because he has always in mind that he is talking of a real particle which is automatically independent of our description of it. It will probably be difficult to convince him that there is an important difference. The obvious interpretation of the groupings of descriptors are 'quarks'.

The primary descriptors are s, e, w. For R&C these are each associated with signs +/- which I do not consider in this paper. R&C write transformations of any groupings in a 3 x 3 array with the charges downwards in the table and the quaternion components i, j, k in the body of the table. Presumably the table corresponds to the interactions in the factors of the product of the quaternion algebras. What are the three columns labelled along the top of the array? R&C identify them with **B, G, R** -the colours used in classifying the quarks. We note that the existence of three of these is a necessity of the algebra. The colours are, however, not descriptors in the same sense as the charges since they appear in classifying the charges. If we require that all the things we can do by associating descriptors be allowed, then we get a group with unit element. There are only two groups of order 4 - the quaternion group and the quadratic group. R&C use the former; we the latter. In the former case transformations take place between coefficients called i, j, k, and one is compelled to introduce a minus sign, for when they operate on themselves. This minus sign complicates the system of descriptors since +/- are in a sense independent descriptors. R&C anticipate the physical application by saying we that with them we have introduced anti-ness or anti-particleness, but I have not established a place for this identification in these notes. Whether there is a case for introducing quaternions as a package deal, or whether we should separate their different aspects (particularly justifying the +/- separately) is a question to be addressed in a further discussion, but what we should not do, arguing as we are now, is simply to say that the two signs are an experimental requirement. Similarly, we must be wary of assuming that if we use quaternions there must also be an automatic physical interpretation of the unit element, There may be one, but such identification would have to be justified by looking at the way we use transformations.

Meanwhile we seem to have lost sight of our 0/1. In fact they are still there because each transformation of descriptor groupings requires that there be a place for a missing descriptor. It is not sufficient that we merely list the descriptors which are present. We are using ordered *strings* with 0 or 1 in each place. R&C are happy with this notation though I believe it involves the continuing 'particle' which I have queried. In a central argument of R&C's in "The Dirac Algebra and SU(5) Symmetry". importance is attached to the condition where one colour has only one unit value. I suggest that the reason must be that if there are more than one unit then transformations between them can happen. If not, not. For diverse reasons it is very important to have this restriction. From a conventional point of view it enables R&C to introduce the concept of a particle by the device of saying that in spite of our logical prohibition (above) on particles we may revert to that language because the particles in question are 'contained'. The argument goes further. R&C say (P.17) "at least one of the charge components will take on unit values in two colours". This must be because one does need to use the system to describe transformations, and there must be a place for them and therefore two unit charges somewhere. I shall assume this to be the cornerstone of their argument because it gives me a starting point. It seems very skeletal, but I don't think it would be right if it were not. Anyway I think this mathematical argument is a central point from which one can start and build the rest. I also surmise that it is original and not to be found in current theory. They proceed from this point to delineate the e and w interactions, but I do not go further for the moment.

3. CORRESPONDING HIERARCHY PRINCIPLES

In the hope that my skeleton agenda for R is recognizable. I now put our alternative (hinted at above). Instead of going from a simple set of operators or descriptors by making a sort of outer product of it with another of the same order, the hierarchy picture required that transformations of the original set should constitute new entities or new units in their own right capable of sustaining operations analogous to the original ones. This produced a new *level*. We imagined a hierarchy of levels of entities.

At that time we already thought in terms of a statistical background of operations going on all the time and supposed that we were studying those which had particularly simple invariant properties which gave rise to physical effects. Vectors had to be diversified by the position of 0's and 1's but we did not postulate things whose existence or non-existence was specified by them individually. My talking earlier about the difficulty of having particles independently of descriptors comes to the surface here. To embrace the logical complexity of having both, we have to have two adjacent levels and talk about them in combination. It is as though each provides an environment for the other. Without such an environment it makes no sense to speak of the elements of a vector as individually defining the presence or absence of a property because if we change the occupation of one position in the vector it becomes a different vector with no relationship to the first. Unless, that is, we can have the reference frame provided by the other level.

Unfortunately we have not investigated the algebra of this kind of relationship between levels. My intuition is that when we have, we shall be able to make serious contact with schemes R&C's complex representations -A,B,C,D,E,. Until that time we cannot claim to be able in detail to use the bridge into the high energy picture and the standard model offered by R&C.

We had to make some sense of what Parker-Rhodes called 'eigenvectors', which were used in his construction, and came slowly to the notion of *discriminate closure*. In the matrices we had normal multiplication of the elements of the strings, and we had addition mod two or symmetric difference to combine the strings elementwise. (There are no other possibilities for the latter if the hierarchy is to be non-trivial). We called this operation *discrimination* because it gave 0 or 1 answers according as the strings which were being brought together were or were not identical. As we said earlier we were talking about a statistical universe in which operations were going on all the time, and with that idea, a certain concept seemed very natural to form our basic physical things. This was *closure* under the discrimination operation. To be precise, the zero string was omitted at each level and therefore discriminate closure was defined as "$a + b$ is in the set, so long as $a \neq b$". The groupings of strings which were to form a new level in the hierarchy were *discriminately closed subsets*. These would have a stability to raise them above the unspecifiable play of the processes.

There followed a long and slow process of basing everything on the discrimination and closure idea and the unsatisfactory features of Parker-Rhodes' picture have now been resolved in a deductive development which is mostly due to Kilmister. This development includes a second iteration correction to the integral (137) value for the electromagnetic coupling reciprocal. K's value is accurate enough (7 or 8 significant figures) to put it into the region where the definition of coupling itself becomes wobbly. Of course anyone confronted with this calculation would assume that in a couple of hours he would be able to find where the numbers were fiddled in, but in reality the deduction proceeds using only the principles already in place for the original calculation.

We draw attention to the way the 'charges' or different interactions emerge from the structure of levels. Relatedly, we note that at the level of the 4-strings (s-charge) there is no apparatus defined to make transformations because these have not yet been constructed. They are at the next level. This result appears in the work of R&C as the 'containment' of the quarks.

For the moment my discussion enables me to give a reasonably neat placing of the bridge between us and R&C. It seems to me quite a strong position to say that we have alternative and hopefully complementary ways of proceeding from the point where you import independent vectors with what I call your descriptors, and where we construct a new level. We can say to the conventional people that if they want a realist particle picture then they can go the descriptor way. If on the other hand they want to understand the origin of the colour descriptors and the s, e, w charges and their strange interaction strengths then they can go the level route.

4. SOME FURTHER PRINCIPLES USED BY R&C WHICH REQUIRE FURTHER THOUGHT TO CONNECT THEM WITH THE HIERARCHY WORK

1. Mass -the principle of charge counting

This principle is very basic. It really explains how quantization appears, and is based on charge counting. In it, mass and charge appear together. This theory is quite remote from conventional ideas on the subject which R&C use elsewhere. This is probably why it appears tucked away in section 29 'Particle mass spectrum'. It all really follows by continuation of the 0/1 principle which is satisfactory. They consider the number of zero units of charge, and this is part of the same argument as that which gets their 'accommodation' of the different charges. I attempt to give some idea of the argument from my own slant with the following rather disconnected comments. R&C introduce the number n_0. This is the number of zero charges under certain specified circumstances. I presume these are the charges which are constrained to be zero by the 'accommodation'. From our point of view we shall expect them to be along the lines of prohibitions arising from higher levels not being in place. R&C speak of 'energy of confinement' which they writes as m_e/α and associate with the prohibition of $0 \Leftrightarrow 1$ relating to electric charge and strong charge and permitting strong charge. This already brings in our identification of the scaling relation between the interaction strengths since for us the containment is due to prohibition or admission of level changes.

2. The symmetry principle and charge accommodation (Charge exclusion theory)

3. The origin of anticharge as an alternative way of achieving symmetry (P.21) and the labels +/- for antiness

4. Baryons and mesons derived from above concept (P.21). Can we get their masses from this possibly using CH?. Also top of P.29 on origin of the meson.

5. Quarks, introduced by the argument (P.27) of interchange of the s charge, assuming that it is necessary to prevent the identification of quark colour.

6. Leptons including the electron. The theory is surprising in making the very existence of the electron dependent on the electroweak interaction mechanism (P.31, bottom) arguments from symmetry. These are very subtle, and require the introduction of three new parameters called red, blue and green to accor7. Obviously there have to be restrictions on the freedom of transition of the charges in order for each to exhibit its proper characteristic behaviour. This need is handled using d with the standard model. In the first place there is no puzzle about the choice of three parameters rather than some other number. In R&C's scheme this is entailed in the requirement that they form a table to connect the charges with the 0/1 principle, and it is the same three appearing again. If this coincides with the empirical requirements of the standard model then that is a success for the theory.

8. The + and - and antiness. Clearly R&C use +/- to specify the potentially complete freedom for any descriptor (spin as well as the charges) to be antied. In the classification schemes the three "charges" all appear fully, and are associated with +/- for all the world as though +/- were descriptors with equal rights to participate. I have hitherto been content to point out that if we follow the 0/1 principle then we have to provide for the +/- by interpreting the - as the removal of a 1, but did not know how this would be done. I have been assuming that +/- was just something added on to the algebraic scheme, and in fact R&C claim to be doing better because the +/- is provided by their classificatory scheme, P. 14 (middle). "The charges are not interconvertible; that is, they are subject to separate conservation laws because they are rotation

asymmetric. Charge units are fixed, so the only possible values for an individual charge component at any point are 0 and 1. However the charge units are imaginary, and so have equal status. Thus a unit charge may be defined as some combination +/-is, +/-je, +/-kw... ." This may be the point at which +/- should be introduced, but for me it requires a basis for the whole idea of antiness. The fact that R&C are not explicit about this may be because of the minus sign's already having appeared in the non-commutativity of the quaternions, and I think this needs more explanation.

Of course the two theories diverge over s, e and w. The important fact is the central prominence of these "charges" in both theories, and the difference is something to be used to good effect. In the hierarchy the association of the charges s, e, and w with the numbers 3, 10, 137, 10^{38} result from the successive coming into play of more algebraic complexity. The first two numbers have in the past been thought to correspond to the varieties of strong interaction, and 137 of course to e. However the last number corresponds with gravitation -not with the weak interaction, and this fact poses a challenge. We have a way of incorporating the weak interaction which uses the fact that though the mapping of the newly created elements at the third level is impossible because there are too many of them to map onto the product space of $256^2 = 65536$ elements, we can imagine some restriction which limits the choices and modifies the probabilities accordingly. This way of thinking give the right value for the weak interaction strength. (See Combinatorial Physics, P.93)

An Investigation of the Higgs Mechanism

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 7ZE, 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 jpc@wincoll.ac.uk

The origin of the Higgs vacuum is sought in the simultaneous requirement of four solutions for both the Dirac equation and the charge accommodation algebra. The fixing of the value of the electroweak mixing parameter, $\sin^2\theta_W$, at 0.25, is related to the existence of four generators for the $SU(2) \times U(1)$ group structure. Possible origins are suggested for the mass values of some fundamental particles, based on the numbers of zero charges involved.

INTRODUCTION

The Higgs mechanism is the most widely used formalism for the generation of masses for both the fermions and the electroweak gauge bosons within the Standard Model. The entire formalism, however, is derived from an intrinsically simple physical source. This is the assumption that the weak interaction operates from a filled, rather than an empty, vacuum. The vacuum is filled in terms of energy (not charge) states, but it is ultimately responsible for the violation of charge conjugation symmetry in the weak interaction. However, although the Higgs mechanism explains the creation of massive particle states through the existence of a massive spin 0 gauge boson, which is coupled to the various fermion states, it does not predict the exact strengths of the couplings which produce the required fermion masses.

Our aim here is to discover first the origin of the filled Higgs vacuum in terms of the two formalisms which we have previously introduced to explain the Standard Model, namely, the Dirac nilpotent wavefunction and the charge accommodation mechanism [Rowlands and Cullerne, 1999, 2000a, b, c]. We will then look briefly at the relation between the Higgs mechanism and the $SU(2) \times U(1)$ electroweak splitting, and, finally, put forward a number of ideas relating to the origins of specific particle masses.

A number of ideas from our previous work will be essential to this discussion. One is our version of the Dirac equation:

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

with nilpotent wavefunction:

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

and four solutions, involving ± E and ± $\mathbf{p}$. Another is the charge accommodation rules, leading to the five tables A-E, shown, in reduced form, below:

A

		B	**G**	**R**
u	$+e$	$1j$	$1j$	$0i$
	$+s$	$1i$	$0k$	$0j$
	$+w$	$1k$	$0i$	$0k$
d	$-e$	$0j$	$0k$	$1j$
	$+s$	$1i$	$0i$	$0k$
	$+w$	$1k$	$0j$	$0i$

B

		B	**G**	**R**
u	$+e$	$1j$	$1j$	$0k$
	$+s$	$0i$	$0k$	$1i$
	$+w$	$1k$	$0i$	$0j$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0i$	$1i$
	$+w$	$1k$	$0j$	$0k$

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	$+e$	$1j$	$1j$	$0j$
	$+s$	$0k$	$0i$	$1i$
	$+w$	$0i$	$0k$	$1k$
d	$-e$	$0i$	$0k$	$1j$
	$+s$	$0j$	$0i$	$1i$
	$+w$	$0k$	$0j$	$1k$

Here, of course, E remains invalid for the full representation, except (it is assumed) at Grand Unification. These tables are the ultimate product of taking the parameter charge to be composed of three individually conserved units, s, e, w, represented by quaternion operators $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$, which are not individually conserved, but are rotation symmetric.

To represent, in algebraic terms, the extra degree of freedom which this rotation symmetry implies, the three quaternion charge operators have been assumed to be operated on by random unit vectors, $\boldsymbol{j}\mathbf{r}_1$, $\boldsymbol{i}\mathbf{r}_2$, $\boldsymbol{k}\mathbf{r}_3$, on the assumption that the extra degrees of freedom are unobservable, because the unit charges only exist in the combinations allowed by the standard model. We accomplish this combination algebraically (in the case of baryons) by taking the scalar products of $(-\boldsymbol{j}\mathbf{r}_1 + \boldsymbol{i}\mathbf{r}_2 + \boldsymbol{k}\mathbf{r}_3)$ with the respective unit vector components, **i**, **j** and **k**. The three expressions

$$(-\boldsymbol{j}\mathbf{r}_1 + \boldsymbol{i}\mathbf{r}_2 + \boldsymbol{k}\mathbf{r}_3).\mathbf{i}$$
$$(-\boldsymbol{j}\mathbf{r}_1 + \boldsymbol{i}\mathbf{r}_2 + \boldsymbol{k}\mathbf{r}_3).\mathbf{j}$$
$$(-\boldsymbol{j}\mathbf{r}_1 + \boldsymbol{i}\mathbf{r}_2 + \boldsymbol{k}\mathbf{r}_3).\mathbf{k}$$

then become the charge allocations for the **R**, **B** and **G** quarks in a ddd baryon, the total (before normalization) being given by $(-\boldsymbol{j}\mathbf{r}_1 + \boldsymbol{i}\mathbf{r}_2 + \boldsymbol{k}\mathbf{r}_3).\mathbf{1}$. Irrespective of the values of $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$ used, there are only five independent solutions for these scalar products, and these are the ones represented in tables A-E.

From this representation, it can be shown that, in the case of the ddd combination (to take the simplest case), the resultant product is of the form

$$(-\boldsymbol{j}\mathbf{r}_1 + \boldsymbol{i}1\mathbf{r}_2 + \boldsymbol{k}\mathbf{r}_3) = (\boldsymbol{i}\mathbf{i}\mathbf{j}\boldsymbol{j} + \mathbf{i}\mathbf{i}\boldsymbol{i}\mathbf{r}_2 - \boldsymbol{i}\mathbf{i}\mathbf{k}\boldsymbol{k}) = -\boldsymbol{i}\mathbf{i}\,(-\mathbf{j}\boldsymbol{j} + \boldsymbol{i}\mathbf{i}\boldsymbol{i}\mathbf{r}_2 + \mathbf{k}\boldsymbol{k}),$$

and the terms $-\mathbf{j}\boldsymbol{j}$, $\boldsymbol{i}\mathbf{i}\boldsymbol{i}$, $\mathbf{k}\boldsymbol{k}$ form a closed cycle of the same form as the quaternion operators $\boldsymbol{k}$, $\boldsymbol{ii}$, $\boldsymbol{ij}$ in the Dirac algebra, although, this time, as the products of vectors and quaternions, they are commutative.

A final result is that our model, using intrinsically integral, rather than fractional charges, requires a weak mixing angle in the GSW $SU(2) \otimes U(1)$ theory, such that

$$\sin^2\theta_W = \frac{\Sigma\, t_3^{\,2}}{\Sigma\, Q^2} = 0.25\,.$$

This is, notably, the same value whether we use quarks or leptons, a fact which is at variance with the predictions of other theories, such as minimal $SU(5)$. Our value of $\sin^2\theta_W$ also leads to exact Grand Unification at the Planck mass between all three nongravitational interactions, with zero electroweak 'hybridization' ($\sin^2\theta_W = 1$) at the Grand Unification energy scale.

THE HIGGS VACUUM

In deriving the origin of the Higgs filled vacuum, we shall find it convenient to refer to our ($SU(5)$) analogy between the components of the Dirac equation E-$\mathbf{p}$-m and the charge structures of w-s-e, as derived from the tables A-E (with s taking the role of the three-dimensional or 'vector' term).

Taking one of the standard forms of the Dirac equation

$$(\alpha.\mathbf{p} + \beta m - E)\ \psi = 0$$

we may represent the left hand side in the form:

$$(\alpha.\mathbf{p} + \beta m - E)\ \psi = \begin{pmatrix} -E & 0 & im & -ip \\ 0 & -E & ip & im \\ -im & -ip & -E & 0 \\ ip & -im & 0 & -E \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = 0 ,$$

where the column vector is the usual 4-component spinor, and the terms E and p are the quantum operators which give the eigenvalues represented by these symbols when applied to the exponential term of the wavefunction.

We have previously shown [Rowlands and Cullerne, 2000a] that the use of 4 × 4 matrices is, in fact, the fundamentally quaternionic nature of the Dirac wavefunction, and that the ultimate origin of this representation is the symmetrical use of a 4-vector space-time in the relativistic energy conservation equation.

The 'vector addition' for conserved charge can be considered as a product of a 4 × 4 matrix and a 4-component column vector in the same way as conserved energy in the Dirac equation:

$$\begin{pmatrix} kw & 0 & -je & -is \\ 0 & kw & -is & je \\ -je & is & -kw & 0 \\ is & je & 0 & -kw \end{pmatrix} \begin{pmatrix} kw + is + je \\ kw + is - je \\ -kw - is + je \\ -kw - is - je \end{pmatrix} \qquad (1)$$

This 4 × 4 matrix is almost identical in form to the matrix for the Dirac differential operator, although the + and – signs are in different places. The s term may be considered to take up the vector-type properties of **p**, and could be represented as a vector with a single well-defined direction. The sign applied to e is that of the charge itself, but e has the added property of isospin, so that the e's on the first and fourth rows of the matrix and on the first and fourth rows of the column vector can be considered as isospin 'up' and the others as isospin 'down'. The opposite states of isospin are not + and – but 1 and 0. So, we should apply to these e terms the matrices:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix};\begin{pmatrix} 0 \\ 1 \end{pmatrix};\begin{pmatrix} -1 \\ 0 \end{pmatrix};\begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

The result of this is that all terms involving e term disappear on multiplication.

The matrix can be written as the sum of three matrices:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} kw$$

$$\begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix} is$$

$$\begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} je\,,$$

and these can be written as the sum:

$$\gamma_0 kw + i\gamma_1\, is + i\gamma_3\gamma_0 je\,.$$

In vector-quaternion notation, $\gamma_0 = i\mathbf{k}$; $\gamma_1 = i\mathbf{i}$; and $\gamma_3 = i\mathbf{k}$. However, since the vector labels are arbitrary (except for their cyclic sense), we could also write γ_1 and γ_3 as $i\mathbf{k}$ and $i\mathbf{i}$. Taking this latter alternative, and multiplying from the left by $-i\mathbf{k}$, we obtain $k\mathbf{k}$, $-ii\mathbf{i}$, and $j\mathbf{j}$ for the respective coefficients of kw, is, and je. This is virtually identical to the algebra of charge accommodation already derived. The extra imaginary sign applied to s in both cases results from the extra vector property arising from the nature of s itself.

When the matrices are multiplied by the column vector, the resulting column vector is a unit column vector times a scalar factor. If we apply the factor i to the s term in the column vector, we derive $w^2 - s^2$, which becomes 0 when $w = s = \pm 1$. (Effectively, the linking of w and s makes one term in s, e, w 'redundant' information in the same way as E becomes 'redundant' information in E-$\mathbf{p}$-m and the table E in the representations A-E.)

Both the charge and 'mass' parameters (w-s-e and E-$\mathbf{p}$-m) require mapping onto a quaternion or 4-vector 4-space. This is manifested in the 4 × 4 matrices required for the Dirac equation and for the specification of the charge states. The matrices require that there must be exactly four 'solutions' in each case (in the accompanying column matrix). The 'mass' and charge parameters, however, present us with alternative problems, for E-$\mathbf{p}$-m offers too few solutions while w-s-e requires too many. In the case of *E*-**p**-*m* we have only the + and – states of **p** (which, as a vector, automatically provides these alternatives), while *E* and *m* are both, strictly speaking, confined to positive values. In the case of the charges, there are eight possible combinations of ±*w*, ±*s*, ±*e*. There is no overall discrepancy, however, since 2 × 8 = 4 × 4, and so, to overcome the problem, each effectively 'borrows' aspects of the other – *E*-**p**-*m* uses charge; *w*-*s*-*e* uses mass. (Alternatively, we may consider it using the fourth (suppressed) term in the matrix representation – angular momentum or mass-energy; expanding *E* restricts angular momentum; contracting *w* expands *m*.)

The Dirac equation introduces the unphysical –*E* to create the extra alternatives for *E*-**p**-*m*. Conventionally, this is taken to be equivalent to producing antiparticles, that is, particles with opposite signs of charge, but the equation itself does not incorporate this. To explain how –*E* was possible physically, Dirac had to assume a filled vacuum for antiparticles in the ground state, and therefore a natural preponderance of matter over antimatter. In effect, – *E* only occurred when a 'hole' was created in the vacuum by the simultaneous production of an electron (or, later, some other antifermion) at an energy above the ground state. The positive nature of physical energy was thus preserved. No other physical explanation has ever been put forward to justify the formalism now universally adopted, and it is in fact the discrepancy between the physical status of *E* and –*E* that is actually responsible for the categorization of half the states as 'antimatter', rather than merely as states with opposite signs of charge.

When we come to *w*-s-*e*, it is apparent that each charge must have equally possible + and – states, producing eight solutions. As in the case of *E*-**p**-*m*, we must have some states which can be categorized as antiparticles of the others. So we decide to make this determined by the sign of *s*. (The charge accommodation rules then make s effectively a vector quantity of the same kind as **p**, with only one 'direction' well-defined.) This then leaves one other charge to adopt + and – signs within matter, producing isospin, as the equivalent of the spin variation produced by the two signs of **p**. Isospin is thus a requirement of a quaternionic mapping. If this property is assigned to *e*, then we have a problem over *w*. We can only resolve this by supposing that some process violates charge conjugation by making the effective signs of *w* for matter and antimatter linked to those of *s*. The mechanism, which is already available from the Dirac theory, is that of the filled vacuum. Dirac had a filled vacuum from which antielectrons could be produced; we can now identify the component involved as the *weak* charge, and the filled vacuum as specifically a *weak* vacuum. So, it is *w*, in effect, that determines the status of matter and antimatter, rather than *s*, though the sign of *s* is linked with the effective sign of *w*. (See Figure 1.)

Figure 1. The origin of the Higgs vacuum.

Exactly four Dirac solutions required:

E	**p**	m
+	+	+
– *	–	
*only with filled fermion vacuum	spin up spin down	

→ antifermions

→ ground state has fermions only

Exactly four charge accommodation solutions required:

w	s	e
+ *	+	+
– *	–	–
+* fermions	fermions	isospin up
–*antifermions	antifermions	isospin down
*effective because of filled weak (fermion) vacuum		

Just as the problem with E-**p**-m was solved by invoking charge, so the problem with w-s-e is solved by invoking mass. Violation of charge conjugation symmetry requires violation also either of parity or time reversal. In effect, for a particle with only weak charge, this would mean a single state of helicity. However, for a particle with any other kind of charge as well as weak, the charge conjugation violation would not be absolute and the alternative state of helicity would be allowed. A finite probability of alternative helicity requires a nonzero rest mass (because the speed must be $< c$), and the amount of mass will be determined by the probability of the state and the strength of the interaction involved (*e*, *s*). Since it is the filled weak vacuum, with nonzero expectation value, or 'Higgs field', that gives rise to the nonzero rest mass of the fermions involved, then the mass of a particle must be determined by the strength of its coupling to this field; and the strength of the coupling will depend ultimately on the degree of symmetry-breaking which the creation of that particle requires. (See Figure 2.)

Figure 2. The Higgs mechanism

Filled weak vacuum

→ Violation of weak charge conjugation symmetry (+ P / T)

→ Fermion with only weak charge in a single state of helicity

→ For fermion with other charge, violation not absolute

→ Finite probability of alternative helicity

→ Nonzero rest mass (speed < *c*)

→ Amount of mass depends on probability of state (e.g. number of zeros)

and strength of the interaction involved (*e*, *s*)

Filled weak vacuum (with nonzero expectation value) = 'Higgs field'

→ nonzero rest mass of fermions involved

→ mass of fermion determined by the strength of coupling to this field

→ strength of the coupling depends on degree of symmetry-breaking
in creation of fermion (e.g. number of zeros)

It is evident from this discussion that the mapping between *w*-*s*-*e* and *E*-**p**-*m* (or charge states onto wavefunctions) derives from the respective 4 × 4 matrices, and that the sign differences derive from the different physical requirements for producing the desired four solutions. In particular, the *E*-*w* connection is required by the filled vacuum; **p**-*s* is required by their shared vector property; also, in the case of leptons, mass occurs only where *e* is present to break the asymmetry (in this case) between the two possible states of helicity.

We can also, finally, use this mapping to produce a more logically-structured version of the charge-accommodation mechanism. Here, we start with the Dirac equation, in which we have to introduce a filled fermion vacuum to create the two-sign degree of freedom required for E. We also now define a particular status for antiparticles beyond the original requirement that each charge-type has two possible signs. When we set up the charge accommodation mechanism, we assume that a particular type of charge, say s, can only be unit in one of the three 'colours' needed to make up an observed state. This excludes charges of the opposite sign, so we take the concept of antistates from the Dirac equation, and so assign $-s$ to the antiparticles. We cannot, however, repeat the same procedure for, say, e, which must have both signs in both states and antistates. So, we preserve the rule that a charge ($-e$ in this case) can be unit in only one of the three 'colours', but make the 'default' position (e, e, e) as opposed to (0, 0, 0) for s, and so produce two signs by creating 'weak isospin', with alternatives (e, e, 0) and (0, 0, $-e$). (Subsequently, we find that using 'weak isospin' actually gives us a suitable zero for the matrix equation for charge, which we find is equivalent to the Dirac equation for E-$\mathbf{p}$-m.) Finally, to accommodate two signs of w, we have to refer to the fact that a filled vacuum, with antiparticles nonexistent in the ground state, violates charge conjugation symmetry for the charge (w) which specifies the fermion state.

THE PSEUDO-DIRAC EQUATION

It is convenient to write (1) in the form:

$$\begin{pmatrix} kw & 0 & -ije\uparrow & -iis \\ 0 & kw & -iis & ije\downarrow \\ -ije\downarrow & iis & -kw & 0 \\ iis & ije\uparrow & 0 & -kw \end{pmatrix} \begin{pmatrix} kw + iis + ije\uparrow \\ kw + iis - ije\downarrow \\ -kw - iis + ije\downarrow \\ -kw - iis - ije\uparrow \end{pmatrix} = 0 \qquad (2)$$

Here ↑ represents isospin up and ↓ isospin down.

If we now create an exponential term $e^{-i(wt - \mathbf{s.r})}$, to produce a state vector for charge, and define $i\partial/\partial t = -iw$ and $-i\nabla = is$, we obtain

$$\begin{pmatrix} ik\partial/\partial t & 0 & -ije\uparrow & -i\nabla \\ 0 & ik\partial/\partial t & -i\nabla & ije\downarrow \\ -ije\downarrow & i\nabla & -ik\partial/\partial t & 0 \\ i\nabla & ije\uparrow & 0 & -ik\partial/\partial t \end{pmatrix} \begin{pmatrix} kw + iis + ije\uparrow \\ kw + iis - ije\downarrow \\ -kw - iis + ije\downarrow \\ -kw - iis - ije\uparrow \end{pmatrix} e^{-i(wt - \mathbf{s.r})} = 0 . \qquad (3)$$

The weak isospin terms cancel, suggesting why this becomes the phase term. We can therefore write equation (2) as:

$$\begin{pmatrix} kw & 0 & 0 & -iis \\ 0 & kw & -iis & 0 \\ 0 & iis & -kw & 0 \\ iis & 0 & 0 & -kw \end{pmatrix} \begin{pmatrix} kw + iis \\ kw + iis \\ -kw - iis \\ -kw - iis \end{pmatrix} = 0,$$

or

$$\begin{pmatrix} kw & 0 & 0 & -iis \\ 0 & kw & -iis & 0 \\ 0 & iis & -kw & 0 \\ iis & 0 & 0 & -kw \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} (kw + iis) = 0 .$$

The left-hand side reduces to

$$\begin{pmatrix} kw + iis \\ kw + iis \\ kw + iis \\ kw + iis \end{pmatrix} (kw + iis) ,$$

in which each row of the column vector becomes

$$- w^2 + s^2 = 0 ,$$

as in the parallel case of the Dirac equation.

Without the 'phase' terms, equation (3) becomes:

$$\begin{pmatrix} ik\partial/\partial t & 0 & 0 & -i\nabla \\ 0 & ik\partial/\partial t & -i\nabla & 0 \\ 0 & i\nabla & -ik\partial/\partial t & 0 \\ i\nabla & 0 & 0 & -ik\partial/\partial t \end{pmatrix} \begin{pmatrix} kw + iis \\ kw + iis \\ -kw - iis \\ -kw - iis \end{pmatrix} e^{-i(wt - \mathbf{s}.\mathbf{r})} = 0 .$$

Here, each term of the resultant column vector becomes a pseudo-Dirac equation:

$$(ik\partial/\partial t + i\nabla)\,(kw + iis)\, e^{-i(wt - \mathbf{s}.\mathbf{r})} = 0 ,$$

in the same way as each term of the resultant column matrix becomes a Dirac equation for the *E*-**p**-*m* combination. This equation provides a convenient representation of the parallel between the mathematics for charge accommodation and that for the Dirac state.

THE ELECTROWEAK MIXING PARAMETER

The electroweak $SU(2) \otimes U(1)$ mixing mechanism requires the existence of four generators, W^-,W^+, Z^0, γ, only one of which corresponds to a purely massless state. It is of interest to consider whether, with the association of E with w, and m with e, the figure of $\sin^2\theta_W$ of 0.25 has some additional fundamental significance, in the association of symmetry-breaking with the creation of massive states. The same proportion is possibly observed in the neutral electroweak transitions which we have defined in our previous work [Rowlands and Cullerne, 2000b], the photon having one neutral isospin transition ($0 \rightarrow 0$) and three hypercharge ($-j \rightarrow k$, $-j \rightarrow 0$, $0 \rightarrow k$), while the Z^0 has the same proportion in reverse.

The number of unit (±) and zero states of electromagnetic charge are equal throughout the full range of quarks and leptons, and this is because of weak isospin. Also there is the same ratio of states having or not having weak isospin, because they are left-handed or right-handed. When we take $\Sigma\, t_3^2 \,/\, \Sigma\, Q^2$, therefore, we find $(½)^2 \,/\, 1^2$ because the quantum number for t_3 is ½ ($SU(2)$) and that for Q is 1 ($U(1)$). However, if we now look just at the lepton states, the same ratio will apply for the 4 lepton-antilepton combinations, which are the basis of W, W, Z, γ, because each lepton has a 1 in 2 chance of having an electric charge component, and so only one of the combinations $e\bar{e}$, $e\bar{\nu}$, $\nu\bar{e}$, $\nu\bar{\nu}$, has no electric charge, i.e. no right-handed component of ν or left-handed component of $\bar{\nu}$, to mix with and produce mass by the Higgs mechanism.

MASSES AND OTHER FUNDAMENTAL CONSTANTS

The Higgs mechanism was introduced originally to account for the spectrum of particle masses. We have previously shown [Rowlands and Cullerne, 2000a] that the breaking of symmetry which produces mass is largely a result of the production of zero states of charge, mass and charge acting in this sense as an invariant in the same manner as space and time. Zero charges represent complete coupling to the Higgs field; nonzero charges represent a reduction of the vacuum state to less vacuum. The complete coupling to a zero charge may be equivalent in energy to one unit of m_e / α.

We have already suggested ways of deriving the masses of the lighter baryons and mesons by counting up the zero charges in the multiplets [Rowlands and Cullerne, 2000a]. Particularly successful was the prediction of the mass of sss (Ω-), at the apex of the baryon decuplet, from the 24 zero charges at $24\, m_e / \alpha = 1.68$ GeV. Similar principles applied to an extended multiplet including the fourth quark (c), would lead to a count of 50 zero charges for ccc, and a mass of $50\, m_e / \alpha = 3.5$ GeV, in line with the assumed mass for the c quark of between 1.0 and 1.6 GeV.

We have also proposed, though much more tentatively [Rowlands and Cullerne, 2000a], that the Higgs boson might represent the zeroing of charge in the complete range of fermion-antifermion combinations within the A-D representations, with a total of 182 GeV (or 227 GeV for A-E). (6 flavours, 6 anti-flavours, 3 colours, 3 charge types for each quark / antiquark, and 2 for each quark-antiquark pairing, over 4 representations, gives a total of 2592 zeros, on assumption that each zero represents a mass energy of m_ec^2/α, gives an approximate total of 182 GeV.) Similar procedures may apply to the electroweak bosons, which require only the calculation of the mass of Z^0. The masses of particles are determined by the strength of their coupling to the Higgs field. Z^0 is completely coupled, γ does not couple. Complete coupling implies full strength of vacuum, i.e. zero charges. If Z^0, γ derived from the reduced D, E representations (no *s*), or A/B/C – D/E, then complete summation of the zeros over 2 representations produces 91 GeV.

The other masses that we would hope to derive are those of the most important fermions, namely the masses of the six quarks (u, d, c, s, t, b) and at least three of the leptons (e, μ, τ). The quarks present a problem, as their masses 'run' like the values of the coupling constants with the energy of interaction. The lepton masses, however, are believed to be constant, and the electron mass, m_e, may perhaps be derivable from a more fundamental derivation of m_e / α or m_e from Grand Unification.

Essentially, the three fundamental constants *G*, *h* and *c* have no intrinsic meaning. There must be numbers which relate the arbitrary units which we choose to assign to space, time and mass, but, since these parameters are fundamental, it is meaningless to look for fundamental significance in the units themselves. Only the numerical values attached to structures, such as the electron, have this kind of intrinsic meaning. Now, if M_X is the Planck mass, then it, too, becomes a fundamental unit, since it is composed entirely from *G*, *h* and *c*. The value of $\sin^2\theta$ is also known from an exact conceptual argument, In setting up the conditions for Grand Unification, then, we have four equations with just five unknowns at any particular energy (μ), namely μ, α, α_2, α_3, and α_G [Rowlands and Cullerne, 2000b, c]. Of course, these equations, as we write them, are merely first-order approximations, but we could, *in principle*, refine them to any degree of exactness. In effect, given any assumed μ, we could have exact predictions for any of the other four constants, with no other empirical input. To go further, it is quite possible that one of the other constants has a theoretically exact value at some particular specified value of μ. The most likely possibility is that $\alpha_3 = 1$ when μ is, say, m_e / α (the energy equivalent for a unit strong charge). At present, agreement is moderately good but not perfect, as $\alpha_3 = 1$ seems to occur for $\mu = 1.5\ m_e / \alpha$ (the muon mass), but the equations here are very sensitive to the approximations employed, and the true value may really be closer to the one we expect. If this is so, then *not even μ need be assumed* to derive the four fine structure constants; we have, rather, a fifth fundamental equation to derive m_e / α or m_e itself.

We have no definite means yet of extending this kind of calculation to other fundamental fermion masses, though Grand Unified theories, in general [Kounnas, 1984], predict that the masses of (d, s, b) become equal with those of (e, μ, τ) at Grand Unification, and that, at the energies within the range of measurement, the masses of (d, s, b) will be approximately three times the Grand Unification value. It is of interest to note that the total mass of the twelve known fermions is approximately 182 GeV, which is the same as the total mass of bosonic states from a vacuum composed of the four representations A-D.

According to the standard treatment of the Higgs mechanism (Aitchison and Hey, 1989), fermion masses *m* are generated by couplings g_f to the Higgs field of the form

$$g_f = \frac{e}{\sqrt{2}\sin\theta_w}\frac{m}{M_w} .$$

Let us suppose that the sum of all twelve fermion masses is generated by the bosonic equivalent of the four quark-lepton representations A-E, while the electroweak transitions are generated by the bosonic equivalent of just two. In numerical terms, this would make the total fermion mass (Σm) equivalent to 182 GeV (which is possibly also the same as the Higgs mass, M_H), while the electroweak boson mass scale (M_Z) would be of order 91 GeV. Since $M_W = M_Z \cos\theta_w$, $\sin\theta_w$ at $M_Z = 0.5$, $\Sigma m = M_H = 2\,M_Z$, and the weak coupling constant, $g = e / \sin\theta_w$, the total coupling to all the fermion states,

$$\Sigma g_f = \frac{g}{\sqrt{2}}\frac{2}{\cos\theta_w} = \sqrt{\frac{8}{3}}g .$$

The total coupling, which produces the fermion masses, is thus *directly determined* by the weak coupling, and the fermion masses are related to M_H in the ratio of the Higgs coupling to the weak coupling:

$$\frac{m}{M_H} = \frac{g_f}{g}\sqrt{\frac{3}{8}} .$$

REFERENCES

Aitchison, I. J. R. and Hey. A. J. G., *Gauge Theories in Particle Physics*, second edition (Adam Hilger, 1989).

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.

Rowlands, P. and Cullerne, J. P. [1999], 'A derivation of particle structures and the Dirac equation from fundamental symmetries', in *Aspects II*, ed. K. G. Bowden, ANPA, 155-91.

Rowlands, P. and Cullerne, J. P. [2000a], 'The Dirac algebra and charge accommodation', in *Participations*, ed. K. G. Bowden, ANPA, 89-111.

Rowlands, P. and Cullerne, J. P. [2000b], '*SU*(5) and Grand Unification', in *Participations*, ed. K. G. Bowden, ANPA, 124-146.

Rowlands, P. and Cullerne, J. P. [2000c], 'The connection between the Han-Nambu quark theory, the Dirac equation and fundamental symmetries', *Nuclear Physics A* (in press).

On Dirac Symmetry and the Rowlands-Cullerne Formalism

Brian D. Koberlein

Department of Physics, SUNY Empire State College,
1475 Winton Road N., Rochester, NY 14609

November 2000

The work of Rowlands and Cullerne is discussed. It is demonstrated that the PRJPC equation for a fermion is mathematically equivalent to the standard Dirac formulation.

I. Introduction

Perhaps no ANPA presentation has sparked more debate in recent years than the work of Peter Rowlands and J. P. Cullerne (PRJPC). Given the potential breakthrough the model could provide particle physics, such debate is quite justified. However much of the discussion seems to focus on the validity of the PRJPC form of the Dirac equation rather than the application of this form to the standard model. This is unfortunate because the PRJPC equation is simply the Dirac equation expressed in an algebraic, rather than a matrix, form. Although this form differs significantly from the common expression, its validity is clear.

In an effort to clarify this equivalence, a derivation of the PRJPC equation is presented which differs from earlier works. In their model, Rowlands and Cullerne derive the equation from various symmetry arguments. This derivation is central to their application of the equation to the standard model. However, it is also possible to take the Dirac equation as a given, and transform it to the PRJPC form. This is the derivation presented here. In Section II, the Dirac equation in its standard form is presented. In Section III, this form will be converted to an equivalent algebraic form. Finally, Section IV expresses the algebraic form in PRJPC notation.

II. The Standard Dirac Equation

Following Hiley, let us assume the standard form of the Dirac equation as [1]

$$\left(i\gamma^\mu \frac{\partial}{\partial x^\mu} - m\right)\psi = 0, \tag{2.1}$$

where γ^μ are the usual Dirac matrices

$$\gamma^0 = \gamma_0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ \gamma^i = -\gamma_i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}. \tag{2.2}$$

Here the $\sigma^i$are the 2×2 Pauli spin matrices and ψ is a four-component matrix vector. The Dirac matrices are generally expressed in this form so as to relate them to Pauli spin. In general, however, there is an infinity of such matrix forms.

Equation (2.1) is a Dirac equation so long as the γ satisfy the anti-commutation relation [2]

$$\{\gamma_\mu, \gamma_\nu\} \equiv \gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = 2I, \tag{2.3}$$

where $\gamma_\mu = (\gamma_0, \gamma_i)$ and I is the identity matrix. By applying Eq. (2.3) directly, it is easily seen

$$\left(\mathrm{i}\gamma^\mu \frac{\partial}{\partial x^\mu} - m\right)\left(\mathrm{i}\gamma^\mu \frac{\partial}{\partial x^\mu} - m\right)\psi = \left(\frac{\partial^2}{\partial t^2} - \nabla^2 + m^2\right)\psi = 0, \tag{2.4}$$

which reduces to $E^2 = p^2 + m^2$ via the usual quantum operator substitution. It is clear then that any matrix set satisfying Eq. (2.3) may be used to satisfy Eq. (2.1). While the use of alternative forms may complicate certain physical interpretations, such complications do not affect the validity of the general equation and may always be resolved by mapping the γs to the form above.

This generality can also be used to express Dirac's equation in a more symmetrical form. For this, we introduce $\gamma_5 \equiv \mathrm{i}\gamma^0 \gamma^1 \gamma^2 \gamma^3$, which also satisfies Eq. (2.3). The (γ^μ, γ^5) terms actually form a mutually anti-commuting set, known as a *pentad*. By multiplying Eq. (2.1) on the left by γ^5, one obtains

$$\left(\mathrm{i}\gamma^5 \gamma^\mu \frac{\partial}{\partial x^\mu} - \gamma^5 m\right)\psi = 0. \tag{2.5}$$

By introducing the notation $(\epsilon^\mu, \epsilon^5) \equiv (\mathrm{i}\gamma^5 \gamma^\mu, -\gamma^5)$, Eq. (3.7) is simplified to

$$\left(\epsilon^\mu \frac{\partial}{\partial x^\mu} + \epsilon^5 m\right)\psi = 0, \tag{2.6}$$

where

$$\{\epsilon_\mu, \epsilon_\nu\} \equiv \epsilon_\mu \epsilon_\nu + \epsilon_\nu \epsilon_\mu = 2I, \tag{2.7}$$

Equation (2.7) is identical to the original constraint (2.3), therefore ϵ and γ are isomorphic. To emphasize this, we express Eq. (2.6) in γ form as

$$\left(\gamma^\mu \frac{\partial}{\partial x^\mu} + \gamma^5 m\right)\psi = 0. \tag{2.8}$$

Equation (2.8) is often referred to as the pentad form of the Dirac equation.

In general any set of γ_μ satisfying Eq. (2.3) generates a set of 16 linearly independent terms [2]

$$\begin{aligned}&\gamma_\mu, \quad \gamma_5 \equiv \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3, \\ &\gamma_5\gamma_\mu, \quad \sigma_{\mu\nu} \equiv \tfrac{i}{2}[\gamma_\mu, \gamma_\nu],\end{aligned} \tag{2.9}$$

and the identity matrix, I. Taken together, these matrices form an orthogonal basis for all complex 4×4 matrices. Any 4×4 matrix A may be expressed as a linear sum of the γ matrices

$$A \equiv \sum_{i=0}^{15} a_i \gamma^i \tag{2.10}$$

where a_i are complex coefficients. This property of the Dirac algebra is central to the concept of the algebraic Dirac equation.

III. The Algebraic Dirac Equation

For a particle in free space, the Dirac wavefunction is reduced to the form

$$\psi = \phi e^{i(p\cdot r - Et)}, \tag{3.11}$$

where ϕ is a 4-component complex coefficient. Substituting Eq. (3.11) into Eq. (2.8) yields

$$(\mathrm{i}\gamma^\mu p_\mu + \gamma^5 m)\phi = 0. \tag{3.12}$$

The left side of this equation is simply a γ-form of a 4×4 matrix. To express ϕ in a γ-form, we introduce a normalized wavefunction χ, such that the outer product

$$\Psi \equiv \phi\chi^* \equiv \phi \otimes \chi^* \tag{3.13}$$

is a 4×4 matrix. Multiplying Eq. (3.12) by χ^* yields

$$(\mathrm{i}\gamma^\mu p_\mu + \gamma^5 m)\Psi = 0, \tag{3.14}$$

which is easily solved by noting

$$(\mathrm{i}\gamma^\mu p_\mu + \gamma^5 m)(\mathrm{i}\gamma^\mu p_\mu + \gamma^5 m) = E^2 - p^2 - m^2 = 0. \tag{3.15}$$

Thus

$$\Psi = (\mathrm{i}\gamma^\mu p_\mu + \gamma^5 m), \tag{3.16}$$

from which our original wavefunction becomes

$$\psi = \Psi\chi e^{i(p\cdot r - Et)} = (\mathrm{i}\gamma^\mu p_\mu + \gamma^5 m)\chi e^{i(p\cdot r - Et)}. \tag{3.17}$$

It is clear Ψ is a free-particle solution to the Dirac equation, from which the standard wavefunction form may be obtained. Thus, Eq. (3.16) may be taken as the general state of a fermion in free space.

With the γs expressed in their usual form, Ψ represents the fermion as a 4×4 matrix. However, this form overlooks the generality of Eq. (3.16). The 16 terms of Eq. (2.9), along with their negative and imaginary cousins, form a finite group with 64 members. As such, the set forms an abstract algebra, of which 4×4 matrices are but one isomorphic form. Only the anti-commutation relation (2.3) is needed to satisfy the Dirac equation, and this constraint is sufficient to generate the finite group. Thus, Ψ represents the free fermion as a vector in a group algebra space. As such, Eq. (3.16) forms a connection between standard quantum theory and the group theory of particle physics. It is this connection which the PRJPC formulation [3, 4] attempts to exploit.

IV. The PRJPC Dirac Equation

Equation (3.16) holds for any free fermion. The group symmetries of fermions can therefore be derived from the symmetry of the finite γ-group. If one considers a composite fermion, such as the proton, the group symmetry of the component quarks must be a subgroup of the general fermion symmetry. Thus, Rowlands and Cullerne argue, [3, 4] the symmetry of the standard model must be somehow encoded in the γ-group symmetry.

Based on general symmetry arguements [5], PRJPC feel the best way to expose the standard model symmetry is through a quaternion/4-vector double notation. This is fully acceptable, since the finite γ-group is isomorphic to the algebra generated by two mutually commuting algebra:

$$\begin{gathered} IJ = -JI = K, \quad I^2 = -1 \text{ etc.}, \\ \sigma_i \sigma_j = -\sigma_j \sigma_i = \mathrm{i}\sigma_k, \quad \sigma_i^2 = 1 \text{ etc.}, \\ I\sigma_j = \sigma_j I \text{ etc.} \end{gathered} \tag{4.18}$$

Although these algebra have the properties of quaternions and Pauli matrices, they are not isomorphic to their usual 2×2 matrix forms. Due to their mutual commutation, a matrix representation for Eq. (4.18) must take a 4×4 form. In general, however, these algebra are not matrices, and are defined only via Eq. (4.18).

The pentad Dirac equation (3.14) can be easily obtained by the isomorphic substitution

$$\mathrm{i}\gamma^0 = K, \quad \mathrm{i}\gamma^j = I\sigma_i, \quad \gamma^5 = \mathrm{i}J, \tag{4.19}$$

which reduces (3.14) to

$$(K\mathrm{E} + \mathrm{i}I\sigma \cdot p + \mathrm{i}J\mathrm{m})\Psi = 0, \tag{4.20}$$

yielding the PRJPC solution

$$\Psi = (K\mathrm{E} + \mathrm{i}I\sigma \cdot p + \mathrm{i}J\mathrm{m}). \tag{4.21}$$

Since Eq. (4.21) represents a free fermion, it is always possible to choose the direction of motion as the primary coordinate axis, thus simplifying the solution to

$$\Psi = (K\mathrm{E} + \mathrm{i}I\mathrm{p} + \mathrm{i}J\mathrm{m}) \tag{4.22}$$

from which Rowlands and Culleme derive many of their results.

V. Conclusions

Whether the work of Rowlands and Culleme can generate a new interpretation of the standard model is yet to be determined. However, it is clear their algebraic form of the fermion is completely consistent with standard derivations. A recognition of this equivalence is necessary if we are to explore the implications of the PRJPC model.

REFERENCES

1. B. J. Hiley, in *Participations: Proceedings of ANPA 21*, ed. K. G. Bowden (Cambridge University, 2000), p. 114.
2. George Arfken, *Mathematical Methods for Physicists, 3rd Ed.* (Academic, San Diego, 1985), p 211.
3. Peter Rowlands and J. P. Culleme, in *Aspects II: Proceedings of ANPA 20*, ed. K. G. Bowden (Cambridge University, 2000), p 155.
4. Peter Rowlands and J. P. Culleme, in *Participations: Proceedings of ANPA 21*, ed. K. G. Bowden (Cambridge University, 2000), p. 89.
5. Peter Rowlands, in *Aspects II: Proceedings of ANPA 20*, ed. K. G. Bowden (Cambridge University, 2000), p. 123.

The Nilpotent Representation of the Dirac Algebra

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 7ZE, 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 jpc@wincoll.ac.uk

In response to comments made in *Participations* (Hiley, 2000), the procedure we have used in generating our version of the Dirac algebra is given in detail, together with an outline of the results (some new) which seem to emerge more naturally from this representation.

THE PROCEDURE

In explaining our work on the Dirac algebra, it is important to stress that its primary purpose is physical, rather than mathematical. We do not believe that an equation which appears to be one of the most fundamental in nature can only be described using a complicated algebra which appears to have no fundamental physical explanation, and which involves the mysterious extra physical concept of spin. Our approach, therefore, has been to seek for the *physical* origin of this algebra, and then use the additional insights given by this approach in reconstructing the equation in a form which makes its physical meaning more transparent.

In the first instance, our reasoning has been inductive, as was Dirac's in 1928. We have found what we consider may be a fundamentally 'physical' version of the algebra, and have used that to construct an equivalent to the Dirac equation from first principles; but we have then used deductive reasoning to relate this to the more familiar forms of the Dirac equation and shown that the equivalence is valid. However, our new form of the equation is meant to be *physically equivalent* to the standard version of Dirac, not algebraically isomorphic, and a literal application of the more restricted algebra used in the standard version will not produce it, any more than a literal application of Heisenberg's version of quantum mechanics will produce Schrödinger's.

The steps in the procedure might be described as follows:

(1) On the assumption that mass-charge is described by a quaternion, while space-time is described by a 4-vector, create a 32-part algebra from their combination. The *i* term of the 4-vector ensures that this is a complexified 4-vector-algebra.

(2) Find a primitive set of components which will generate the whole algebra that will match the gamma matrices of the Dirac algebra. This turns out to be an anti-commuting pentad, but it is significant that it can exist in more than one form.

(3) Invoke the presumed full symmetry between space-time and mass-charge to make the 4-vector 'quaternion-like', which means that the vector part becomes multivariate, or isomorphic to the set of Pauli matrices. Multivariate vectors

have a 'full' product that is equivalent to the scalar product plus *i* times the vector product.

(4) Use existing results involving multivariate vectors applied to the momentum operator in the Schrödinger equation (Gough, 1990) to hypothesize that spin is generated by the multivariate nature of this operator, and does not require *extra* spinor terms in the wavefunction. The spin here comes from the extra vector product term in the full product.

(5) Apply the first four terms of the pentad (equivalent to γ^0, γ^1, γ^2, γ^3) to the standard form of the Dirac equation to produce a vector-quaternion form of the differential operator, with an unspecified wavefunction.

(6) Multiply this equation throughout from the left by γ^5 or ij.

(7) Recognize that one can redefine the γ^0, γ^1, γ^2, γ^3 terms using the variation allowed within the anticommuting pentads to create a new form of the Dirac equation in which the mass term is preceded by ij. In this form, the differential operator, when reduced to eigenvalues, becomes a nilpotent of the form $(kE + ii\,\mathbf{p} + ij\,m)$.

(8) Recognize that, if this equation is valid, and a plane wave solution is applied, then the wavefunction must also be a nilpotent incorporating the term $(kE + ii\,\mathbf{p} + ij\,m)$.

(9) Recognize that four solutions are immediately made possible by the use of $\pm E$ and $\pm \mathbf{p}$, with the multivariate nature of $\mathbf{p}$ allowing us to interpret it as either $\mathbf{p}$ or $\sigma.\mathbf{p}$ (if we can show that the correct value of spin is generated from our equation).

(10) Use the classical energy-momentum-rest mass equation to derive simultaneously all four solutions by the same quantization procedure of replacing eigenvalues by differential operators, as for the Schrödinger equation.

(11) Recognize that this derivation allows us to describe the variation in the four solutions *either* by varying the exponential term or by varying the differential operator (but not both). (Elaborated below.) (This is related to the Feynman description of negative energy particles being equivalent to positive energy particles going backwards in time, with reversal of the spin / momentum terms being equivalent to reversal in position coordinate.)

(12) Demonstrate that a matrix version of the equation allows us to vary the differential operator (and that we can even reduce to a single differential operator, if we construct our matrices in a particular way). (Elaborated below.)

(13) Recognize that varying the differential operator rather than the exponential, and incorporating the spin concept into a multivariate $\mathbf{p}$, allows us to construct a wavefunction made up of a column vector with terms $(kE + ii\,\mathbf{p} + ij\,m)$, $(kE - ii\,\mathbf{p} + ij\,m)$, $(-kE + ii\,\mathbf{p} + ij\,m)$, $(-kE - ii\,\mathbf{p} + ij\,m)$, and a *single* exponential term. This liberates the wavefunction from being confined to being an ideal, although a nilpotent could be taken as being an extreme case of an ideal.

(14) Recognize that, since the wavefunction operator and the eigenvalue of the differential operator are in every case identical, then we can assume that $\mathbf{p}$ means $\mathbf{p}$, $\sigma.\mathbf{p}$, in general or in any specified direction, without loss of generality, and that it can be for a field-free particle or a $\mathbf{p}$ (or E) involving field terms.

(15) Recognize that, if this description is valid, we can now generate all required wavefunctions using a purely operator approach. Everything emerges from

using a bra vector with four operator terms, or a ket vector with four operator terms, or combinations of these.

(16) Demonstrate that this form of the equation produces the correct spin and helicity relations.

(17) Demonstrate that one can start from this form of the equation, and, by various transformations, arrive at the standard Dirac equation.

(18) Demonstrate that one can start from the Dirac equation in any of its forms, and arrive at this new form of the equation. Inevitably, such derivations will always require some specification of the representation, since the two equations are *physically equivalent*, not isomorphic, but, in any individual case, such a representation can always be made.

(19) Define the procedure for normalisation in the form $\psi\psi^*$.

(20) Using the adjoint wavefunction, derive the bilinear covariants to show that the Dirac current is zero in the absence of an external field, and construct the appropriate Lagrangian.

(21) Construct annihilation and creation operators, and show that second quantization is unnecessary since the new form of the equation can be shown to be derivable from the quantum field integrals.

(22) Demonstrate that routine results, such as the hydrogen atom, follow just as easily from this model as from any other.

Except for (20)-(22), these steps are explicit or implicit in our previous papers [Rowlands and Cullerne, 1999, 2000a]. (20)-(22) can be found, with much other material, in Rowlands and Cullerne [2000b].

EXPLANATIONS

From these steps, we believe that we can produce the following:

(1) A vacuum wavefunction, which is *k* times the column vector and exponential term for the fermion, with a vacuum *operator*, which is just *k* times the column vector. All fermion states which may be produced may then be considered as acting on the vacuum wavefunction, and the exponential part of the fermion wavefunction be regarded as, in origin, a vacuum term, expressing all possible space and time variations of a state in the vacuum.

(2) C, P and T transformations using the three quaternions operators, with consequent demonstration of CPT symmetry, etc.

(3) Immediate Pauli exclusion for identical fermions.

(4) A clear reason why the Dirac equation cannot apply to bosons, although the Klein-Gordon equation can be applied to fermions. A nilpotent wavefunction requires a differential operator with nilpotent eigenvalue to produce a zero product.

(5) An operator for the antifermion that reverses the *E* signs in the fermion operator, and that can be conveniently arranged as a bra vector if the fermion operator is a ket, and vice versa.

(6) Production of boson states by allowing a bra antifermion operator to act on a ket fermion state. The product is always a scalar.

(7) Differentiation between vector and scalar bosons according to whether the signs of the fermion and antifermion **p** terms are the same or different.

(8) An explanation of why vector bosons may be massless, but scalar bosons may not.

(9) An expression for the scalar wavefunction of a Bose-Einstein condensate, made of two fermions with opposite **p** states.

(10) An expression for the wavefunction of a baryon, with three symmetric and three antisymmetric terms, as required.

(11) Correct parity values for ground state baryons and bosons.

(12) A connection with approaches to the Dirac algebra that are constructed using quaternion or 4-vector operators to represent the four solutions, and an *explanation* for the existence of four solutions, and the validity of a 4 × 4 matrix for the differential operator.

(13) Annihilation and creation operators for the quantum field that are identical to the nilpotent operators and have the required algebraic relations. Our formalism already thus incorporates the quantum field, and can be shown equivalent to it.

(14) Supersymmetry operators that are identical to the bra and ket vectors used for fermions and antifermions.

(15) Infinite vacuum operation by a fermion state which is identical to an infinite alternating series of virtual boson and fermion states, as required for renormalization. Similar vacuum operation by bosons.

(16) Propagator terms for QED that immediately relate the Dirac propagator to the Klein-Gordon propagator, using the bra or ket operator for the fermion, and not requiring any averaging over helicity states.

(17) Explanation of the *origin* of the Dirac state, in reducing the eight original components of the algebra (1, *i*, *j*, *k*, *i*, **i**, **j**, **k**), derived from the physical mass, charge, time, and space, to a Dirac pentad, in which mass, time and space acquire, by combination, the characteristics of the three 'dimensions' of charge, when this disappears as an independent entity.

(18) The consequent origin of quantized rest mass in the Dirac state.

(1)-(12) can be found in Rowlands and Cullerne, [1999, 2000a]. (13)-(15) and (17)-(18) are in Rowlands and Cullerne, [2000b]. For (16), see below.

The main advantage in our approach, we believe, is that it *does away* with much of the algebra usually associated with the Dirac equation, and allows all tasks to be performed with a single vector operator. Everything can be made explicit; there are no mysterious wavefunctions or spinors hidden in the formalism. There is also the strong possibility that the group symmetry for the Dirac algebra may be identical or closely related to that for the Standard Model. We do not believe these results are accidental; the approach we have used generates physical results because it is itself ultimately based on physical arguments.

THE DIFFERENTIAL OPERATOR

In view of previous comments on our procedure (Hiley, 2000), we believe it worthwhile to give a fuller explanation of our approach to the Dirac differential operator. Replacing the four solutions

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

with common differential operator:

$$(kE + ii\,\mathbf{p} + ij\,m) = \left(ik\frac{\partial}{\partial t} + i\nabla + ijm \right) ,$$

by the set:

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

using the four differential operators:

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm \right) \qquad \left(ik\frac{\partial}{\partial t} - i\nabla + ijm \right)$$

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

requires a matrix version of the Dirac equation, such as:

$$(\alpha.\mathbf{p} + \beta m - E)\,\psi = \begin{pmatrix} -E & 0 & im & -i\mathbf{p} \\ 0 & -E & i\mathbf{p} & im \\ -im & -i\mathbf{p} & -E & 0 \\ i\mathbf{p} & -im & 0 & -E \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = 0 ,$$

with the terms E and $\mathbf{p}$ representing the usual space and time differentials. However, even this equation can be reduced to one with a single differential operator, for we can reduce the matrix to an eigenvalue by multiplying the matrix terms by assigned sets of row and column quaternion coefficients, the eigenvalue being a multiple of

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

The column vector here is a ket matrix of the form:

$$\begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = \boldsymbol{i}\,\psi_4 - \boldsymbol{j}\,\psi_2 + 1\,\psi_3 + \boldsymbol{k}\,\psi_1$$

where the row coefficients are identical to the column coefficients of the bra matrix. Since each product results in zero, multiplying the terms of the ket from the left by the eigenvalue is equivalent to multiplying each of them by the terms of a bra matrix, with each preceded by the same quaternion operator. That is, the set of equations:

$$(-kE + ii\,\mathbf{p} + ij\,m)\,k\,(kE + ii\,\mathbf{p} + ij\,m) = 0$$
$$(-kE + ii\,\mathbf{p} + ij\,m)\,(-j)\,(kE - ii\,\mathbf{p} + ij\,m) = 0$$
$$(-kE + ii\,\mathbf{p} + ij\,m)\,(-kE + ii\,\mathbf{p} + ij\,m) = 0$$
$$(-kE + ii\,\mathbf{p} + ij\,m)\,i\,(-kE - ii\,\mathbf{p} + ij\,m) = 0$$

is identical, in principle, to the set:

$$k\,(-kE + ii\,\mathbf{p} + ij\,m)\,k\,(kE + ii\,\mathbf{p} + ij\,m) = (kE + ii\,\mathbf{p} + ij\,m)\,(kE + ii\,\mathbf{p} + ij\,m) = 0$$
$$j\,(-kE + ii\,\mathbf{p} + ij\,m)\,(-j)\,(kE - ii\,\mathbf{p} + ij\,m) = (kE - ii\,\mathbf{p} + ij\,m)\,(kE - ii\,\mathbf{p} + ij\,m) = 0$$
$$(-kE + ii\,\mathbf{p} + ij\,m)\,(-kE + ii\,\mathbf{p} + ij\,m) = (-kE + ii\,\mathbf{p} + ij\,m)\,(-kE + ii\,\mathbf{p} + ij\,m) = 0$$
$$i\,(-kE + ii\,\mathbf{p} + ij\,m)\,i\,(-kE - ii\,\mathbf{p} + ij\,m) = (-kE - ii\,\mathbf{p} + ij\,m)\,(-kE - ii\,\mathbf{p} + ij\,m) = 0\,.$$

In other words, we can represent the same operation, by using a bra and ket matrix, with equivalent differential and wavefunction operators, without explicit use of quaternion operators in the matrix, or by an eigenvalue using a single differential operator acting on a ket matrix, with explicit use of quaternion components. And, of course, we can transform the sign of E in both operators to produce a more familiar set of relations, based on the eigenvalue $(kE + ii\ \mathbf{p} + ij\ m)$. The combination of eigenvalue and ket matrix now becomes:

$$(kE + ii\,\mathbf{p} + ij\,m)\begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\\ \psi_4\end{pmatrix} = (kE + ii\,\mathbf{p} + ij\,m)\,(1\ \psi_1 + \ i\ \psi_2 + k\ \psi_3 - j\ \psi_4)$$

This is the operator version. In differential terms, it can be written:

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm\right)\begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\\ \psi_4\end{pmatrix} = (kE + ii\,\mathbf{p} + ij\,m)\,(1\ \psi_1 + \ i\ \psi_2 + k\ \psi_3 - j\ \psi_4)$$

The result of this analysis is that we can use the same exponential term $e^{-i(Et\,-\,\mathbf{p.r})}$ for all the components of the Dirac spinor. The significant fact is that there are exactly four terms because there are four possible multipliers to create the different states of ± E and ± $\mathbf{p}$ in the quaternion system. The Klein-Gordon differential operator,

$$\left(ik\frac{\partial}{\partial t} + i\nabla + ijm\right)\left(ik\frac{\partial}{\partial t} + i\nabla + ijm\right) = \left(\frac{\partial^2}{\partial t^2} - \nabla^2 + m^2\right),$$

of course, applies irrespective of the signs of E and $\mathbf{p}$ in the exponential, because of the squaring, and boson wavefunctions carry over the single exponential term produced by applying an antifermion creation operator to a fermion state (or vice versa), with a scalar product which is a sum of such terms as

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

DIRAC AND KLEIN-GORDON PROPAGATORS

We include this as an example of a new preliminary result, to demonstrate that the method is effective in many areas, even when the term $(kE + ii\mathbf{p} + ijm)$ is no longer a nilpotent. In QED, we write the Dirac propagator

$$S_F(x - x') = (i\gamma^\mu \partial_\mu + m)\, \Delta_F(x - x')\,,$$

where $\Delta_F(x - x')$ is the Klein-Gordon propagator [Kaku, 1993]. In our notation, we can write this in the form:

$$S_F(x - x') = ((kE + ii\,\mathbf{p} + ij\,m)\,\ldots)\, \Delta_F(x - x')\,,$$

where $((kE + ii\,\mathbf{p} + ij\,m)\,\ldots)$ is the bra matrix with the terms:

$$\begin{array}{l} (kE + ii\,\mathbf{p} + ij\,m) \\ (kE - ii\,\mathbf{p} + ij\,m) \\ (-kE + ii\,\mathbf{p} + ij\,m) \\ (-kE - ii\,\mathbf{p} + ij\,m)\;. \end{array}$$

This is exactly what we would expect in transferring from boson (Klein-Gordon field) to fermion (Dirac field), using our single vector operator. Adapting the usual procedure, using the Green's function for the plane wave solutions, for the case in which variation over space and time (including the time-reversed solutions produced by $-E$ states) is transferred to the differential operator, we can then simply write

$$S_F(x - x') = \int d^3p \sqrt{\frac{m}{E}}\, \frac{1}{2E}\, (2\pi)^{-3/2}\, \theta(t - t')\, \Psi(x)\, \bar{\Psi}(x')\,,$$

where

$$\Psi(x) = ((kE + ii\,\mathbf{p} + ij\,m)\,\ldots)\, \exp(ipx)$$

and the adjoint term,

$$\bar{\Psi}(x') = ((kE - ii\,\mathbf{p} - ij\,m)\,\ldots)\, \exp(-ipx'),$$

with $((kE - ii\,\mathbf{p} - ij\,m)\,\ldots)\,(ik)$ now a ket. No averaging over spin states or 'interpreting' $-E$ as a reversed time state is necessary; the 'reversed time' state occurs with the t in the operator $\partial / \partial t$. Reinterpreting $\Psi(x)$ and $\bar{\Psi}(x')$ as the vacuum expectation values of quantized spinor fields, say $\psi(x)$ and $\bar{\psi}(x')$, we obtain results of the form:

$$i\, S_F(x - x')_{ab} = \langle 0|\, T\, \psi(x)_a\, \bar{\psi}_b(x')\, |0\rangle\,.$$

In effect, multiplying bra terms of the form $(kE + ii\,\mathbf{p} + ij\,m)\,\ldots)$ with ket terms of the form $((kE - ii\,\mathbf{p} - ij\,m)\,\ldots)(ik)$ results in a scalar multiple of the bra term, while the exponential multiple takes the form $\exp(ip(x - x'))$.

REFERENCES

Gough, W. [1999], 'Mixing scalars and vectors – an elegant view of physics'. *Eur. J. Phys.*, **11**, 326-33.

Hiley, B. [2000], 'Comments on the Dirac algebra and charge accommodation', in *Participations*, ed. K. G. Bowden, ANPA, 114-123.

Kaku, M. [1993], *Quantum Field Theory* (Oxford University Press).

Rowlands, P. and Cullerne, J. P. [1999], 'A derivation of particle structures and the Dirac equation from fundamental symmetries', in *Aspects II*, ed. K. G. Bowden, ANPA, 155-91.

Rowlands, P. and Cullerne, J. P. [2000a], 'The Dirac algebra and charge accommodation', in *Participations*, ed. K. G. Bowden, ANPA, 89-111.

Rowlands, P. and Cullerne, J. P. [2000b], 'The Dirac algebra and its physical interpretation' (36 pp), xxx.lanl.gov/quant-ph/00010094.

A Note on the Role of Idempotents in the Extended Heisenberg Algebra.

B. J. Hiley.

Theoretical Physics Research Unit, Birkbeck, Malet Street, London WC1E 7HX.

b.hiley@bbk.ac.uk

Abstract.

Motivated by a process explanation of quantum phenomena, we explore an algebraic representation-free approach to the quantum formalism. We show that the Schrödinger equation can be written in a representation-free way provided the appropriate algebra contains idempotents. To fit into the general scheme, the nilpotent Heisenberg algebra must be extended to include suitable idempotents. We show that the extended boson algebra, which includes the projector to the vacuum, can be used to generate these new idempotents using the Heisenberg and metaplectic groups. We briefly discuss some of the consequences of this approach.

1. Introduction

In this note I want to re-examine the Dirac algebraic approach to quantum mechanics. Like Dirac (1956) I feel the representation free approach to quantum mechanics offers greater possibilities. For reasons which have to do with my speculations about the nature of quantum processes, I want to make the algebraic structure primary and work for as long as possible in this algebraic structure. If you like, I want to take up the suggestion of Haag (1992) (Local Quantum Theory) "The intrinsic structure of the [quantum field] theory is fully characterised by the algebraic relations in the net of abstract algebras (as opposed to their representative algebras on a Hilbert space.) The divorce of the basic description of the theory from Hilbert space brings a tremendous additional freedom, that is it allows the introduction of thermodynamics at the quantum level." Dirac (1956) expressed a similar sentiment, but did not mention thermodynamics. He also claimed the approach offered a more general way of dealing with situations where it was not possible to represent results in Hilbert space. His concern was that there did not seem to be any readily available interpretation to explain this approach. I have claimed elsewhere (Hiley 2000) that a generalised notion of process which is basic to the implicate order begins to provide an explanation. I will not justify this proposal here.

Since the unfolding process described in Hiley (2000) takes the form of an inner automorphism, I start with the Heisenberg picture and consider the time development operator to be a one-parameter automorphism of a general element of the appropriate algebra in the usual way. Thus $\exists$ an $M(t)$ such that

$$M(t)\colon A \rightarrow A \qquad \forall A \in \mathcal{A}$$

such that $\quad A(t) = M(t)^{-1}A(0)M(t) \qquad$ with $M(t) = \exp[iHt]$

Here H of course is the Hamiltonian. The infinitesimal transformation gives the Heisenberg equation of motion

$$i\frac{dA}{dt} = [A, H] \tag{1}$$

We then compare this with the corresponding classical equation of motion with the Poisson brackets replacing the commutator. Thus

$$\frac{dA}{dt} = \{A, H\} + \frac{\partial A}{\partial t} \tag{2}$$

where we have included the explicit time dependence. if we also allow for the explicit time dependence in the Heisenberg equation of motion, then (1) becomes

$$i\frac{dA}{dt} = [A, H] + i\frac{\partial A}{\partial t} \tag{3}$$

Physicists usually derive this equation by first starting from the Schrödinger picture. Because of my starting point, I want to reverse this procedure and arrive at the Schrödinger equation by starting from the algebra

We know that in the usual approach the Hamilton acts on a vector in Hilbert space. But we want to remain in the algebra without first going to an abstract Hilbert space. Therefore we need to look for a natural vector space in the algebra. In fact there is such a natural vector space in the algebra, namely, a left ideal, $\mathcal{I}_{L\omega}$[1]. There is also exists a dual

---

[1] Recall that a left ideal is defined by $\mathcal{I}_{L\omega} = \{A \in \mathcal{A} : \omega(B^*A) = 0, \forall B \in \mathcal{A}\}$. While the right ideal is defined by $\mathcal{I}_{R\omega} = \{A \in \mathcal{A} : \omega(A^*C) = 0, \forall C \in \mathcal{A}\}$. Here ω is a mapping $\omega : A \rightarrow \Re$. ω is called the

vector space, the right ideal, $I_{R\omega}$. Unfortunately if we act from the right on an element of the left ideal we will in general leave the left ideal. Thus we cannot put an element from a left ideal into equation (3) and remain in that left ideal. However if we form a two-sided ideal such that any element belonging to it can be written in the form $A = BC$, with $B \in I_{L\omega}$, and $C \in I_{R\omega}$. Such an element can be put into equation (3) and it will remain in the two-sided ideal.

We must restrict the choice of A further and assume $dA/dt = 0$. Our reason for introducing this restriction will become clear from equation (10). We these assumptions we write equation (3) in the form

$$i\frac{\partial A}{\partial t} = [H, A] \tag{4}$$

Since $A = BC$ equation (4) becomes

$$i\left(\frac{\partial B}{\partial t}\right)C + iB\left(\frac{\partial C}{\partial t}\right) = (HB)C - B(CH).$$

Multiplying this equation from the left by B^{-1} and from the right by C^{-1}, we find

$$i\,B^{-1}\left(\frac{\partial B}{\partial t}\right) - B^{-1}HB = -i\left(\frac{\partial C}{\partial t}\right)C^{-1} - CHC^{-1}.$$

Since B is any element of the left ideal C any element of a right ideal, we can write

$$i\left(\frac{\partial B}{\partial t}\right) = HB \tag{5}$$

and

$$-i\left(\frac{\partial C}{\partial t}\right) = CH \tag{6}$$

We shall see immediately that equation (5) and equation (6) have the same *form* as the Schrödinger equation and its conjugate counterpart. However it should be noted that here B and C are elements of the algebra and not elements of a Hilbert space. Because of this we will call B and C wave operators.

---

state of the system. It is defined by $\omega(A_i) = \alpha_i$, $\forall A_i \in \mathrm{A}$ where $\alpha_i \in \Re$. The set $\{\alpha_i\}$ is determined by the particular physical problem under consideration.

Thus we see that the algebraic content of the Schrödinger equation is equivalent to the one-parameter transformation on ideals in the algebra

$$M(t)\colon B \to B \qquad \forall B \in I_{L\omega} \qquad \text{such that} \qquad B(t) = M(t)^{-1}B(0).$$

While the conjugate equation is equivalent to

$$M(t)\colon C \to C \qquad \forall C \in I_{R\omega} \qquad \text{such that} \qquad C(t) = C(0)M(t).$$

It should be noted in coming to this conclusion, we are using equation (3) and noting that, since $dA/dt = 0$, and B and C are independent, $dB/dt = dC/dt = 0$.

We now need to show that equations (5) and (6) become the standard Schrödinger equation and its conjugate. In order to do this we need to make a closer connection with the Dirac notation. For this reason we will write $A = B\rangle\langle C$.[2] Further we will assume $B\rangle = B(X, t)\rangle \in I_{L\omega}$ [3]. Then in order to project this element onto a Hilbert space, we follow Dirac (1947) and define

$$\langle x | B\rangle = B(x,t)$$

which is now the usual wave function, conventionally written as $\Psi(x, t)$[4]. It is easy to show that equation (4) becomes the Schrödinger equation

$$i\frac{\partial\Psi(x,t)}{\partial t} = \int H(x,x')\Psi(x',t)dx' \qquad (7)$$

When $H(x',x)$ is diagonal, we can write this in the more usual form

---

[2] NB. The symbol $\rangle$ (without the line $|$) is NOT the ordinary ket, $|\ \rangle$. Dirac calls this the standard ket. For further details see Dirac (1947, p. 79-83.) We re-introduced this notation to emphasise that B in an element of the left ideal, which, we emphasise again, belongs to a vector sub-space of the algebra itself. The usual ket $|a\rangle$ is an element of an abstract vector space, which is labelled by an eigenvalue of some appropriate observable.

[3] We use upper case letters for elements of the algebra and lower case for the eigenvalues of these elements.

[4] Not all elements of a left ideal produce state functions that are physically meaningful. We will not discuss these restriction here (See Ballentine 1990).

$$i\frac{\partial\Psi(x,t)}{\partial t}=H(x)\Psi(x,t) \tag{8}$$

The conjugate equation can be derived by assuming $\langle C=\langle C(X,t)$ and now multiplying from the right by the usual position ket $|x\rangle$,we find

$$\langle C(X,t)|x\rangle=C^*(x,t)=\Psi^*(x,t).$$

So that equation (5) becomes

$$-i\frac{\partial\Psi^*(x,t)}{\partial t}=\int\Psi^*(x',t)H(x',x)dx'=H(x)\Psi^*(x,t) \tag{9}$$

In this way we make contact with the usual approach.

There is one more relationship worth pointing out here to clarify the notation we are using. In the x-representation, $A=B\rangle\langle C$ becomes

$$\langle x|A|x'\rangle=\langle x|C\rangle\langle B|x'\rangle=\Psi^*(x,t)\Psi(x',t)=\rho(x,x',t). \tag{10}$$

$\rho(x,x',t)$ is immediately recognised as the density matrix in the x-representation. Thus a sub-class of these the two-sided ideals are playing the role of density operators. It is now clear why we imposed the assumption $dA/\mathrm{dt}=0$. It is introduced to ensure the density operator will satisfiy Liouville's theorem. Further equation (4) is immediately recognised as the Liouville equation.

By assumption that $A=\Psi\rangle\langle\Psi$ in the above, we have restricted ourselves to pure state density operators. It is straightforward to generalise this procedure to mixed states but we will not carry that out here.

From the mathematical point of view the status of the symbols $\rangle$, $\langle$, and $\rangle\langle$ can be considered as unsatisfactory. Dirac simply introduced them to distinguish between elements that could be operated on from the left, $\rangle$, and right, $\langle$. He did not consider the role of $\rangle\langle$. As we will show the symbol $\rangle\langle$ plays the role of an idempotent. An idempotent can be used to define left and right ideals. To see this, consider an algebra with a non-trivial idempotent, $\varepsilon=\varepsilon^2$. We can then construct a left ideal in the following manner,

$$\mathbb{I}_{L\varepsilon}=\{A\in\mathbb{A}\colon \mathbf{A}=B\varepsilon,\ \forall\, B\in\ \mathbb{A}\} \tag{11}$$

Also the right ideal can be defined through

$$\mathrm{I}_{\mathrm{R}\varepsilon} = \{A \in \mathrm{A}: \mathbf{A} = \varepsilon B,\ \forall\, B \in \ \mathrm{A}\} \tag{12}$$

Note that since $\mathrm{I}_{\mathrm{R}\varepsilon}: \quad \mathrm{I}_{\mathrm{L}\varepsilon} \rightarrow \mathrm{C}$,

$$\langle B(X,t) \bullet B(X,t)\rangle = \int \langle B(X,t)|x\rangle\langle x| B(X,t)\rangle dx \in \Re$$

Thus in an algebra with an idempotent, the idempotent can be used to define a state, which can be chosen to be equivalent to an ω defined in footnote (1). Furthermore if A contains a non-trivial idempotent ε, we can identify ε with $\rangle\langle$ so that $A = B\rangle\langle C = B\varepsilon C$[5]. Thus $\rangle\langle$ is part of the algebra. The reason why Dirac was forced to bring in this symbol was because the Heisenberg algebra, being nilpotent has no non-trivial idempotents. We will show how to remedy this in the next section.

Before doing this, however, I would like to summarize the position we have reached so far. From the point of view we are adopting in this note, writing the Schrödinger equation in the form (5) has the advantage that it is independent of the representation we ultimately choose. What we will also do is to assume that equations (5) and (6) are independent. This then allows us to form two new equations, the first by subtracting the two equations and the second by adding the two equations.

Subtracting the two equations gives us back the Liouville equation (4). As is well known this gives us an equation for the conservation of probability.

The second equation gives us a new equation. If we write $B\rangle = Re^{iS}\rangle$This is can be written in the form

$$\rho_R \frac{\partial S}{\partial t} + \frac{1}{2}[\rho, H]_+ = 0 \tag{13}$$

where $\rho_R = R^2$. This equation describes the time development of the phase operator and if the energy is well defined then the equation is an expression for the conservation of energy. It is, in fact the quantum generalisation of the Hamiltonian-Jacobi equation. This

---

[5] If there are more than one idempotent in the algebra, this identification is not unique. As we will show in the next two sections, we can generate many equivalent idempotents in the extended enveloping Heisenberg algebra. In most cases in physics we do not need to identify these idempotents and this is why the Dirac notation is so useful to physicists.

can be seen by multiplying from the left by <x| and from the right by |x> which produces the equation

$$\frac{\partial S(x)}{\partial t} + H(x) + Q(x) = 0 \tag{14}$$

where $Q(x) = -\frac{1}{2m}\frac{\partial^2 R(x)}{\partial^2 x}\frac{1}{R(x)}$ is the quantum potential. If we regard $\hat{H} = H + Q$ as being the effective Hamiltonian, then the classical canonical formalism gives

$$\dot{x} = \frac{\partial \hat{H}}{\partial p} \qquad \dot{p} = -\frac{\partial \hat{H}}{\partial x} \tag{15}$$

The first equation in equation (14) is just the 'guidance condition' usually written in the form

$$\dot{x} = \frac{1}{m}\,\mathrm{Im}\frac{\partial \psi}{\partial x}\frac{1}{\psi} \tag{16}$$

Thus the representation-free formalism enables us to obtain a different insight into the Bohm interpretation. Some further implications of this equation are discussed in detail in Brown and Hiley (2000).

2. The Nilpotent Heisenberg Algebra.

Clearly I would like to apply the above techniques to the Heisenberg algebra, but as we have already remarked the Heisenberg algebra is a nilpotent algebra of degree three under the product $[A, B]$. As a consequence of a well-known theorem, nilpotent algebras do not contain any non-trivial idempotents and if we have no non-trivial idempotents, we cannot construct non-trivial left ideals.

To remedy the absence of such an idempotent in the Heisenberg algebra, Frescura and Hiley (1984) followed a suggestion made by Schönberg (1958) who showed that it was possible to generalise the (enveloping) Heisenberg algebra by adding to it a new fundamental idempotent. Unfortunately his papers were long and the methods far from clear. In this paper we simplify his ideas and bring out the full implications of his method using techniques that will be recognised by physicists.

We begin by noting that the Heisenberg algebra, HA, is isomorphic to the boson algebra, BA, defined through creation and annihilation operators satisfying the commutation relations $[a, a^\dagger] = 1$. (For simplicity we will consider one pair of operators only.) This isomorphism can be realised through the Bargmann transformation, which formally establishes the well-known relations

$$a = Q + iP, \quad \text{and} \quad a^\dagger = Q - iP,$$

What physicists actually work with in physics are the respective enveloping algebras, EHA and EBA. Clearly the EBA is also nilpotent. However this has not caused any problems for physicists because they have introduced the projector onto the vacuum state, V, which satisfies the conditions

$$aV = 0, \qquad Va^\dagger = 0 \qquad \text{and} \qquad V^2 = V \tag{17}$$

The algebra generated by $\{1, a, a^\dagger, V\}$ is no longer nilpotent because $[a, V] = Va$ and $[a^\dagger, V] = a^\dagger V$ so that higher order commutators do not vanish. We will call this algebra the extended enveloping boson algebra, EEBA. [6]

All of this is very straightforward once we realise that in the Dirac notation, $V = |0\rangle\langle 0|$, where $|0\rangle$ is the vacuum state. We can write an element in the form

$$\sum_{m,n} c_{mn} V^{mn} = \sum_{m,n} c_{mn} (a^\dagger)^m |0\rangle\langle 0| a^n = \sum_{m,n} c_{mn} |m\rangle\langle n|,$$

which, once again, is easily recognised as the density operator, $\rho(m,n)$. This emphasises that it is the density operator that lies at the heart of the algebraic description. If we require a pure state then consider the special case with $\rho^2 = \rho$.

Notice, we could have written $B(X,t)\rangle = B|\ \rangle$ in analogy with $F(a^\dagger,t)\rangle = F(a^\dagger,t)|0\rangle$ in which case we see the standard ket plays a role in the Heisenberg algebra that the vacuum state plays in the boson algebra. Furthermore, as we have already remarked, $\rangle\langle = \varepsilon$, plays the role of an idempotent, the projector onto the standard ket, which again is analogous to V. (Frescura and Hiley (1984) have already obtained this result.) Since ε is an idempotent equation (11) tells us that $B = A\varepsilon$, $\forall A \in \mathrm{A}$, generates the left ideal.

---

[6] It is interesting to note that V is the limit point $V = \lim_{\beta\to\infty} \exp[-\beta a^\dagger a]$

In the structure we have outlined so far, the idempotent ε is not well defined since we not know the product rule for ε with P. We need an extra defining relation and the simplest rule would be $P\varepsilon = 0$. If we assume $\varepsilon^{\dagger} = \varepsilon$, then $\varepsilon P = 0$ and ε projects the EEHA onto the polynomial sub-algebra $F(X)$.

By symmetry we could also extend the EHA by introducing a different idempotent Π such that, $\Pi^2 = \Pi$ and $X\Pi = 0$. Again if $\Pi^{\dagger} = \Pi$ then $\Pi X = 0$, thus Π projects the EEHA onto the polynomial sub-algebra $\Phi(P)$.

We could also extend the EHA by including both idempotents and make some assumption about their product, for example, $\varepsilon\Pi = \Pi\varepsilon$. Such an assumption is arbitrary and, as we will show shortly, is incorrect. What we now show is how these idempotents can be derived from the EEBA, which includes V. We will see that it is the presence of V that enables these new idempotents to be generated.

3. Relations between the idempotents.

3.1 The Heisenberg Group.

So far we have provided no way to systematically generate the idempotents that we require. To show we can do this we must consider the structure of EHA a little more carefully.

The EHA contains two groups of inner automorphisms in which we are interested. The first is the Heisenberg group generated by linear elements of the algebra. The general element of this group can be written as

$$H(a,b,c) = \exp[\alpha X + \beta P + \varepsilon Z]$$

where $Z \in \Re$. Details of the structure of the Heisenberg group will be found in Guillemin and Sternberg, (1984).

To illustrate the method we choose a specific element of this group

$$\exp[\beta P - \varepsilon X)] = \exp[\alpha a^{\dagger} - \alpha^{*} a] = D(a)$$

where $\alpha, \beta, \varepsilon \in C$. Some will immediately recognise that $D(a)$ generates coherent states, states that are well known in quantum optics. We can also use this operator to generate a new idempotent, the projector onto the coherent states, in the following manner

$$\Omega = |\alpha\rangle\langle\alpha| = D(a)VD^{-1}(a)$$

Umezawa has an interesting way of looking at the coherent state $|\alpha\rangle$. He regards it as the α-vacuum state, which is a superposition of states with many particles and can be thought of a condensation of many particles. This statement becomes more transparent by noting that

$$|\alpha\rangle = \exp[-\tfrac{1}{2}|\alpha|^2]\exp[-\alpha a^{\dagger}]|0\rangle$$

As the properties of the coherent states are well known and we will not discuss them further here, particularly as these are not the idempotents that we need.

3.2 The Metaplectic Group.

The second group of inner automorphisms contained in the EHA that is of importance to our discussion can most easily be seen by first noting that the EHA is isomorphic to the symplectic Clifford algebra. (See Crumeyrolle 1990.) This algebra is analogous to the usual orthogonal Clifford algebra, which is known to contain the spin group formally known as the Clifford group. The Clifford group is generated by bilinear combinations of the generating elements of the algebra. The corresponding spin group generated by bilinear combinations of the symplectic Clifford algebra is known as the metaplectic group which double covers the symplectic group. We expect this metaplectic group to give rise to some symplectic spinor properties.

The generators of the metaplectic group can be written in the following form

$$M = \exp[\alpha X^2 + \beta P^2 + \varepsilon(XP + PX)].$$

Such elements are not unfamiliar in quantum optics and form the basis of the description of squeezed states. These states are generated by operators of the form

$$M_s = \exp[\tfrac{i}{2}(a^2 - a^{\dagger 2})]$$

I will not discuss the properties of squeezed states here. I merely introduce them to show that elements of the metaplectic group have physical significance. What I now want to do is to show how elements of the metaplectic group enable us to generate the idempotents we need from the vacuum projector, V.

We find

$$\varepsilon = \sqrt{N}\exp[X^2/2]V\exp[X^2/2] \qquad \text{and} \qquad \Pi = \sqrt{N'}\exp[P^2/2]V\exp[P^2/2]$$

where N and N' are suitable constants to ensure $\varepsilon^2 = \varepsilon$ and $\Pi^2 = \Pi$. It is a straightforward but tedious task to show that $P\varepsilon = \varepsilon P = 0$ and $X\Pi = \Pi X = 0$. Furthermore we can show that

$$\varepsilon\Pi = N_1\exp[X^2/2]V\exp[P^2/2]$$

and

$$\Pi\varepsilon = N_2\exp[P^2/2]V\exp[X^2/2].$$

We will show below that the expressions on the right hand side of these last two equations are essentially the idempotent E that we have used elsewhere in Frescura and Hiley (1984) and in Hiley and Monk (1993). We now call the conjugate $E^\dagger = \Delta$ and will show that Δ is the algebraic equivalent of the Dirac delta function $\delta(x)$.

4. More Idempotents.

Consider the two idempotents defined by

$$E = \sqrt{N}\exp[X^2/2]V\exp[P^2/2] \quad \text{and} \quad \Delta = \sqrt{N'}\exp[P^2/2]V\exp[X^2/2].$$

Where N and N' are again some numbers introduced to ensure that E and Δ are idempotent. For convenience we will omit writing these numbers from the discussions below since they do not alter the results.

We first start by recalling an old suggestion made by Weyl (1930). He noticed that in quantum mechanics we are interested in ray representations so that we can consider iQ and iP as infinitesimal unitary rotations of the ray field. The Heisenberg commutation rules show that these ray rotations are commutative. Weyl then suggested that we should explore the properties of Abelian groups of unitary rotations in the ray field of an n-dimensional space. This suggestion has been taken up by Schwinger (1960). Davies et al

(1982), and Monk and Hiley (1993) explored this structure, which was called the discrete Weyl algebra, and showed how to construct a discrete space with interesting properties.

This algebra is generated by $\{1, A, B\}$ subject to

$$AB = \omega BA \qquad A^n = 1, \qquad B^n = 1, \qquad \text{with } \omega = \exp[2\pi i/n].$$

Weyl shows that we can write $A = \exp[i\xi P]$, and $B = \exp[i\eta Q]$. In the limit as $n \to \infty$ this algebra approaches the Heisenberg algebra. However the discrete Weyl algebra is not nilpotent and therefore has idempotents that are not difficult to construct from elements of the algebra.

In Hiley and Monk (1993) we show how these idempotents can be constructed. One such idempotent can be written in the form

$$\varepsilon_{00} = \frac{1}{n}\sum_{\beta} \exp[i\beta Q] \quad \to \quad \frac{1}{2\pi}\int d\beta \exp[\beta Q] \quad \to \quad \delta(q) \tag{18}$$

We have followed the limiting process described by Weyl. Thus our idempotent ε_{00} is playing the role of a point at the origin of the co-ordinates. The limiting process suggests that we need to extend the Heisenberg algebra by adding the idempotent, Δ, that is to play the role of the algebraic equivalent of the Dirac delta function at the origin. Thus our EEHA is generated by $\{1, Q, P, \Delta\}$. We will show that Δ is exactly the idempotent defined above and we will evaluate the multiplication rules with other elements of the algebra.

Before doing this, we will recall in more detail how the ideas discussed in Hiley and Monk (1993) are related to Weyl's discussions. With the idempotent defined in equation (10), we can translate to a new point through the relation $\varepsilon_{\alpha\alpha} = A^{-\alpha}\varepsilon_{00}A^{\alpha}$. If we act only on one side with $A^{-\alpha}$, this will corresponds to the first expression in the Weyl equation (15.2), $A^s : x'_k = x_{k+s}$, which in the limit gives $\psi(x) \to \psi(x - \alpha)$. Thus $A^{-\alpha}$ produces a translation through α. Our equation (3.6), namely, $B^{\beta}\varepsilon_{\alpha 0} = \omega^{\alpha-\beta}\varepsilon_{\alpha 0}$ corresponds to the second part of the Weyl equation (15.2), viz: $B^t: x'_k = \varepsilon^{kt}x_k$, which in the limit corresponds to $\psi(x) \to \exp[i\beta x].\psi(x)$. Thus we see how the Schrödinger representation actually emerges from our algebraic structure as a limiting feature.

In order to identify the algebraic structure of Δ we have to show that $A^{\beta} = \exp[i\beta P]$ translates Δ to Δ_{β} such that

$$\Delta_\beta = A^{-\beta}\Delta A^{\beta} = \exp[-i\beta P]\Delta\exp[i\beta P]$$

and if we let $X\Delta = 0$ then

$$X\Delta_\beta = X\exp[-i\beta P]\Delta\exp[i\beta P] = \exp[-i\beta P](X+\beta)\Delta\exp[i\beta P] = \beta\Delta_\beta$$

If we write $\Delta = \exp[P^2/2]V\exp[X^2/2]$ then $X\Delta = 0$ so we can label $\Delta = \Delta_0$. Then

$$\Delta_x = \exp[-ixP]\Delta_0\exp[ixP]$$

Thus we have generated a continuum of delta functions producing a straight line. Thus our conditions for Δ are

$$X\Delta = 0, \qquad \Delta P = 0, \qquad \text{and} \qquad \Delta^2 = \Delta.$$

Then the $E = \Delta^\dagger$ introduced by Frescura and Hiley (1984) satisfies

$$EX = 0, \qquad PE = 0, \qquad \text{and} \qquad E^2 = E.$$

This completes our identification of the idempotents that we have used in our previous work.

5. Conclusions.

We have shown that if we extend the boson algebra by incorporating the limit point V, which physicists identify with the projection onto the vacuum, we use V to generate idempotents to extend the Heisenberg algebra to EEHA. We have shown these idempotents are generated by using elements of the Heisenberg and the metaplectic groups. The consequence of this is that EEHA with its new set of idempotents is isomorphic to EEHA just as EBA is isomorphic to EHA.

Within the EEHA we can now discuss the Heisenberg equation and Schrödinger equation in a representation-free manner. We then have the possibility of approaching quantum mechanics in a more general way. As we show in Brown and Hiley (2000) this leads us to see the Schrödinger equation as describing simultaneously both the conservation of energy and the conservation of probability. These equations can also be extended to the relativistic Dirac equation.

The conservation of energy equation can also be interpreted as giving phase information which leads to a very simple account of gauge effects. This approach also enables us to see the Bohm interpretation in a new light. It shows that the non-commutative structure enables us to construct "shadow phase spaces". The recent work of de Gosson (2001) has shown how these ideas are related to the work on Lagrangian quantisation. This shows the key role the metaplectic group is playing in quantum mechanics.

As a further example of the role of the metaplectic group, Fernandes and Hiley (2000) have shown how the symplectic spinor, which is central to the symplectic Clifford algebra can account for the Guoy phase in optics and the discontinuous change of phase when light passes through a focal point.

In this paper we have also re-examined how the discrete Weyl algebra can be extended to the continuum limit. The novelty of our approach is that the points of the continuum are carried in the algebra itself. This is done by constructing the special idempotent, Δ, which in the algebra plays the role of the Dirac delta function. This additional element has enabled us to complete the discussion of how the discrete Weyl algebra approaches the continuum limit.

References.

Ballentine, L. E., 1990, *Quantum Mechanics*, pp. Prentice Hall, New Jersey.

Bohm, D. J., Davies, P. G., and Hiley, B. J.,1982, *Algebraic Quantum Mechanics and Pre-geometry*, unpublished.

Brown, M. R. and Hiley, B. J., 2000 *Schrödinger revisited: the role of Dirac's 'standard' ket in the algebraic approach.* quant-ph/0005026.

Crumeyrolle, A. 1990 *Orthogonal and Symplectic Clifford Algebras: Spinor Structures*, p. 233 Kluwer, Dordrecht.

De Gosson, M 2001, *The Mathematical Principles of Newtonian and Quantum Mechanics*, World Scientific, to be published.

Dirac, P. A. M. 1947, *The Principles of Quantum Mechanics*, Oxford University Press, Oxford.

Dirac, P. A. M., 1965 *Quantum Electrodynamics without Dead Wood*, Phys. Rev., **139B,** 684-90.

Fernandes, M. and Hiley, B. J., 2001, *The Metaplectic Group, the Symplectic Spinor and the Guoy Phase*, (To appear in 2001).

Frescura, F. A. M. and Hiley, B. J. 1984, *Algebras, Quantum Theory and Pre-Space*, Rev. Brasil. Fis., Volume Especial, Os 70 anos de Mario Schönberg, 49-86.

Guillemin, V. and Sternberg, S. 1984, *Symplectic Techniques in Physics*, Cambridge University Press, Cambridge.

Haag, R., 1992, *Local Quantum Physics*, Springer, Berlin.

Hiley, B. J. 2000, *Non-Commutative Geometry, the Bohm Interpretation and the Mind-Matter Relationship*, Proc. Fourth Int. Conf. on Anticipatory Systems, CASAY 2000, Liege, Belgium, August 2000. (Accepted. To appear in 2001)

Hiley, B. J. and Monk, N. 1993, *Quantum Phase Space and the Discrete Weyl Algebra*, Mod. Phys. Lett., **A8,** 3225-33.

Schönberg, M., 1958, *Quantum Mechanics and Geometry*, An. Acad. Brasil. Cien., **30**,1-20.

Schwinger, J., 1960, *Unitary Operator Basis*, Proc. Natl. Acad. U. S. A., **46**, 570-579.

Weyl, H. 1931, *The Theory of Groups and Quantum Mechanics*, Dover, London.

Complexity Categorified

PETER EISENHARDT; DAN KURTH

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

I Introduction

The aim of this paper is to show that n-category theory or *categorification* is a proper tool for modeling - at least the first steps - of the non-topological mapping underlying nature's process of *complexification.*

II From Discontinuity ...

Continuous transformations or mappings are well known and deep-rooted ways to represent nature's processes. Think of the calculus, especially the concept of derivation, or recall differentiable manifolds. It seems as if almost everything could be described as an *iteration of morphisms*, that is structure-preserving mappings. But how to take into consideration that *natura facit saltus* ? (Nature is connected discontinuously yet doesn't jump discretely.) What about the abundant and fundamental discontinuities and instabilities which have to be covered? In rough outline the following alternative suggests itself:

1. 'External Approximation'
Selecting a specific iteration of morphism one tries to get closer and closer to discontinuities, for example by folding of smooth manifolds and considering a trajectory of some sort of 'test-particle' that goes along the ripples of the manifold, by which the change of an external parameter is represented. That's realized in catastrophe theory, see figs. 1a and 1b.

2. 'Internal Analytical Solution'
Yet if taking the discontinuities seriously one have to build them in the theory in an *inherent* way. The most radical possibility consists then in replacing morphisms by antimorphisms, that means for example fundamental changes of topology. We do prefer this method. But before we will present our ansatz, let us have a look at a similar thought.

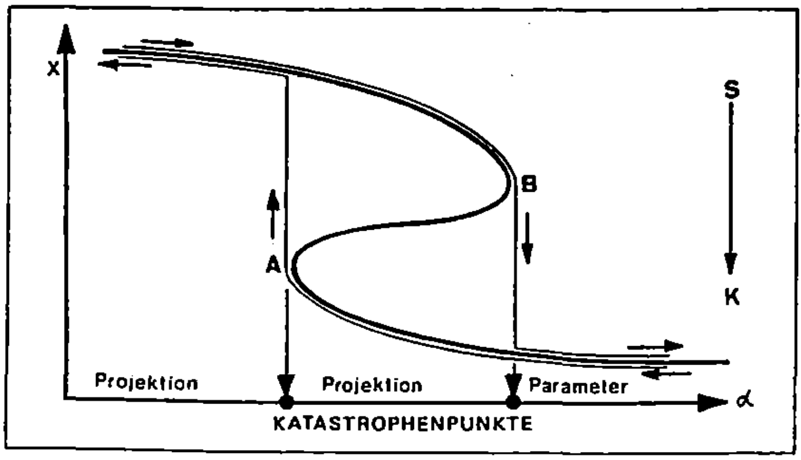


Fig. 1a: Fold Catastrophe

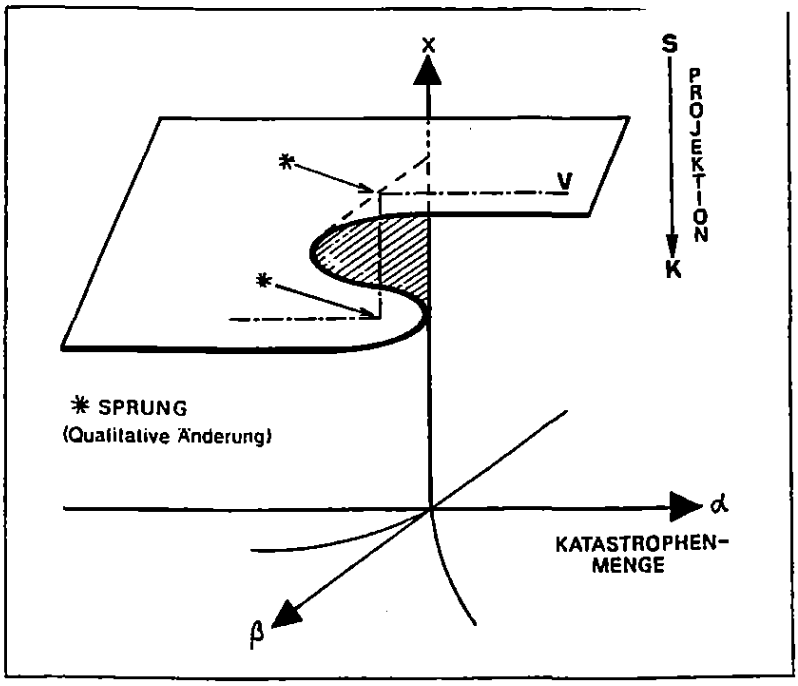


Fig 1b: Cusp Catastrophe

Hermann Haken, the founder of Synergetics, baptized his theory several years ago „General Theory of Emergence“. The essential object of such a theory is structural change in contrast to continuous transformation.

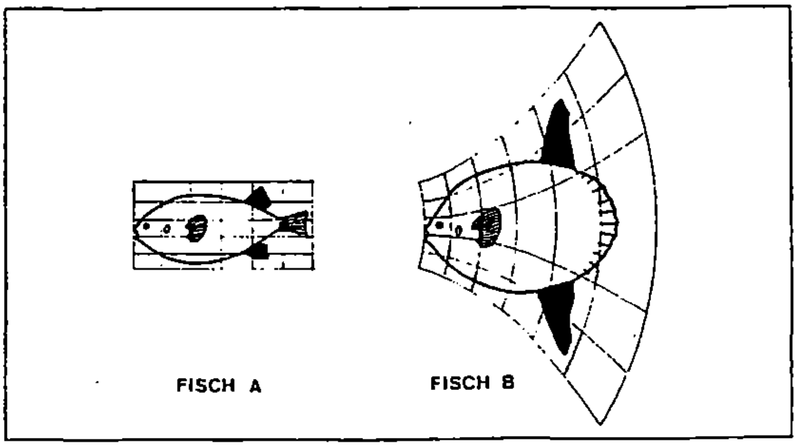


Fig. 2: Homeomorphic transformation

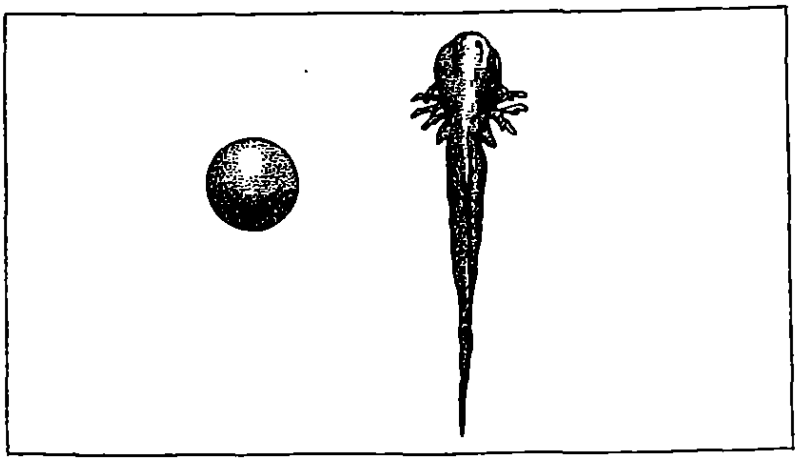


Fig. 3a: Non-homeomorphic transformation

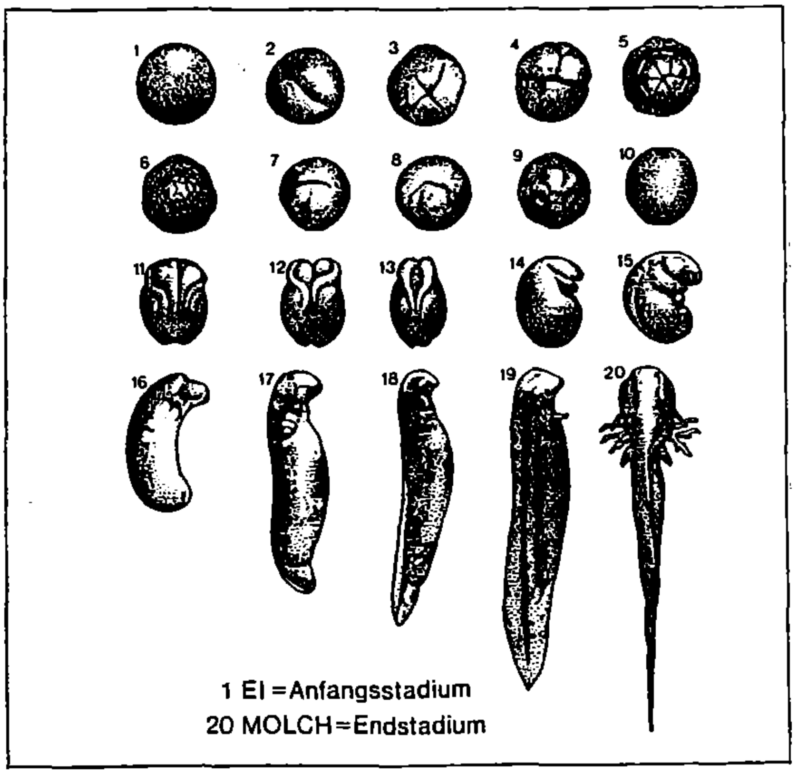


Fig. 3b: Non-homeomorphic transformation (detailed version)

The first homeomorphic transformation (fig. 2) stays within an *equivalence class*, the second (figs. 3a and 3b) transcends this class. What does it mean? HAKEN reminds us of the fact that a homeomorphism (a one to one and continuous mapping from point sets to point sets) establishes an equivalence relation (for a set of elements A, B, C, ...and a relation ~ one has reflexivity [A ~ A], symmetry [if A ~ B, then B ~ A], and transitivity [if A ~ B and B ~ C, then A ~ C]. In connection with this fact the theorem holds that every equivalence relation gives a definite disjunct classification of the elements (partition) and vice versa, for if one has two elements A and B belonging to different equivalence classes and assumes that an element C joins both classes, one comes because of C ~ A and C ~ B to A ~ B, which contradicts the assumption of the disjointness of A and B.

Therefore we arrive at a partition of the space of *dynamical systems* (fig. 4) - which are classified into equivalence classes, namely homeomorphisms.

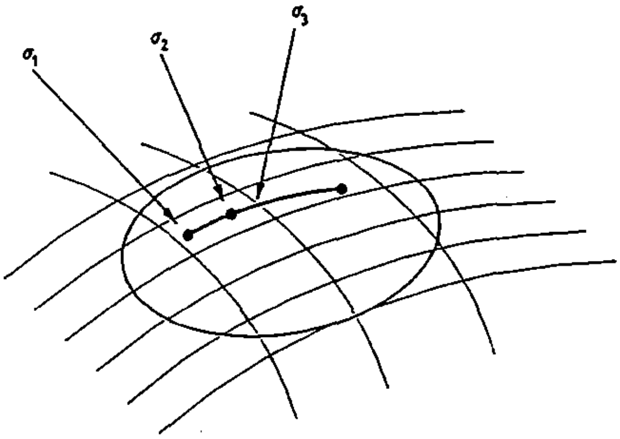


Fig. 4: Non-homeomorphic transformations in the space of dynamical systems

Following a path through this space one stays for a while within some equivalence class but suddenly it is left because the trajectory crosses the border and invades another class, a difference which already could have been seen in figs. 3a and 3b. The continuous transformation shown in fig. 3a is characterized by its properties of staying in the same equivalence class respectively by performing the homeomorphic mapping from A to B and thus sticking to the *same topology*. In the contrary fig. 3b shows a crossing from one equivalence class to another. So it is a rather graphic example of an antihomeomorphic mapping (and therefore inevitably a change of topology) compared with fig. 4 which is a slightly more abstract one of that kind.
Earlier we represented such an antihomeomorphic mapping as follows.

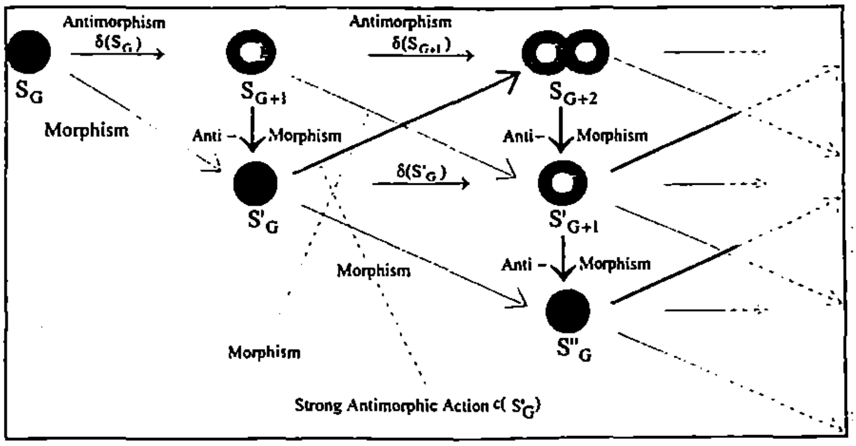


Fig. 5: Increase of the topological genus by an iteration of non-homeomorphic mappings

As our standard example we took a topology change of a sphere to a torus (see fig. 6). One can easily see that fig. 6 simply displays an accentuation of the first step in fig. 5.

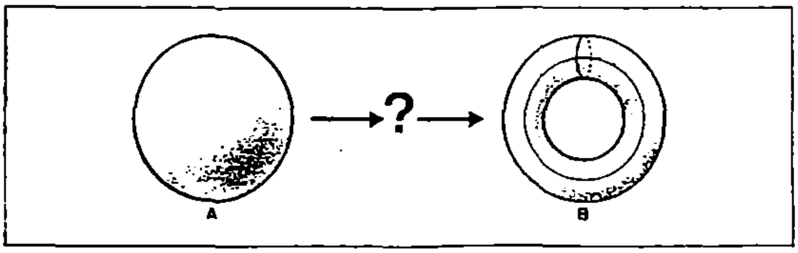


Fig.6: Elementary heterotopic transformation[i]
by a non-homeomorphic mapping

Here we have again the bifurcation of the theoretical methods of 'external approximation' and 'analytical solution' mentioned above. Synergetics as all the other theories of self-organization or dynamical systems comply in the end with 'external approximation' in tackling the problem of including discontinuity in a mathematical model. The theoreticians pin their hopes on *non-linearity* of differential equations in connection with specific selection rules of parameters to grasp *discontinuity*. (The transition of ordinary lamp-light to laser-light exemplifies this thinking.)

Yet we resolved to tread a new - or rather old - path firstly revealed by GOTTFRIED WILHELM LEIBNIZ, who is well known for the *Principle of Continuity*, which he has conceived. But less surprisingly as one might think LEIBNIZ with the same right could be named as well as the founder of the *Principle of Discontinuity*.

This *Principle of Discontinuity* follows (by means of LEIBNIZ) in three steps:

- At first we take from the Principle of Continuity the notion of an infinite partition of matter.
- Then secondly follows from that Principle of Continuity - together with what could be called a *postulate of an excluded state of rest* - a *postulate of infinitely small motion* - (by LEIBNIZ dubbed: *conatus*).
- And then finally comes the intrinsically discontinuous nature of that *conatus* by its structural equivalence with the *contiguous nature of space*, which LEIBNIZ fervently pointed out in his *Theoria motus abstracti*.

The Principle of (Dis)Continuity

I **The Principle of Continuity:**

„Le moindre corpuscule est actuellement subdivisé à l'infini, et contient un monde de nouvelles creatures, dont l'Univers manqueroit, si ce corpuscule étoit un Atome, c'est à dire un corps tout d'une piece sans subdivison."
(Streitschriften zwischen Leibniz und Clarke, Postscribt zu Leibniz' viertem Schreiben, in: G.W.Leibniz, Die Philosophischen Schriften (ed. G.J.Gerhardt), Bd.VII, p 377/378)

II **From *Specimen dynamicum*, pars II:**

"... nulla est unquam quies vera in corporibus, nec a quiete aliud nasci potest quam quies;" (G.W.Leibniz, specimen dynamicum, pars II, in: Mathematische Schriften, (Hrsg. C.J.Gerhardt), Bd.VI p 252) „... there is never actual rest in material bodies, and nothing else can emerge from a state of rest than just merely rest [i.e. from a state of rest merely nothing can emerge, therefore it must be conceived as impossible that there ever will be or ever had been such a state of rest, D.K.]."

III **From *Theoria motus abstracti (Fundamenta praedemonstrabilia)*:**

***Nullum est minimum in spatio aut corpore,* seu cujus magnitudo vel pars sit nulla...**

Dantur indivisibilia seu inextensa...

***Punctum* non *est, cujus* pars nulla est...sed cujus *extensio nulla est,* seu cujus partes sunt indistantes...**

Conatus est ad motum, ut punctum ad spatium, seu ut unum ad infinitum, est enim initium finisque motus.

(Die Philosophischen Schriften von G. W. Leibniz, ed. Gerhardt, vol. IV, pp 228/229)

And now - at last - an English version, which nicely brings together some aspects of the previously said:
"The law of continuity takes "rest as infinitely small motion (that is, as equivalent to a particular instance of its own contradictory), coincidence as infinitely small distance, equality as the limit of inequalities" (GM IV 93: L544)" (The Cambridge Companion to Leibniz, ed. N.Jolley, Cambridge 95, p 186)

A slightly more modern version of this **Principle of (Dis)Continuity** then might go as follows:
In the realm of the infinitely small a morphism is the 'limit' of antimorphisms as equality is the limit of inequalities, and as rest is infinitely small motion. Thus the ground state of nature consists of infinitely small perpetual antimorphic motion or topology change, for it has to be considered as a change in a strictly inextensive state.

Fig.7: Leibniz' Principle of (Dis)Continuity

After having spent these thoughts on discontinuous transformations in general, or antimorphisms, as we named them earlier, we now want to come to a special case of these transformations.

III ... to Emergent Complexity

In previous papers we always had our focus on the entanglement of emergence and complexity, namely on the emergence of complexity.
And as well we tried to make this concept of *emergent complexity* fit into a consistent mathematical framework, namely that of topology and category theory.
The most elementary level of a required structural enrichment which is needed to achieve a proper model of emergent complexity can be seen in fig. 5 which shows what we had called *topological enrichment*, namely the increase of the topological genus by iterated heterotopic transformations (formerly known as "strong antimorphic actions").
Obviously such an elementary level of the required structural enrichment, namely the multiplication of holes won't be enough to characterize even the most basic levels of natural complexity. Therefore we introduced the respective concepts of linking and interlocking to represent higher level emergent complexity (fig. 8).

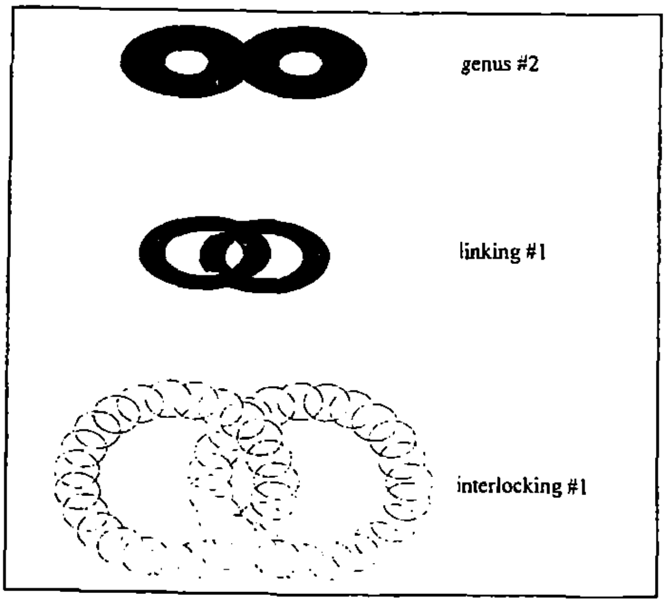


Fig. 8: Elementary levels of complexity characterized by genus, linking and interlocking

Here one should keep in mind that linking and interlocking work by exactly the same topological mechanism as the original increase of the topological genus, especially the initial emergence of a first toroid up from a respective prestruc-

ture. Thus in the following the emphasis will mostly be laid on different yet complementary aspects of that initial emergence.
As we already have mentioned the categorical aspect of heterotopic transformations has already been implicit in the topological enrichment (produced by the increase of the topological genus), as it can be seen in figs. 5, 6 and 9.
But there had been some criticism referring to that ominous initial emergence of a first toroid up from a respective prestructure, which is displayed on the very left hand of the first line of fig. 5.
That criticism was relating to the fact that by starting that course of topological enrichment there would have been a somewhat arbitrary act of "punching a hole in the initial globe or sphere".
Frankly spoken, we then didn't take this criticism very seriously, - and actually we still do not so now. The main reason for that seeming stolidity was the fact that that criticized arbitrariness was not just ours. Rather can a similar arbitrariness already be found in M-theory. There the necessary initial heterotopic transformation from an infinitesimal globe or sphere to a respective loop or torus had been described as the result of a quantum topological uncertainty. In string theory (version M) this characteristic topology change plays a prominent rôle.

Fig. 9: An illustration of a non-homeomorphic mapping from String Theory[ii], showing a mapping from a torus to a sphere.

Obviously that solution of the problem how to start the process of topological enrichment is itself not a genuine mathematical but rather a physical one. Nevertheless this would be perfectly in line with our original claim to give a formal description of the emergence of complexity, i.e. of the transition from one level of complexity to the next level of higher complexity. This is evidently something different from the question of how a *first* level of complexity initially came into existence.

IV Categorification

But to be honest our true objective had in fact been to give a mathematically that is category theoretically self-consistent representation of those enfolding heterotopic transformations resp. non-topological mappings *up from their starting point.*

And thanks to n-categories resp. to the operation of *categorification* we eventually found a model which lets us redeem our original pretension to give a natural morphism from a sphere to a torus, i.e. this time strictly by categorical means, as will become obvious fig. 13. As a preparation for this the following overview of various categorificational approaches might be helpful.

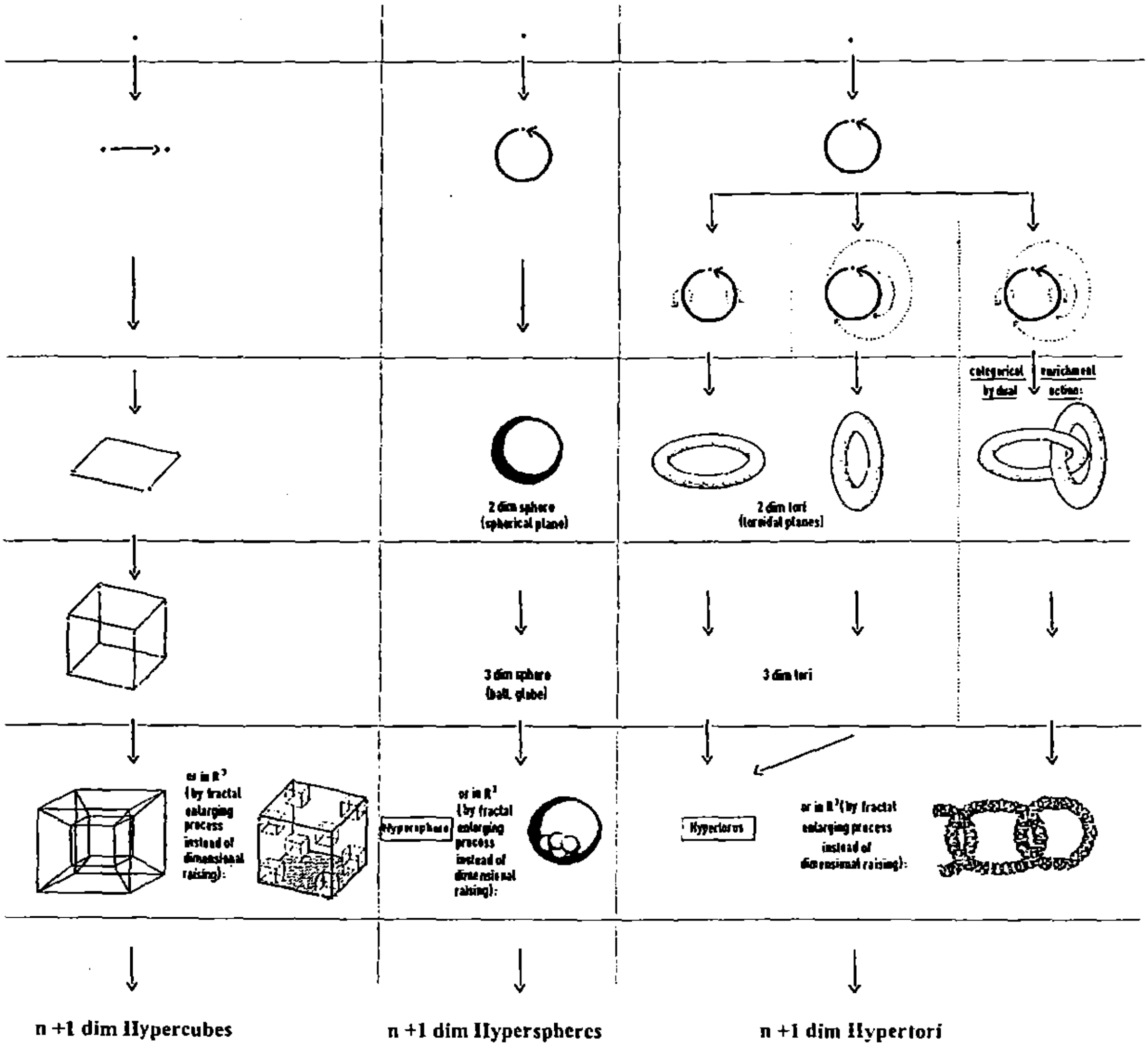
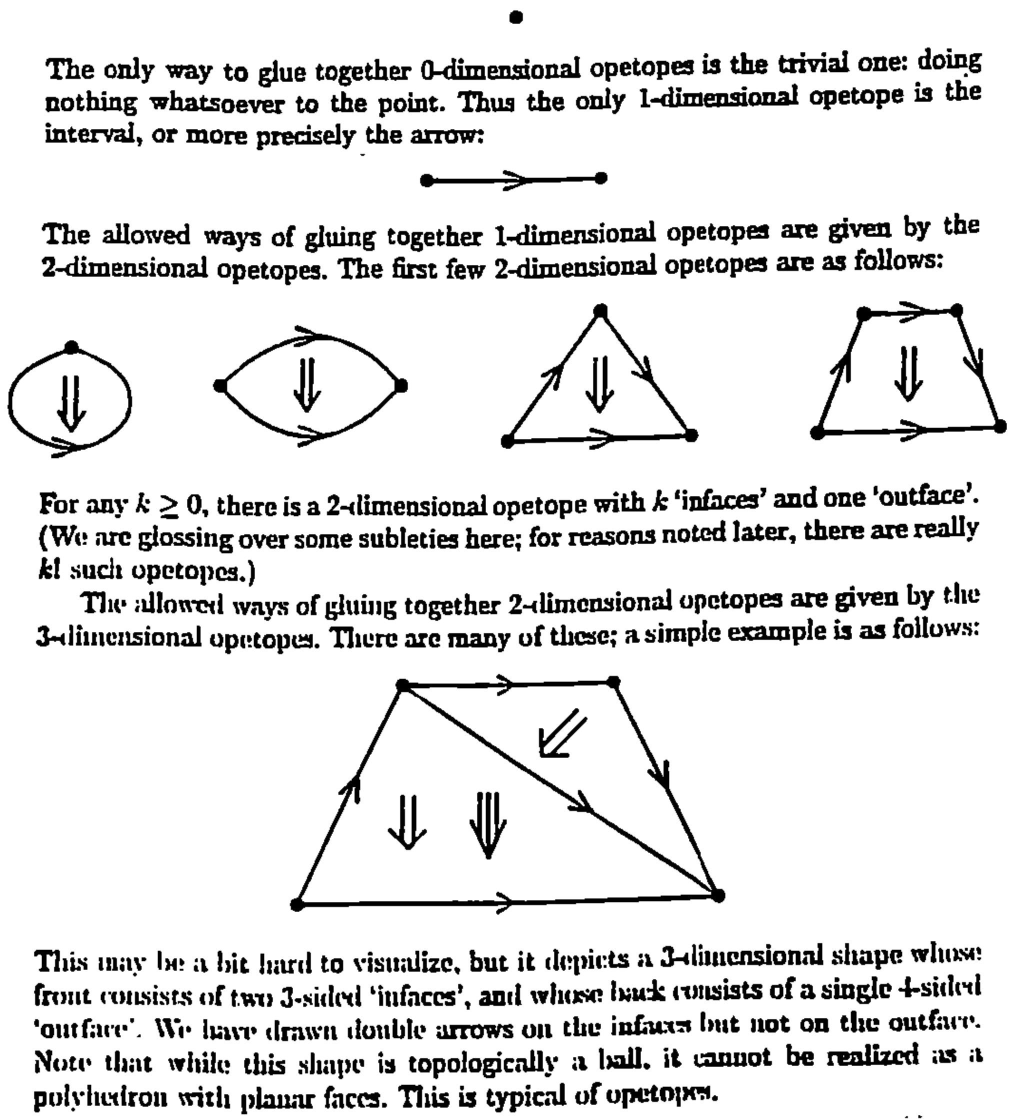


Fig. 10: The cubical, spherical and some toroidal ways
of increasing n-categorical dimensions

The rationale behind this attempt of exploiting categorification for the purpose of complexification goes as follows:
Emergence (of complexity) is in an appropriately abstractive and therefore sufficiently general way describable by means of dimension increasing morphisms,

i.e. n+1 morphisms. The best way to represent such a process (or class of processes) consists in commutative diagrams showing the morphisms.



The only way to glue together 0-dimensional opetopes is the trivial one: doing nothing whatsoever to the point. Thus the only 1-dimensional opetope is the interval, or more precisely the arrow:

The allowed ways of gluing together 1-dimensional opetopes are given by the 2-dimensional opetopes. The first few 2-dimensional opetopes are as follows:

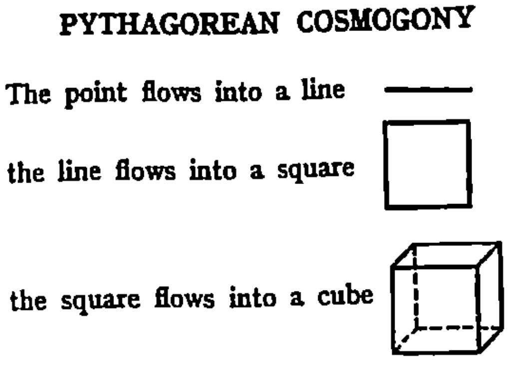
For any $k \geq 0$, there is a 2-dimensional opetope with k 'infaces' and one 'outface'. (We are glossing over some subleties here; for reasons noted later, there are really $k!$ such opetopes.)

The allowed ways of gluing together 2-dimensional opetopes are given by the 3-dimensional opetopes. There are many of these; a simple example is as follows:

This may be a bit hard to visualize, but it depicts a 3-dimensional shape whose front consists of two 3-sided 'infaces', and whose back consists of a single 4-sided 'outface'. We have drawn double arrows on the infaces but not on the outface. Note that while this shape is topologically a ball, it cannot be realized as a polyhedron with planar faces. This is typical of opetopes.

Fig. 11: The opetopic approach (by J.Baez)[iii]

Yet instead of "gluing" the respective figures together we rather walk along Pythagorean paths in using the 'fluxion method'



PYTHAGOREAN COSMOGONY

The point flows into a line

the line flows into a square

the square flows into a cube

Fig. 12: The cubical approach (Plato's version)[iv]

to let emerge the figures of increasing complexity in the cubical way (see the very left column of fig. 10).
Arranging the respective figures in commutative diagrams

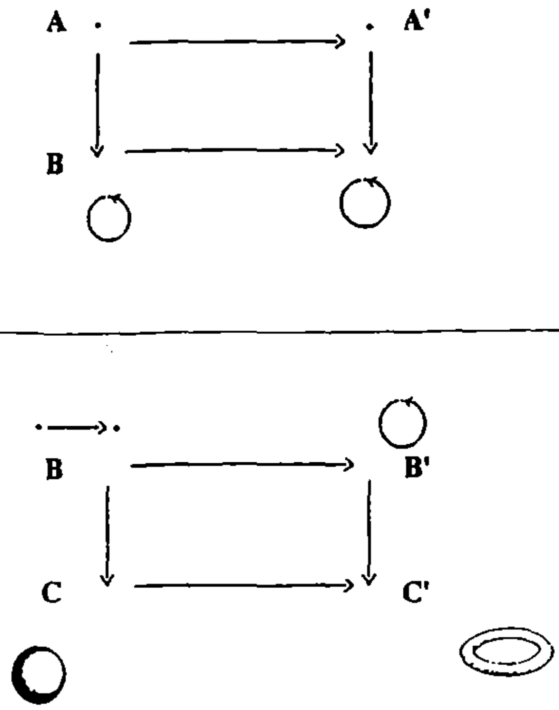


Fig. 13: Heterotopic mapping categorified

means to organize an increasing order of categories - n, n+1, ... - in the framework of weak n-category theory which is about the enrichment of mathematical structure (in contrast to strict n-category theory which refers to the increase of categoricality). Let us quote from Baez:

> " An *n*-category is an algebraic structure of a collection of 'objects', a collection of 'morphisms' between objects, a collection of '2-morphisms' between morphisms, and so up to *n*-morphisms, with various reasonable ways of composing *j*-morphisms. A 0-category is just a set, while a 1-category is just a category."[v]

> "The classic example of a 2-category is Cat, which has categories as objects, functors as morphisms, and natural transformations as 2-morphisms."[vi]

And thus that certain arbitrariness of starting the course of non-topological mappings, (previously called 'antimorphisms' lately instead: 'heterotopic mappings') as shown in fig. 5, incidentally happened to be eliminated - by strictly categorical means, see the commutative diagram at the bottom of fig. 13.

Baezed on that n-categorical ansatz let us now *fingere una hypthesis* about the interplay of categorification on the one hand and the emergent process of building nature's hierarchy on the other.

HYPOTHESIS

There is an equivalence of categorification and structural enrichment in nature, more precisely there exists a correspondence between the n-categorical and the natural levels, respectively. In a sense emergence *is* categorification and reduction *is* decategorification.

Especially we can represent the fundamental (dynamical) discontinuity of the emergent process in an inherent way because we have *difference* and *sameness* of morphisms *simultaneously*. The arrows n fig. 14 pointing from A to B, B to C etc., and from A' to B', B' to C' etc. are morphisms between objects, the arrows from A to A', B to B', C to C' etc. are morphisms of different n-categories, i.e. functors. But - and this is what we wanted to show you in the first place - these functors are in fact antimorphisms as well. That is they are antimorphisms quite in a sense we had antimorphisms in the realm of the infinitely small of Leibniz.
The antimorphism from A to A' can even actually be seen as an authentically Leibnizian one. And the higher dimensional ones like B to B', C to C' and so on now just show you in some enlargement the characteristic topology change - not easily been conceived in that antimorphic automorphism of an inextensive point to 'itself'.
But by means of these antimorphisms we can also add to the central feature of n-category theory, the art of refining the concept of sameness. Now in a rather stratified dimension namely by analyzing the relative equivalence of topologically different structures. Here this relative equivalence is one with respect to their levels of complexity. Let us hear the words of Baez and Dolan about the philosophical significance of that kind of subtlety:

> "One philosophical reason for categorification is that it refines our concept of 'sameness' by allowing us to distinguish between isomorphism and equality. In a set, two elements are either the same or different, In a category, two objects can be 'the same in a way' while still being different. In other words, they can be isomorphic but not equal. Even more importantly, two objects can be the same in more than one way, since there can be differ-

> ent isomorphisms between them. This gives rise to the notion of the 'symmetry group' of an object: its group of automorphisms."[vii]

In the higher dimensional cases the introduction of antimorphisms may add to this endeavor of refining the concept of sameness in a new way. It yields - so to speak - in a level perspective the transformation of apparently topologically different structures.

And so it brings to light that the morphisms which in a categorificational manner (by increasing the dimension) generated these topologically different structures can themselves seen as being 'isomorphic but not equal'. Or to put it in an even more simplified manner: it brings to light that in the course of emergence isomorphic processes didn't lead to same (or even 'identical') structures.

---

[i] The authors owe the term "heterotopic" as a characterisation of that "genus increasing" transformation to Hans van den Berg. Our earlier expressions of "antimorphism" resp. of "antimorphic action" seemed to have been somewhat deviating from the common mathematical terminology.

[ii] Cf. Brian Greene: Das elegante Universum, Berlin 2000 (= The Elegant Universe, german edition), p. 377. Obviously the kind of non-homeomorphic transformation shown in fig. 9 goes just in the other direction as our initial heterotopic transformation. In String Theorythis transformation has to be understood as a lower dimensional analogy.

[iii] J. Baez, An Introduction to n-Categories, in: 7th Conference on Category Theory and Computer Science, ed. E. Moggi and G. Rosolini, Berlin-New York 1997

[iv] Cf. F.M.Cornford, Plato and Parmenides, London 1958, p. 12

[v] J. Baez, An Introduction to n-Categories

[vi] J. Baez and J. Dolan, Categorification, math.QA/9802029 5 Feb 1998, p.8

[vii] Ibid., p.7

Sameness and Oppositeness in Quantum Information

O J E Maroney
Theoretical Physics Research Unit
Birkbeck College, University of London
o.maroney@physics.bbk.ac.uk

December 2000

Abstract

Surprising things happen when information processing is generalised to quantum, rather than classical systems. This paper reviews some recent results about the nature of information stored in quantum bits (qbits). An 'informational equivalence' between 'sameness' and 'oppositeness' in classical information fails to hold when those notions are generalised to quantum information. This has consequences for the sharing and flow of information within quantum networks. Draft v2.0

1 Informational Equivalence of Duplicate Classical Bits

Alice and Bob are typical quantum information theorists: they have well stocked laboratories, with state of the art measuring devices, computers and so forth, and have access to an inexhuastible supply of quantum and classical objects. They have three noise free communication channels - one of which allows then to send classical information, in bits, the second for quantum objects and the third is a telephone to talk to each other about what they're doing. When a message from either of the first two channels gets received, it is not 'read' immediately, but is stored in a box (or a memory circuit, or something like that). The receiver then has a number of options as to what kind of operation s/he can perform upon the received 'message', one of which is "open the box and have a look" (of course, if it's a quantum message, the receiver must also decide *how* to open the box). For the moment, we will assume Alice and Bob are dealing with purely classical bits - which have two states - 0 or 1. If the two bits are both in state 0 or both in state 1, we say they are in the *same* state. If one is 0 and the other is 1, they are in *opposite* states.

1.1 Same Bit Test

Alice sends Bob a single bit in a box, but does not let Bob know what state it is in. Bob has to return two bits, each in the same state as Alice's original bit. Bob can pass Alice's bit through whatever logic gates he requires, but he is not allowed to open the box and look at what the bit is (we don't want to make it too easy!).

This test is still quite easy, because Bob just needs one logic gate to solve this perfectly: the Controlled NOT (CNOT) gate.

A	B	C	D
0	0	0	0
0	1	0	1
1	0	1	1
1	1	1	0

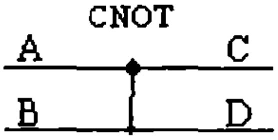


Figure 1: Controlled NOT Gate

With the input $B = 0$ this is often referred to as a FANOUT gate, and with $A = 1$ it is a NOT gate.

Bob takes Alices bit, and a second bit prepared in the state 0. Alice's bit is fed into the A-input, and Bob's second bit is fed into the B-input of the CNOT gate. The two output bits will now be in the same state as Alice's initial bit.

1.2 Opposite Bit Test

The second test requires Bob to return a second bit in the opposite state to Alice's original bit. Again, however, the solution is very simple. Bob requires only a third bit, now prepared in the state 1. After performing the same operation as for the Same Bit test, Bob simply sends the D-output into the B input of a second CNOT gate, with the third bit entering the A-input.

1.3 Informational equivalence

Although these tests seem trivial, they illustrate an important feature of information. Suppose Alice sends Bob a single bit, without specifying the state of the bit. Bob may have a physical system, but has acquired, as yet, no information (Bob has complete uncertainty about the state of the bit). If Alice sends Bob a second, unspecified bit, Bob has two physical systems, but still no information. However, now suppose Alice phones up Bob and tells Bob that, whatever state the first was in, the second was in the *same* state (we will assume Alice is not lying!). How much information has Bob acquired? What if Alice said the second was in the *opposite* state?

The answer is that Bob acquires 1 bit of information. By passing the two bits through the CNOT gate, A is in the initial state, but D is always in state 0. Although Bob has no information about the state Alice sent, he has an absolute certainty about the state of D. For the second case, Bob performs a NOT upon B, before passing it through the CNOT. By informing Bob of the correlation between A and B, Alice has reduced the number of the

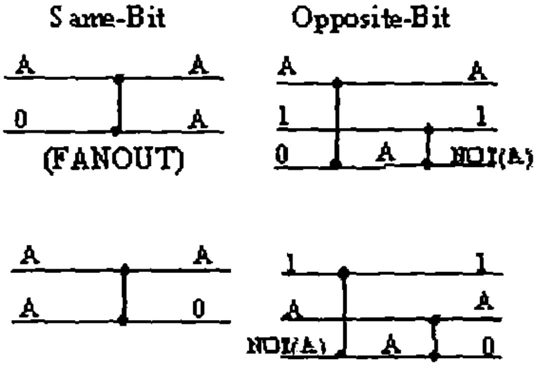


Figure 2: Informational Equivalence

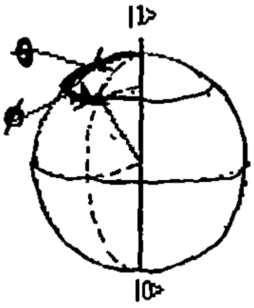


Figure 3: The Bloch Sphere

possible states (or Bob's uncertainty) by 1 bit. Bob converts this correlation information into a certainty about the state of one of the bits ('using up' the correlation). Bob can switch between these states but always has exactly 1 bit of information about the joint system. The three states are "informationally equivalent". Creating duplicate bits, or flipping those duplicate bits, cannot be used to increase or decrease the information available. This is why Bob found the tests so easy: Alice was not requiring Bob to supply any more information back to her, than she had sent to him in the first place.

2 Quantum Bits

Now Alice and Bob try to do the same thing, but with quantum bits. A quantum bit (or qbit) is a two dimensional Hilbert space, with basis states $|0>$ and $|1>$, The general state of qbit is:

$$\Psi = e^{i\frac{\phi}{2}} \cos\left(\frac{\theta}{2}\right)|1> + e^{-i\frac{\phi}{2}} \sin\left(\frac{\theta}{2}\right)|0>$$

often shortened to

$$\Psi = \alpha|1> + \beta|0>$$

The qbit may be represented by a point on a sphere of unit magnitude (Bloch sphere). The inner product of two qbits is

$$<\Psi|\Psi'> = \alpha^*\alpha' + \beta^*\beta'$$

For any given qbit state, there is exactly one other qbit state with which it has a zero inner product. This other state corresponds to the point in exactly the opposite direction on the Bloch sphere. $|1>$ and $|0>$ are 'opposite' to each other, and as the choice of a basis for $|1>$ and $|0>$ is arbitrary, this may be taken as the quantum generalisation of 'oppositeness' of bits. A qbit 'opposite' to $|n>$ will be represented by $|-n>$. For the inner product to be 1, the two qbits must at the same point on the Bloch sphere, and will be in the 'same' quantum state. If there is an ensemble of qbits, prepared in different states, the density matrix corresponds to a point lying inside the sphere.

Now we need the logical operations Alice and Bob can perform upon qbits, and it turns out that a single gate is sufficient to represent all of them - the quantum controlled-not[1].

$$\begin{aligned} C(a,b,c) &= |0><0|_A \otimes I_B + |1><1|_A \otimes U(a,b,c)_B \\ U(a,b,c) &= \begin{pmatrix} e^{ia}\cos b & -e^{ic}\sin b \\ e^{-ic}\sin b & e^{-ia}\cos b \end{pmatrix} \end{aligned}$$

If we take the basis states of $|0>$ and $|1>$ as the 'classical' states 0 and 1, we can reconstruct classical logic gates from this[1]. In particular, the classical CNOT gate is equivalent to $C(0,\pi/2,\pi)$.

How much information is there in a single qbit? There is a lot more possibilities for the qbit, than for the bit. Surely, this gives Alice a much wider range of signals she can send Bob.

However Bob has a problem. Alice sends Bob a single, unknown qbit in a box. It could be pointing anywhere on the Bloch sphere. When receiving a classical bit, Bob can simply open the box, find out what it is, and gain one bit of information. However, for a qbit, Bob must make a measurement, along a particular axis. If he chooses the conventional axis, he gets the probabilities

$$\begin{aligned} \Pr(|0>) &= \sin^2(\frac{\theta}{2}) \\ \Pr(|1>) &= \cos^2(\frac{\theta}{2}) \end{aligned}$$

and afterwards the qbit is in state $|0>$ or $|1>$ respectively.

---

[1] In classical logic, there are three input gates which cannot be built from reversible two input gates. These three input gates can be constructed from two-input quantum gates that are not equivalent to two input classical gates!

Although a large number of identically prepared qbits will eventually yield up the value of θ, Bob doesn't get any information about ϕ and Alice could be sending qbits pointing in different directions. As the measurement described has only two outcomes, it turns out Bob can get, at most, one classical bit of information from the transmitted qbit. With an infinite amount of qbits, Bob can find θ exactly, but this is the same as the infinite amount of bits necessary to specify a continuous parameter. The average information conveyed by each qbit is still one bit.[2][3][4]

2.1 Same Qbit Test

This suggests that there is no difference, in information content, between the classical and quantum bits. However, if we look again at the trivial tests Alice set, we find Bob has a much harder task. We will judge Bob's success at playing the game by a FIDELITY test. If Bob is supposed to produce a qbit in the state $|u>= \alpha|1> +\beta|0>$, Alice will measure the state he actually produces in the $|u>, |-u>$ basis, and the fidelity is the probability of Bob's qbit passing the test. In order to test a given strategy, we average this over all the possible states Alice could have chosen.

Bob has a number of strategies.

2.1.1 Random Guesswork

Bob is feeling bored with this silly game. He throws Alice's bit away and sends her two bits, each of which he prepares in some random, but identical state, $|n'>= \alpha'|1> +\beta'|0>$. When Alice measures Bob's qbits, the probability of passing is $\cos^2\left(\frac{\theta}{2}\right)$ where θ is the angle between the Alice's and Bob's directions. The mean fidelity of each bit is $F = 1/2$ and the joint fidelity (the probability of getting both right) $F_J = 1/3$. While random guesswork is not a good solution it gives us a baseline to measure the success of other methods.

2.1.2 Measure and Copy

Bob measures the bit in some basis (for convenience, we use $|1>, |0>$), and sends back two bits in the direction the measurement gives. This gives a density matrix $\rho = |\alpha|^2 |11><11| + |\beta|^2 |00><00|$ and average fidelities $F = 2/3$, $F_J = 1/2$

2.1.3 FANOUT

Bob thinks the problem might be because he is opening the box to measure the qbit, and this disturbs the quantum system. So he uses in logic gates (FANOUT) to copy the qbit, without opening the box. This produced a perfect solution with classical bits. However, with qbits, the result is $F = 2/3$, $F_J = 1/2$ and is no better than "measure and copy"! What went wrong?

The answer is that FANOUT fails to copy the qbit - instead it creates an entangled state between the output bits:

$$FANOUT(\alpha|1>+\beta|0>)|0>=\alpha|1>|1>+\beta|0>|0>$$

As a density matrix this is

$$\rho = |\alpha|^2 |11><11| + |\beta|^2 |00><00| + \alpha^*\beta|00><11| + \alpha\beta^*|11><00|$$

with diagonal elements equivalent to the 'Measure and Copy' approach. In fact the FANOUT gate bears a lot in common with the process of measurement, and is sometimes referred to as a 'measurement' gate.

2.1.4 Quantum FANOUT and "no-cloning"

Can we build a quantum FANOUT? If we take an initial, unknown qbit, and a auxiliary system, prepared in a known state, does there exist any unitary operation of the form:

$$CLONE(|n>|Aux0>) = |n>|n>|Aux(n)>$$

where $|Aux0>$ is the initial auxiliary system, and $|Aux(n)>$ is an n dependant 'junk' output, which works for all values of n? The answer to this was answered in the negative by Wootters and Zurek[5]. The proof of the non-existence of $CLONE$ can be found from the unitary operation preserving the inner product between states

$$<n|n'><Aux0|Aux0>=(<n|n'>)^2<Aux(n)|Aux(n')>$$

However, the inner product of two states obeys the relation

$$|<i|j>| \leq 1$$

with equality holding only when $i = j$. The required relationship can only hold when either $n = n'$ or when $<n|n'>= 0$, but cannot hold for general values of n . An obvious case for $< n|n' >= 0$, is where $n = 1$, $n' = 0$. So classical information can be cloned (which is fortunate, as we already have a FANOUT gate that does this!)

However, it is possible to build imperfect cloning machines, that produce a fidelity better than simply 'measure and copy'.[6][7][8][9] An example of an optimal quantum cloning, in which the fidelity of the output is independant of the input state, is given by the following unitary operation:

$$\begin{aligned} |0>|00> &\rightarrow \sqrt{\frac{2}{3}}|000> + \sqrt{\frac{1}{6}}|011> + \sqrt{\frac{1}{6}}|101> \\ |1>|00> &\rightarrow \sqrt{\frac{2}{3}}|111> + \sqrt{\frac{1}{6}}|010> + \sqrt{\frac{1}{6}}|100> \end{aligned}$$

For a general input qbit of $|n>$ in the first position, this produces output qbits in the first and second positions of $\rho = 5/6|n >< n| + 1/6| - n >< -n|$. The third qbit is the 'junk' auxiliary output. The fidelty is $F = 5/6$, with a joint fidelity of $F_J = 2/3$

2.2 Opposite Qbit Test

What of Alice's second test? What if Bob has to produce two qbits, but in opposite directions?

Bob has the same strategies available to him. If he measures Alice's bit, and sends back opposite qbits $|10>$ or $|01>$, he gets the same fidelity of as 'measure and copy' for producing the same qbits. If he runs the qbit through a FANOUT and a NOT gate, he get the fidelity of 'FANOUT' for same qbits.

What if he uses an optimal cloner and NOT? Surprisingly, our joint fidelity is worse and our opposite bit is terrible!

$Fs = 5/6$, $Fo = 7/12$, $Fj = 2/3$

The reason for this failure is that our NOT gate is failing to work in the way we desired:

$$NOT(\alpha|1> + \beta|0>) = \beta|1> + \alpha|0>$$

It is easy to show that $NOT(|n>)$ is opposite to $|n>$ only if $n = 0$ or $n = 1$ (if the input qbit is part of classical logic). Bob needs an operation that performs:

$$\begin{aligned} OPP(|n>) &= |-n> \\ OPP(\alpha|1> + \beta|0>) &= \beta^*|1> - \alpha^*|0> \end{aligned}$$

Such an operation is not forbidden by the conservation of the inner product, as $< n|m > = < -n| - m >$. However if $OPP(|1>) = |0>$ and $OPP(|0>) = |1>$, then NOT is the only linear operation that satisfies these conditions (the 'no-spin-flip theorem').

Still, perhaps there is an imperfect OPP, in the same manner that there is an imperfect $CLONE$? It turns out the best that can be done is to build an anti-cloning machine, that takes an input qbit $|n>$ and attempts to make output qbits $|n>|-n>$, succeeding with fidelity $F = 2/3$, joint fidelity $F_J = 5/8$. [10][11][12]

Now this is particularly interesting - not only is Bob failing Alice's test, but trying to produce an *opposite* second bit is failing worse than a *same* second bit. Somehow it seems *opposite* is not informationally equivalent to *same*? Rather than examine proofs of the no-cloning and no-spinflipping theorems, let us look at the states we are trying to produce - the duplicate qbits.

3 Duplicate Quantum Bits and Informational Equivalence

Suppose Alice sends Bob a qbit prepared in a state unknown to Bob. Bob's uncertainty is at a maximum, as he has no information on the state of the bit. Now Alice sends Bob a second qbit, also unknown, but prepared in the same state as the first qbit. How much more information does Bob possess?

In the classical case, we saw that the answer was one bit. However, that was clearly related to the fact that we could run both bits through the FANOUT gate, and put one of them into a definite state. This is clearly not possible for qbits: if it were, we could simply reverse the process, and clone a qbit[15].

If we take an ensemble of qbits, in different states eg.

$$|a|^2|00><00| + |b|^2|11><11|$$

the information known about the ensemble is given (in bits), by

$$H = 1 + Tr(\rho log_2 \rho)$$

For an N-qbit system, the information known is

$$H_N = N + Tr(\rho log_2 \rho)$$

For a 1 qbit system, a general qbit is described by:

$$\rho = \begin{pmatrix} |\alpha|^2 & \alpha^*\beta \\ \alpha\beta^* & |\beta|^2 \end{pmatrix}$$

The 'general' qbit could have been a point anywhere on the Bloch sphere, with uniform probability. We average α and β uniformly over the Bloch sphere, and get a density matrix

$$\rho = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

In this case the information is zero, which represents the fact that we know nothing about where on the Bloch sphere a general qbit points. If we knew the state $|n>$ the qbit was prepared in, we can express the density matrix in the basis $|n>, |-n>$, where it becomes

$$\rho = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and the information is 1 bit, representing a complete knowledge of the state of the qbit.

For two qbits prepared in the same state $|n, n>$, the density matrix, when $|n>$ is integrated over the Bloch sphere, is:

$$\overline{\rho(|n,n>)} = \begin{pmatrix} 1/3 & 0 & 0 & 0 \\ 0 & 1/6 & 1/6 & 0 \\ 0 & 1/6 & 1/6 & 0 \\ 0 & 0 & 0 & 1/3 \end{pmatrix}$$

which can be expressed as

$$\overline{\rho(|n,n>)} = \frac{1}{3}(|u,u><u,u| + |u+><u+| + |-u,-u><-u,-u|)$$

where $|u>$ is an arbitrary point on the Bloch sphere and
$|u+>= \frac{1}{\sqrt{2}}(|u,-u> + |-u,u>)$.

This is not complete uncertainty. A total lack of knowledge is represented by

$$\rho_{\min} = \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}$$

which has information of zero. Instead, the knowledge that we have two bits that are in the same state gives us $H = (2 - \log 3) = 0.415$ bits. This is less than the 1 bit that correlated classical bits gives us, but more than complete ignorance. A classical correlation would have given us $H = (2 - \log(2))$ bits.

What if they are in opposite states - how much information have we gained? Two qbits in unknown, opposite states have a averaged density matix of

$$\overline{\rho(|n,-n>)} = \begin{pmatrix} 1/6 & 0 & 0 & 0 \\ 0 & 1/3 & -1/6 & 0 \\ 0 & -1/6 & 1/3 & 0 \\ 0 & 0 & 0 & 1/6 \end{pmatrix}$$

or

$$\begin{aligned} \overline{\rho(|n,-n>)} &= \frac{1}{6}(|u,u><u,u| + |u+><u+| + |-u,-u><-u,-u|) \\ &\quad + \frac{1}{2}(|u-><u-|) \\ &= \frac{1}{2}\overline{\rho(|n,n>)} + \frac{1}{2}(|u-><u-|) \end{aligned}$$

with $|u-> = \frac{1}{\sqrt{2}}(|u,-u> - |-u,u>)$ This has $H = (2 - \frac{1}{2}\log 12) = 0.208$ bits, exactly half the information of the same states.

3.1 Why don't we have 1 bit of correlation information?

If we expand the pure states, in the conventional basis, we obtain:

$$\begin{aligned} \Psi(|n,n>) &= \alpha^2|11> + \sqrt{2}\alpha\beta\left(\frac{|01> + |10>}{\sqrt{2}}\right) + \beta^2|00> \\ \Psi(|n,-n>) &= \alpha\beta^*|11> + \left(\frac{|\alpha|^2 - |\beta|^2}{\sqrt{2}}\right)\left(\frac{|01> + |10>}{\sqrt{2}}\right) \\ &\quad + \left(\frac{|\alpha|^2 + |\beta|^2}{\sqrt{2}}\right)\left(\frac{|01> - |10>}{\sqrt{2}}\right) - \alpha^*\beta|00> \end{aligned}$$

When measured in a different basis to the preparation basis, the same-state qbits may yield opposite results, while the opposite-state qbits can give the same results!

This clearly is a property of non-orthogonality in the quantum measurement process - even if we are sure the states were prepared in same (or opposite states), we cannot be sure they will both pass/fail (or the opposite) if the measurement is in a different basis. The essential feature of this is the non-orthogonality of the states the qbits *may have been* prepared in. If we are told that the qbits are prepared in a *particular* basis, then we can simply switch our logic gates to operate on that basis, and all our results of classical logic apply.

3.2 Why does oppositeness convey less information than sameness?

Although the separation between the two cases is guaranteed by the no-spin flip theorem, this does not explain why oppositeness conveys so much less information than sameness.

The wavefunctions and density matrices above, were expressed in the basis

$$\Phi_1 = |11> \qquad \Phi_2 = \tfrac{1}{\sqrt{2}}(|10> + |01>)$$
$$\Phi_3 = \tfrac{1}{\sqrt{2}}(|10> - |01>) \qquad \Phi_4 = |00>$$

In $\overline{\rho(|n,n>)}$, the probability of finding the state Φ_3 is zero, while for $\overline{\rho(|n,-n>)}$, it is one half.

Using the notation

$$|1>_X = \frac{|1>_Z+|0>_Z}{\sqrt{2}} \qquad |0>_X = \frac{|1>_Z-|0>_Z}{\sqrt{2}}$$
$$|1>_Y = \frac{|1>_Z+i|0>_Z}{\sqrt{2}} \qquad |0>_Y = \frac{i|1>_Z+|0>_Z}{\sqrt{2}}$$

we can construct the Φ basis from a superposition of opposite states:

$$\begin{aligned}
\Phi_1 &= \frac{(1-i)}{2}(|01>_Z - |10>_Z) + |10>_X - i|10>_Y \\
\Phi_2 &= \frac{1}{\sqrt{2}}(|10>_Z + |01>_Z) \\
\Phi_3 &= \frac{1}{\sqrt{2}}(|10>_Z - |01>_Z) \\
\Phi_4 &= \frac{(1+i)}{2}(|10>_Z - |01>_Z) - |10>_X - i|10>_Y
\end{aligned}$$

however, we can only construct 3 out of the 4 basis from same states:

$$\Phi_1 = |11>_Z$$

$$\begin{aligned}\Phi_2 &= \sqrt{2}|11>_X - \frac{1}{\sqrt{2}}(|11>_Z + |00>_Z) \\ \Phi_4 &= |00>_Z\end{aligned}$$

The space of the two qbits $SU(2) \times SU(2)$ has two invariant subspaces, under global rotations: a symmetric subspace, of dimension 3, and an antisymmetric subspace, of dimension 1. 'Sameness' means that the qbits can only be found within the symmetric subspace, and are evenly distributed throughout it. While in the classical case, the correlation information restricts the bits to a two dimensional subspace (and therefore represents an ignorance of log(2) bits) in the quantum case the restricted subspace is 3 dimensional, and the ignorance is log(3) bits. The opposite qbits are distributed throughout the entire state space - but they are likely to be found in the antisymmetric subspace, so 'oppositeness' does give some correlation information.

What is remarkable, however, is that these differences can only appear when one looks in an entangled basis, *even though the qbits themselves are always prepared in product states!* The measurement is of a joint property of the qbits - we cannot relabel the parts of the apparatus that apply only to the second particle, because there are no such parts. If we were to 'flip' the second qbit in the 'opposite state' expansion of the of the Φ basis, the result would no longer be an orthonormal basis, and does not correspond to a valid measurement (see also [13]). This strange phenomena - entangled state measurements yielding more information than any combination of local measurements, even when made on ensembles of product states - has been dubbed 'non-locality without entanglement'[14]

4 Information and reversible computing

The theory of reversible computation was developed following the discovery of Landauer's principle[16], that only logically irreversible operations implied an irretrievable loss of energy (prior to that, it was thought that each logical operation involved a dissipation of $kT\ln(2)$ per bit). The amount of lost energy is directly proportional to the Shannon measure of the information that is erased.

It is often defined as a requirement to 'do work' to perform the erasure. This is not strictly accurate. It requires an *investment* of $kT\ln(2)$ free energy, per bit of information that is stored. At any time in the computation, any bit that is in a known state can have this free energy recovered. A known state is one that is in a particular value, regardless of the choice of input state, (we may extend this to include always in the same state as an initial input state). When we examine a computational network, given the program, and the input state, we can recover all the free energy from the bits that are known. Other bits may be in determinate states, well defined functions of the input. It may be argued that these are, therefore, 'known' but, as these states are non-trivially dependant upon the input state (eg. (A OR NOT B) AND (C XOR D)), to extract the energy requires one to find the value of the bit from the input state ie. to recapitulate the calculation on a second system, which requires an investment of an equivalent amount of free energy - so no gain is made in terms of recoverable energy. The objective of reversible computing is to reduce the amount of the free energy invested into the calculation that cannot be recovered at the end without losing the results of the computation.

A reversible calculation may be defined as one which operates, upon an input state i and an auxiliary system, prepared in an initial state $Aux0$, to produce an output from the calculation $O(i)$, and some additional 'junk' information $Aux(i)$:

$$F : (i, Aux0) \rightarrow (O(i), Aux(i))$$

in such a manner that there exists a complementary calculation:

$$F' : (O(i), Aux(i)) \rightarrow (i, Aux0)$$

The existence of the 'junk' information corresponds to a history of the intervening steps in the computation, so allowing the original input to be reconstructed. A computation that did not keep such a history, would be irreversible, and would have lost information on the way. The information lost would correspond to an amount of free energy invested into the system that could not be recovered.

However, $Aux(i)$ is not generally known, being non-trivially dependant upon the input, i, and so represents free energy that cannot be recovered. A general procedure for discovering the complementary calculation F' can be given like this: take all the logical operations performed in F, and reverse their operation and order. As long as all the logical operations in F are reversible logic gates, this is possible. It is known that the reversible Fredkin-Toffoli gates are capable of providing all classical logical operations. So it is always possible to make a computation reversible. However, this is not immediately very useful: although we could recover the energy by reversing the computation, we lose the output $O(i)$ in doing so.

Bennett[17][18] showed that a better solution was to find a different reverse calculation F''

$$F'' : (O(i), Aux(i), AuxO) \rightarrow (i, Aux0, O(i))$$

The only additional unknown information is $O(i)$, which is simply the output we desired (or extra information we needed to know). A general procedure for F'', is: copy $O(i)$ into a futher auxiliary system $AuxO$ by means of a FANOUT gate, then run F'on the original system. This has also been shown to be the optimal procedure[19][20] for F''. We call such a calculation, G, TIDY. All classical reversible computations are TIDY.

Straight away, we should notice a problem! The universal FANOUT gate does not exist for a quantum computation.

Clearly, in the case where the output states from a quantum computer are in a known orthogonal set, then the quantum computation can be made tidy. In fact, for other reasons, having orthogonal output states was initially taken as a requirement on a quantum computer, as it was deemed necessary for reading out the output. This was suggestive not of a general quantum computation, but of limited quantum algorithmic boxes: each connected by classical communication. However, developments in quantum information theory have suggested that distributed quantum information may be desirable - in particular, a more general conception of quantum computation may be required which takes inputs from different sources, and/or at different times. In Figure 5 we see an example of this - Alice performs some quantum computation, and stores the result of it in a 'quantum data warehouse'. At some later time, Bob takes part of these results as an input into his own computation. We are going to take our definition of a quantum computation as: [2]

$$U_C(|i> |Aux0>) \rightarrow |O(i)> |Aux(i)>$$

so that the ouput is always in a separable state (in other words, we regard the 'output' of the computation as the subsection of the Hilbert space that is

---

[2] There is further complication when entanglement enters the problem. When part of an entangled state is transmitted, the loss of free energy is always greater than the entropy of the reduced density matrix. The minimum loss of free energy requires knowledge of an accurate representation of the resulting density matrix - which may not be possible without explicitly calculating the output states.

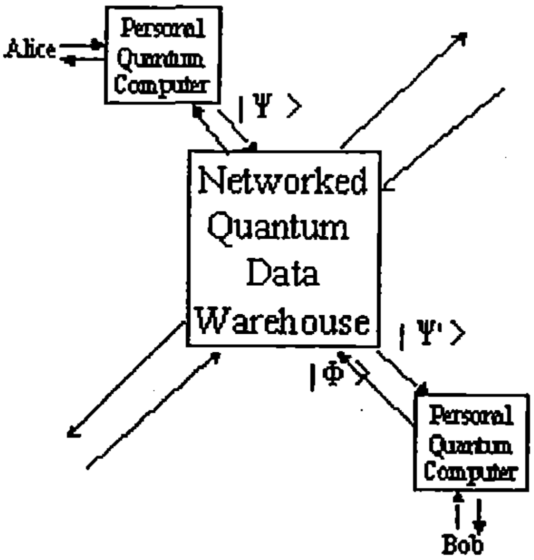


Figure 4: Distributed Quantum Computing

interesting, and the 'auxiliary' as everything that is uninteresting. If the 'output' were entangled with the 'auxiliary' space, then there would be additional information relevant to the 'output', contained in the super-correlations between 'output' and 'auxiliary' spaces). As any quantum computation must be performed by a unitary operation, all quantum computers must be reversible. But are they TIDY?

If this model of computation is classical, then each time data is sent to the central database, the local user can FANOUT the data before sending it, and tidy up their computer as they go along. The only energy commitment is: total input, plus stored data. The difference between connected classical algorithmic boxes and a single classical computation is a trivial distinction, as the computation may be tidied along the way.

Considering a general quantum operation, unitarity requires that the inner products between different input states and between the corresponding output states is unchanged by the computation. Reversibility must always hold.

$$\begin{aligned} REVERSIBLE &: \ <i|j><Aux0|Aux0>= \\ & \quad <O(i)|O(j)><Aux(i)|Aux(j)> \\ TIDY &: \ <i|j><Aux0|Aux0><AuxO|AuxO>= \end{aligned}$$

$$< i|j >< O(i)|O(j) >< Aux0|Aux0 >$$

We can eliminate $< Aux0|Aux0 >= 1$ and $< AuxO|AuxO >= 1$, leaving only three cases.

4.1 Orthogonal Outputs

The output states are orthogonal set:

$$< O(i)|O(j) >= \delta_{ij}$$

Reversibility *requires* the input states to be an orthogonal set $< i|j >= 0$, and the TIDY condition will hold. This is not too surprising, as an orthogonal set of outputs *can* be cloned, and so can be tidied using Bennett's procedure.

4.2 Orthogonal Inputs

The input states are orthogonal set $< i|j >= \delta_{ij}$, but the output states are not. To satisfy unitarity, the *auxiliary* output states must be orthogonal.

$$< Aux(i)|Aux(j) >= \delta_{ij}$$

There is a unitary operator for tidying the computation, without losing the output. However, this tidying computation is not Bennett's procedure. If we cloned the auxiliary output, and run the reverse operation, we would lose the output, and be left with the 'junk'! Whether there is an equivalent general procedure for obtaining F'' is not known.

One obvious method is to examine the resulting auxiliary output states, construct a unitary operator from

$$U_G|Aux\,(i)\,,O\,(i) >= |Aux0,O\,(i) >$$

and decompose U_G into a quantum logic circuit. However, it is not clear whether the operator can be constructed without explicitly computing each of the auxiliary output states - which may entail running the computation itself, for each input, and measuring the auxiliary output basis. Alternatively, examine the form of the auxiliary output (eg. (A OR NOT B) AND (C XOR D))) and devise a logic circuit that reconstructs the input state from this. This simply restates the problem: although some such circuit (or U_G) must exist, is there a general procedure for efficiently constructing it from only a knowledge of U_C?

4.3 Non-orthogonal Inputs

The input states are a non-orthogonal set. This corresponds to Bob's position in the quantum distribution network of Figure 5.

If we look at the requirements for a tidy computation, this leads to:

$$< O(i)|O(j) >= 1$$

The output is always the same, regardless of the input! Obviously for a computation to be meaningful, at least some of the output states must depend in some way upon the particular input state. So there does not exist *any* non-trivial ($|O(i) > \; \neq |O(j) >$) computations of the form

$$G : |i > |Aux0 > |AuxO > \rightarrow |i > |Aux0 > |O(i) >$$

for which $< i|j > \; \neq \delta_{ij}$. [3]

It should be clear: this does NOT mean useful, reversible quantum computations of the form

$$F : |i > |Aux0 > |- > |Aux(i) > |O(i) >$$

do not exist when $< i|j > \; \neq \delta_{ij}$ - simply that such computations cannot be 'tidy'. For such computations, not only is the free energy used to store the auxiliary output unrecoverable, but also the input state cannot be recovered, except through losing the output. For our distributed network, this means that not only can Bob not 'tidy' his computation, but he cannot restore Alice's data to the database.

5 Summary

We have examined the notion of 'sameness' and 'oppositeness' when applied to quantum information and found that the 'informational equivalence' of these in the classical case no longer hold. Quantum information cannot be copied or duplicated, in the manner of classical information.

This has a surprising consequence for computation. The flow of information in a classical computation can be broken down into separate algorithms, with these algorithms passing classical information between them. Such algorithms can be reversibly, and tidily, implemented. If the overall calculation

---

[3] It is interesting to note that the 'no-cloning' theorem is a special case of this.

requires input data in separate places and times, it can easily be broken down into separate algorithms at each place and time, with classical communications between them. This is only because such classical information can be duplicated in an 'informationally equivalent' manner.

Existing quantum algorithms have been designed on the basis of replacing similar classical algorithms. They therefore take a set of classical inputs, at one place and time, and produce a set of classical outputs, and so can be implemented in a tidy manner. However, each quantum algorithm itself cannot be broken down into sub-algorithms.

A more generalised conception of the flow of information in a quantum system appears necessary. Information enters and is shared at separate times and places, and cannot necessarily be processed by tidy sub-algorithms, as the information exchanged is not necessarily classical in nature. Even where a tidying procedure can exist, it is not clear that a general and/or efficient program for implementing this procedure is available.

"Oppositeness" and "Sameness" are well defined, conceptually simple, relationships between qbits, yet there are no physical systems that can implement these as operations such as OPP and $CLONE$. We must therefore be very careful before assuming which logical ideas can still be relied upon when trying to understand the nature of information in quantum processess.

References

[1] DiVincenzo DP, cond-mat/9407022

[2] Schumacher B, Phys Rev A, Vol 51 No 4, 2738-2747 (1995)

[3] Jozsa R, Shumacher B, J Mod Optics, Vol 41 No 12, 2343-2349 (1994)

[4] Holevo AS, Prob Inf Trans, Vol 9, 110, 177 (1973)

[5] Wootters WK, Zurek WH, Nature Vol 299, 802-803 (1982)

[6] Buzek V, Hillery M, Phys Rev A, Vol 54 No 3, 1844-1852 (1996)

[7] Gisin N, Massar S, Phys Rev Lett, Vol 79 No 11, 2153-2156 (1997)

[8] Buzek V, Braunstein SL, Hillery M, Bruss D, Phys Rev A, Vol 56 No 5, 3446-3452 (1997)

[9] Bruss D, DiVincenzo DP, EKert A, Fuchs CA, Macchiavello C, Smolin JA Phys Rev A, Vol 57 No 4, 2368-2378 (1998)

[10] Gisin N, Popescu S, Phys Rev Lett, Vol 83 No2, 432-435 (1999)

[11] Buzek V, Hillery M, Werner RF, Phys Rev A, Vol 60 No 4, R2626-R2629 (1999)

[12] Song DD, Hardy L, quant-ph/0001105

[13] Massar S, quant-ph/0004035

[14] Bennett CH, DiVincenzo DP, Fuchs CA, Mor T, Rains E, Shor PW, Smolin JA, Wootters WK, Phys Rev A, Vol 59, 1070 (1999)

[15] Pati AK, quant-ph/9911090

[16] Landauer R, IBM J Res Develop, Vol 5, 183-191 (1961)

[17] Bennett CH, IBM J Res Develop, Vol 17 525-532 (1973)

[18] Bennett CH, Int J Theor Phys, Vol 21, 905-940 (1982)

[19] Li M, Tromp J, Vitanyi P, Physica D, Vol 120 No 1-2,168-176 (1998)

[20] Li M, Vitanyi P, Proc Roy Soc Lon A, Vol 452, 769-789 (1996)

Discrete Coherent States For Spiral Galaxies

D.F.Roscoe, Dept of Applied Mathematics
Sheffield University, Sheffield S3 7RH
UK

April 9, 2001

Abstract

In Roscoe (1999a), it was described how the modelling of a small sample of optical rotation curves (ORCs) given by Rubin, Ford & Thonnard (1980) with the power-law $V_{rot} = AR^{\alpha}$, where where the parameters (A, α) vary between galaxies, raised the hypothesis that the parameter A (considered in the form $\ln A$) had a preference for certain discrete values. This specific hypothesis was tested in that paper against a sample of 900 spiral galaxy rotation curves measured by Mathewson, Ford & Buchhorn (1992), but folded by Persic & Salucci (1995), and was confirmed on this large sample with a conservatively estimated upper bound probability of 10^{-7} against it being a chance effect.

However, caution dictates that, although this analysis gave an extremely powerful positive result, the discrete-state hypothesis is so unexpected and difficult to comprehend that further analyses of fresh data had to be considered a priority. In this paper, we review the earlier work as a preliminary to describing the analyses of two additional samples; one sample, of 1200+ Southern sky ORCs, was published by Mathewson & Ford (1996) whilst the other, of 300+ Northern sky ORCs, was published by Courteau (1997). These analyses provide overwhelmingly compelling confirmation of what was already a powerful result.

When these results are combined with those of Roscoe (1999b), the net conclusion is that an individual spiral galaxy appears to be confined to evolve over one of a set of discrete state planes in the three-dimensional (M, S, α) space, where M is absolute magnitude, S is absolute surface brightness and α is the exponent in the power law referred to above.

Whilst we have confidence that the presented analysis confirms a new astrophysical phenomenon beyond all reasonable doubt, this phenomenon is so far beyond the explanatory powers of any extant conventional theory that we feel unable to offer anything other than the most superficial theoretical observations.

1 Introduction

This paper describes the analyses of three large optical rotation curve samples to show how the hypothesis that 'spiral galaxies are constrained to occupy discrete coherent states' is supported by the data as a statistical certainty. The result is so unexpected, that a short review of already published material (Roscoe 1999a) is likely to be useful to the present reader.

Table 1: Twelve Rubin, Ford & Thonnard 1980 spirals

Galaxy	$\ln A$	Galaxy	$\ln A$
N3672	3.6	U3691	3.6
N3495	4.0	N4605	4.0
I0467	4.1	N0701	4.1
N1035	4.1	N4062	4.5
N2742	4.5	N4682	4.5
N7541	4.6	N4321	4.9

Table 2: $\ln A$ data

RFT scale	Pred value with MFB scale	Actual value MFB scale
3.5	3.81	3.85
4.0	4.22	4.24
4.5	4.63	4.72
5.0	5.04	5.06

We began with the hypothetical working model that, simplistically, the optical structure of a spiral galaxy could be considered as arising from a balance between the gravitational forces generated by a small dense spherically symetric central core, and the forces generated by a large scale rotation and that, within the context of this model, the properties of the 'maximal disc' were to be assumed. We were then led to consider the possibility that the disc component of optical rotation curves (which is given an operational definition later in this text) might be reasonably described by power laws in the form $V_{rot} = AR^{\alpha}$, with the parameters A and α being determined empirically for each galaxy in turn. As a means of gaining familiarity with this idea, we considered the small sample of 21 ORCs published by Rubin, Ford and Thonnard (1980) from this point of view. Of this sample of 21 ORCs, only twelve manifested reasonably monotonic behaviour and so were selected *on these grounds alone* as reasonable candidates for a power law analysis. Subsequently, a linear regression of the model $\ln V_{rot} = \ln A + \alpha \ln R$ onto each of the twelve ORCs provided twelve sets of parameter-pairs $(\alpha, \ln A)$. The first clear result of this mini-analysis was that α and $\ln A$ appeared to be very strongly correlated - and this particular aspect has now been analysed in great detail using the Persic & Salucci (1995) sample of 900 folded ORCs (Roscoe 1999b).

However, as reference to Table 1 shows (the entries of which have been rounded to the nearest decimal), a curious numerical coincidence arose - specifically, that every one of the twelve $\ln A$ values lay between ±0.15 of an integer or half-integer value - a coincidence that has odds around 1:500 of being a chance occurrence. Of course, the integer/half-integer values themselves can be of no possible significance since, if Rubin et al (1980) had estimated distance scales using a value of H significantly different from the 50km/sec/Mpc they actually used, then a completely different set of $\ln A$ values would have resulted. So, the coincidence was simply that of regularity in spacing which would probably have not been noticed with, say, H = 70km/sec/Mpc. Anyway, curiosity provided a sufficient motivation to consider the matter further, using the Persic & Salucci (1995) sample of 900 ORCs. The Persic & Salucci (1995) sample had its distance-scaling determined by a

Tully-Fisher relationship calibrated by Mathewson et al (1992) which, as it happens, gives a scaling approximately equivalent to using $H = 85$km/sec/Mpc, so that the integer/half-integer hypothesis for $\ln A$ is not appropriate. However, a simple analysis (described in Appendix B of Roscoe 1999a, and relying on the investigation of the (α, $\ln A$) correlation given in Roscoe 1999b), reveals the relation

$$\ln A_{MFB} \approx 0.82 \ln A_{RFT} + 0.94$$

where A_{MFB} denotes the value of A determined using the Mathewson et al (1992) scaling, whilst A_{RFT} denotes its value determined using the Rubin et al (1980) scaling. Using this latter relation, the integer/half-integer values of $\ln A$ in the Rubin et al scaling transform into their corresponding value in the Mathewson et al (1992) scaling according to the first two columns of Table 2. The actual $\ln A$ distribution of the 900 ORCs of the Mathewson et al (1992) sample folded by Persic & Salucci (1995) are as given in Figure 1. The vertical solid bars indicate the *predicted* positions of the peaks, as in the second column of Table 2, whilst the vertical dotted lines indicate actual peak centres, as in the third column of Table 2. The correspondence between the peak positions, predicted on the basis on the twelve Rubin et al (1980) galaxies of Table 1, and the actual peak positions is clearly remarkable.

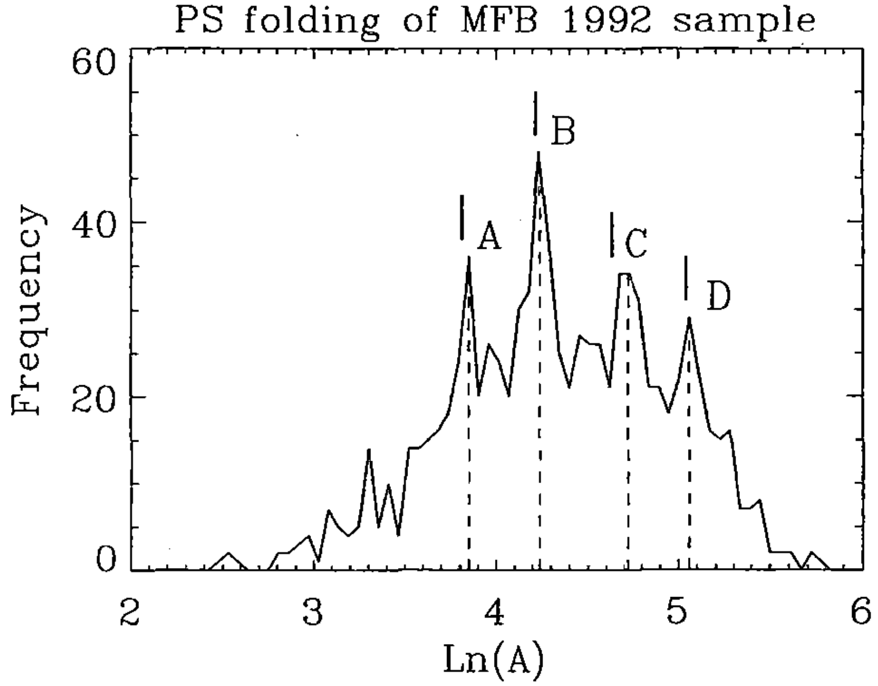


Figure 1: $\ln A$ distribution for the Mathewson et al (1992) sample with Persic & Salucci folding. Vertical dotted lines indicate actual peak centres. Vertical solid bars indicate *predicted* peak centres

A combination of extensive Monte-Carlo simulations and non-parametric statistics allowed a conservative upper bound estimate of the probability of the peaks in the distribution of Figure 1 occurring by chance, given the original hypothesis defined on the small Rubin et al (1980) sample, to be given in Roscoe (1999a) as 10^{-7}.

Table 3: Comparison of the three samples

Sample	Mean distance km/sec	Mean apparent magnitude	% Late type spirals
MFB 1992	3651	12.3 (I-band)	43%
SC 1997	5854	13.5 (R-band)	45 %
MF 1996	6311	13.4 (I-band)	18%

The implications of this result are so profound, that it has become essential to test the specific hypothesis against new samples. This is done using two additional samples in the following sections, and the results are overwhelmingly in favour of the hypothesis. That is, it very much appears as though we are seeing evidence for discrete states in spiral galaxies.

2 The Samples

The basic relevant characteristics of the three samples analysed are given in Table 3, listed in order of probable quality as judged by either mean apparent magnitude, or by % of late-type spirals (which have a higher hydrogen content than early-type spirals, and are therefore likely to be associated with more accurate H_α Doppler shift measurements). We discuss, and analyse, the samples in order of likely quality.

2.1 The Original Sample, Mathewson et al (1992)

In the period 1988-90, Mathewson et al (1992) measured H_α and N_{II} rotation curves for 965 Southern sky spirals on the 2.3m telescope at Siding Spring Observatory, whilst the corresponding I-band photometry was obtained using the 1m and 3.9m telescopes. The N_{II} observations were used to provide an estimate of the internal measurement accuracy of the H_α observations, and these estimates, in the form of a parameter varying on the range (0, 1), were provided for each velocity measurement on every ORC.

Persic & Salucci took this sample of 965 ORCs and subsequently produced a sample of 900 good-to-excellent quality folded ORCs (Persic & Salucci 1995), suitable for their purpose of modelling the internal dynamics of spiral galaxies. It was on this sample that the 'discrete states' hypothesis was originally tested (Roscoe 1999a), and which is represented in Figure 1.

2.2 The Second Sample, Mathewson & Ford 1996

The second sample of 1200+ ORCs was obtained by Mathewson & Ford (1996) in the period 1991-93 as part of the same observing program that gave the original 965 ORCs of Mathewson et al (1992). The main differences between the Mathewson & Ford (1996) and Mathewson et al (1992) samples are given in Table 3: It is clear that the Mathewson & Ford (1996) sample is, on average, 73% more distant than the Mathewson et al (1992) sample, meaning that, on average, we only receive 1/3 as much light (all other things being equal) from each of the objects; this is consistent with the fact that there is an average of a 1.1 apparent magnitude difference between the samples. This large difference in 'light received' indicates that we can expect ORC measurements on the Mathewson

& Ford (1996) sample to be significantly less accurate than those on the Mathewson et al (1992) sample.

Furthermore, the Mathewson et al (1992) and Courteau (1997) sample consists of 43% and 45% respectively of late-type spirals, whilst only 18% of the Mathewson & Ford (1996) sample consists of late-type spirals. Since late-type spirals are significantly richer in hydrogen than are early-type spirals, and since ORCs are measured primarily in H_α, then we can expect the quality of velocity measurements in the Mathewson & Ford (1996) sample to rank behind that of the Mathewson et al (1992) and Courteau (1997) samples for this reason also.

2.3 The Third Sample, Courteau 1997

The third sample, of 300+ ORCs, was selected by Courteau from a sample of *Sb*, *Sc* field galaxy ORCs (Courteau 1997) for a linewidth/Tully-Fisher study. The original observations were made using the Shane $3m$ telescope at Lick Observatories and the du Pont $2.5m$ telescope at Las Palmas. Thus, the importance of this second sample for the present investigation is its total independence of the Mathewson et al (1992) and Mathewson & Ford (1996) samples.

As reference to Table 3 shows, the Courteau (1997) sample is almost as distant as is the Mathewson & Ford (1996) sample, but it contains a similar proportion of late-type spirals to that contained in Mathewson et al (1992). Thus, we would expect the quality of these ORCs to be midway between that of Mathewson et al (1992) and Mathewson & Ford (1996) - all other things being equal.

3 Linewidth estimation

The results of this study are based on using Tully-Fisher methods to set the distance scales and these methods are critically reliant on reliable optical linewidth estimates.

For the samples of Mathewson et al 1992 and Mathewson & Ford (1996), Mathewson et al used an eye-ball case-by-case method to estimate optical linewidths, and so the two largest samples are analysed here using these subjectively derived estimates. By contrast, the Courteau analysis (Courteau 1997) was explicitly designed as a study of objective black-box methods of optical linewidth estimation. He tests a variety of methods, and we present results using those two which he judges to be the best and the worst respectively.

4 The Folding Methods

In the following, a brief description for each of the two folding methods that have been used is given.

4.1 The method of Persic and Salucci

Persic & Salucci (1995) were primarily interested in using rotation curves for studies of the interior dynamics of spiral galaxies and so, by their own criteria, had a requirement for a large sample of particularly accurately folded ORCs. They took the 965 ORCs of Mathewson et al (1992) and used an eye-ball method of folding to produce a sample of 900 good-to-excellent quality folded ORCs; as a qualitative measure of the effort expended to produce this sample, we can note that it took these two authors about a year to process it (private communication). Every velocity measurement in the Mathewson et al (1992) sample came provided with a parameter (varying on the range (0, 1)) which estimated the relative internal accuracy associated with the measurement. Persic & Salucci (1995) found that the accurate folding of any given ORC required the *rejection* of any individual velocity measurement for which the associated accuracy parameter was ≤ 0.35. In the present context, only the Mathewson et al (1992) sample has been folded with this method.

4.2 The auto-folder method of Roscoe 1999c

This method was developed in anticipation of the need accurately to fold the Mathewson & Ford (1996) sample of 1200+ ORCs on a reasonable time-scale. The details of this method are described in Roscoe (1999c) but, briefly, it is based on the formal minimization of the symmetric components in Fourier representations of ORCs with respect to variations in the two folding parameters.

The folding method was developed on the Mathewson et al (1992) sample of 965 ORCs and, corresponding to the experience of Persic & Salucci (1995), we found that the optimal trade-off point between the quality of individual velocity measurements, and the volume of good-quality data available for the automatic folding method, required the prior rejection of any individual velocity measurement which had an associated relative accuracy parameter ≤ 0.4. This folding method has been used here on the samples of Mathewson at al (1992), Mathewson & Ford (1996) and Courteau (1997).

The auto-folder was able to fold 866 of Mathewson et al's 965 ORCs, 1085 of Mathewson & Ford's sample of 1200+ ORCs and 283 of Courteau's 305 ORCs.

5 The computation and representation of $\ln A$ on the disc component of optical rotation curves

5.1 The computation of $\ln A$

The 'discrete states' hypothesis is a statement which *specifically* concerns the values assumed by the set of $\ln A$ parameters, computed for each ORC in turn. It is therefore necessary to state clearly how this parameter is computed.

Roscoe (1999b) showed how the dynamics on the interior sections of ORCs differs objectively and significantly from the dynamics on the exterior sections (note: we are specifically excluding the flat H_I extensions by restricting the discussion to ORCs), and this phenomenon is reviewed in detail in §6 here. Accordingly, since the exterior section largely coincides with the optical disc, we define it as the 'disc component' of the ORC; it was then found that the disc components of the 900 ORCs in the Persic & Salucci (1995) sample (ie the Mathewson et al (1992) sample with Persic & Salucci (1995) folding) are described, to extremely high statistical precision, by the power laws $V_{rot} = AR^{\alpha}$ where the parameters (A, α) vary between galaxies. The $\ln A$ values shown in the distribution of Figure 1, and in every other $\ln A$ frequency diagram in this paper excepting Figure ??, are computed from this model applied to the disc component (operationally defined in §8) of the rotation curve.

5.2 The representation of $\ln A$

All the $\ln A$ frequency diagrams shown in this paper are obtained using the same bin-width ($\Delta \ln A = 0.055$) and initial point ($\ln A = 2.2$) that were used in the original paper, Roscoe (1999a), on this topic. There are therefore no hidden degrees of freedom available to enhance the signals being discussed.

6 The partitioning of optical rotation curves into two objectively defined distinct dynamic zones

In the following, we give a rationale for why it might be expected that ORCs are partionable into distinct dynamic sections and then go on to show how the proposed dynamical partition has an objectively defined reality.

6.1 Basic rationale

The initial study (Roscoe 1999b), from which the present work has flowed, was concerned purely with the hypothesis that the disc dynamics in spiral galaxies can be accurately described in term of a power-law $V_{rot} = AR^{\alpha}$, where the parameter pair (A, α) varies from galaxy to galaxy. The restriction of this hypothesis to the *disc* part of spiral galaxies presented a practical problem, since pure disc spirals are extremely rare - if they exist at all; the reality is that the structure of ORCs arises from a complex interplay of core, disc and halo dynamics - with a question mark hanging over the role, if any, of dark matter. It follows that any specific quantitative approach to the problem of how one might attempt to isolate (even approximately) the disc-dominated dynamics must be based on some form of modelling assumption about the nature of spiral galaxies. For this purpose, we made the assumption of maximal discs, so that the generally rising behaviour of ORCs can be qualitatively accounted for by the visible material using conventional theory.

Given maximal discs then, for the purpose of studying ORCs, we assumed - as a working model - that an idealized spiral could be considered to consist of a very dense gravitational core embedded within a luminous spherical bulge which is itself embedded in a luminous disc. Given this picture, we can conclude that the corresponding ORCs can be considered to extend across two distinct dynamical regimes - one dominated by the core the other dominated by the disc with a transition region ocurring somewhere between the dense core and the luminous bulge/disc boundary. Since the original hypothesis concerned only the behaviour of dynamics in the disc, then the requirement to test it reduces to the problem of finding some means of partitioning the ORC into core-dominated and disc-dominated sections. We describe how this is accomplished, and the effectiveness of what is accomplished, in the following sections.

6.2 Overview of the ORC dynamical partitioning process

The algorithmic details of the dynamical partitioning process are given in Appendix A whilst, in the following subsections, we discuss two methods of assessing its efficiency and effectiveness. These two methods establish with virtual certainty the truth of the statements that:

- the innermost parts of ORCs exhibit behaviour which is qualitatively sharply distinguished from that exhibited by the outermost parts of ORCs;
- the size of such innermost sections can be quantified in terms of a radial measure, R_{min} say, which can be shown to be extremely powerfully correlated with the independently defined optical radius, R_{opt} of the disc. Since R_{opt} carries physical information about the system, then we must conclude that the algorithmically estimated R_{min} likewise carries physical information about the system.

Given the quality of the statistics involved, these two points are entirely sufficient to establish that R_{min} does, in fact, define a real boundary between distinct dynamical regimes which, in turn, gives a concrete justification to the technique by which it is estimated. We interpret R_{min} as a tracer for the gravitational radius of the core on the basis of the circumstance that there appears to be no other possible interpretation.

6.3 A Test of dynamical partitioning

Suppose we define R_{min} as the innermost radial measurement on any given folded ORC and R_{opt} as the optical radius (here, as given by Persic & Salucci 1995). We argued in Roscoe (1999b) that, if the R_{min} value given by the dynamical partitioning process (cf Appendix A for the algorithm) really was a tracer for the gravitational radius of the core, then we might expect to find a positive correlation between R_{min} and R_{opt} - on the grounds that galaxies with large cores might be expected to have large optical radii etc.

Figures 2 and 3 show the $(R_{opt} : R_{min})$ scatter plots for Mathewson et al data before and after applying the dynamical partitioning process respectively. The difference between the two diagrams is dramatic: whilst there is no obvious correlation in Figure 2, the post-partioning diagram of Figure 3 shows an extremely strong positive $(R_{opt} : R_{min})$ correlation. Figures 4 and 5 show an equally strong effect for Mathewson & Ford (1996) data, whilst figures 6 and 7 show the same effect for the much smaller Courteau sample. The foregoing considerations lead to the following conclusions:

- The application of the dynamical partitioning process produces a very powerful $R_{min} : R_{opt}$ correlation confirming that R_{min} (as computed by dynamical partitioning) is a powerful tracer for R_{opt};
- The computed value of R_{min} defines a physical transition boundary between core-dominated dynamics and disc-dominated dynamics.

Taking these items together, and noting the absence of any other obvious interpretation, we conclude that R_{min} almost certainly represents a dynamically derived tracer of the gravitational radius of the core.

7 The Analysis of the Mathewson et al (1992) Sample

The Mathewson et al (1992) sample, like the Mathewson & Ford (1996) sample, is drawn from an area of the sky which Lynden-Bell, Lahav & Burstein (1989) believe contain the Great Attractor (GA) and approximately one half of the Mathewson et al (1992) and Mathewson & Ford (1996)

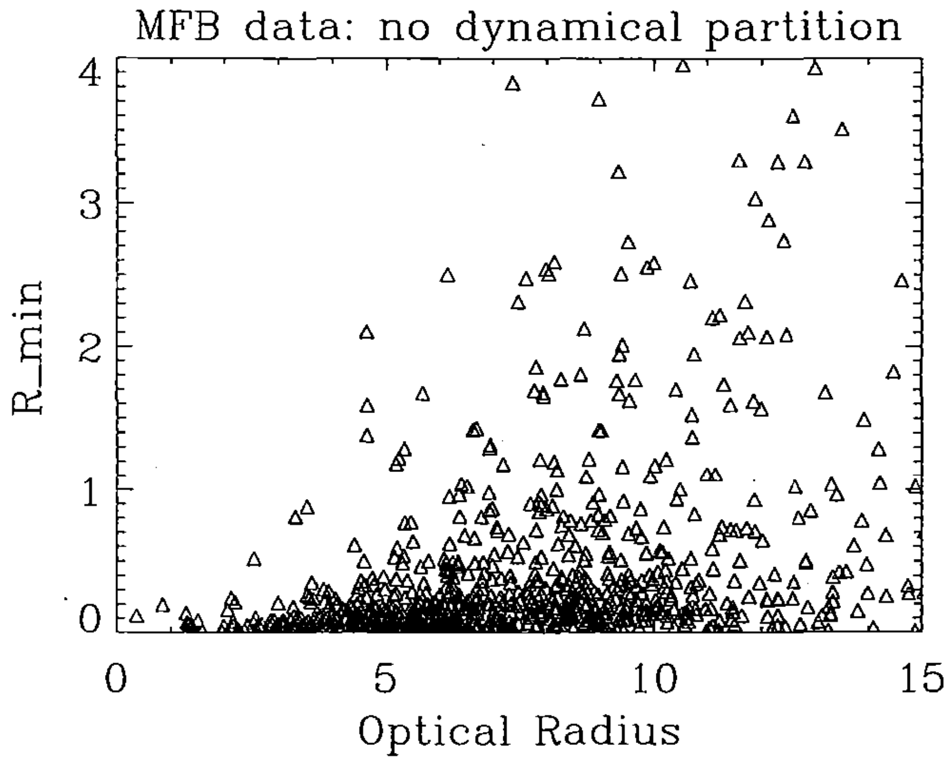


Figure 2: Scatter plot of R_{min} : R_{opt} for MFB data without dynamical partitioning

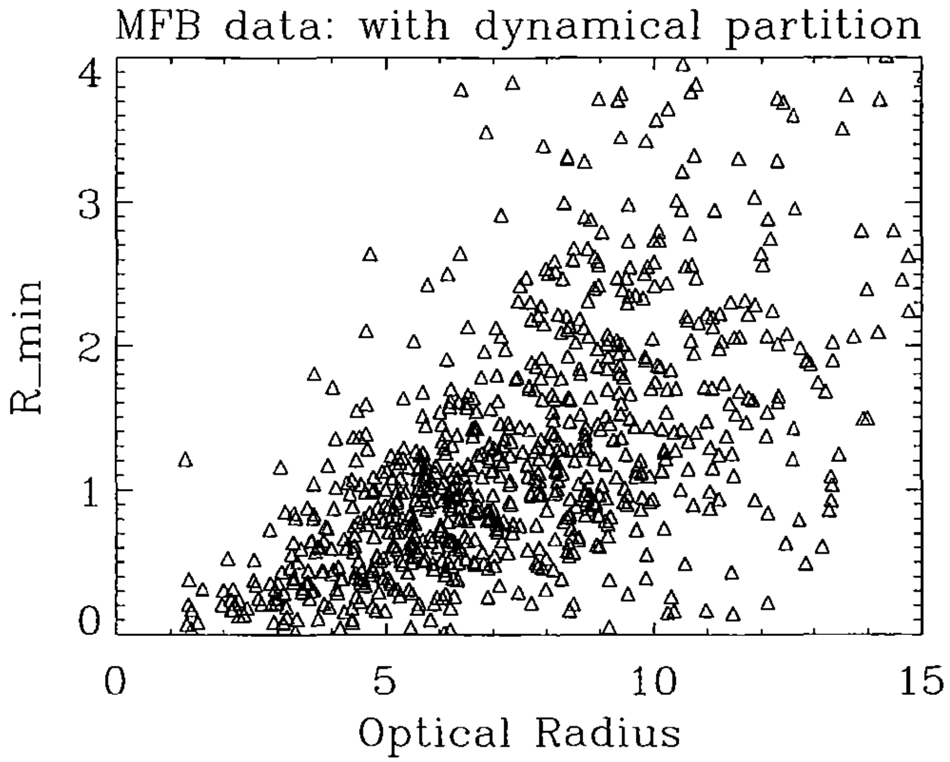


Figure 3: Scatter plot of R_{min} : R_{opt} for MFB data with dynamical partitioning

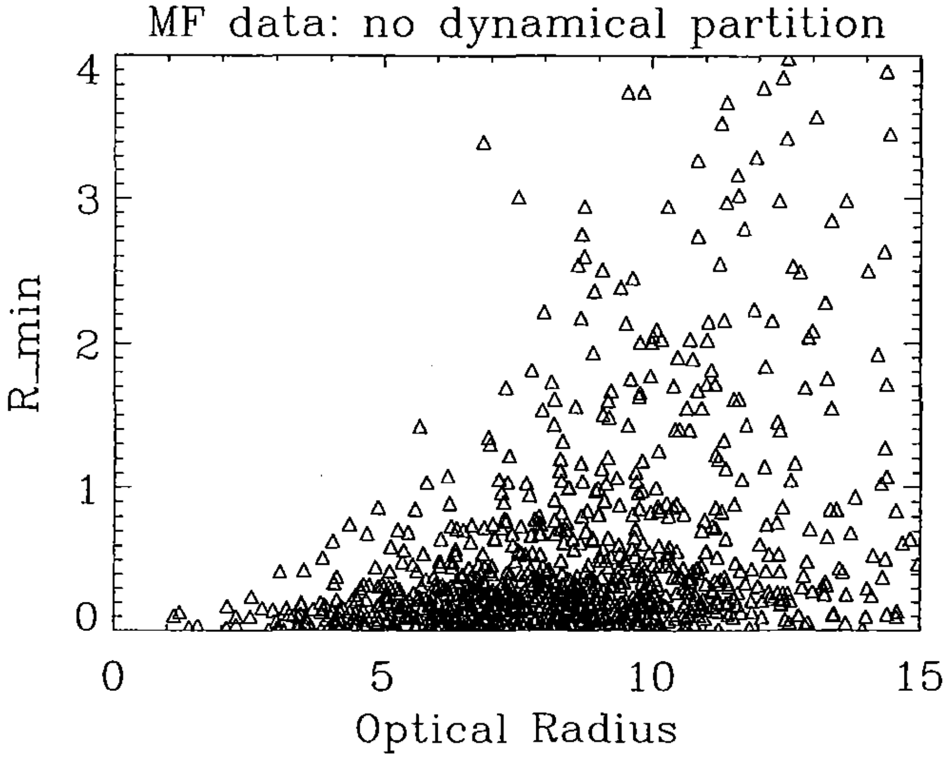


Figure 4: Scatter plot of R_{min} : R_{opt} for MF data without dynamical partitioning

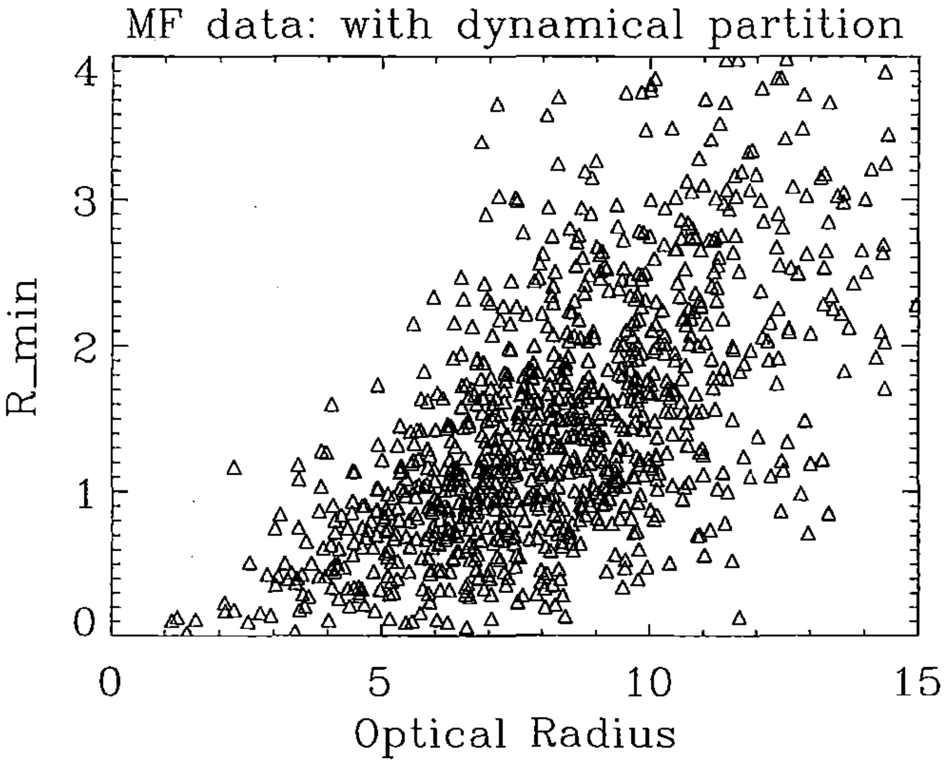


Figure 5: Scatter plot of R_{min} : R_{opt} for MF data with dynamical partitioning

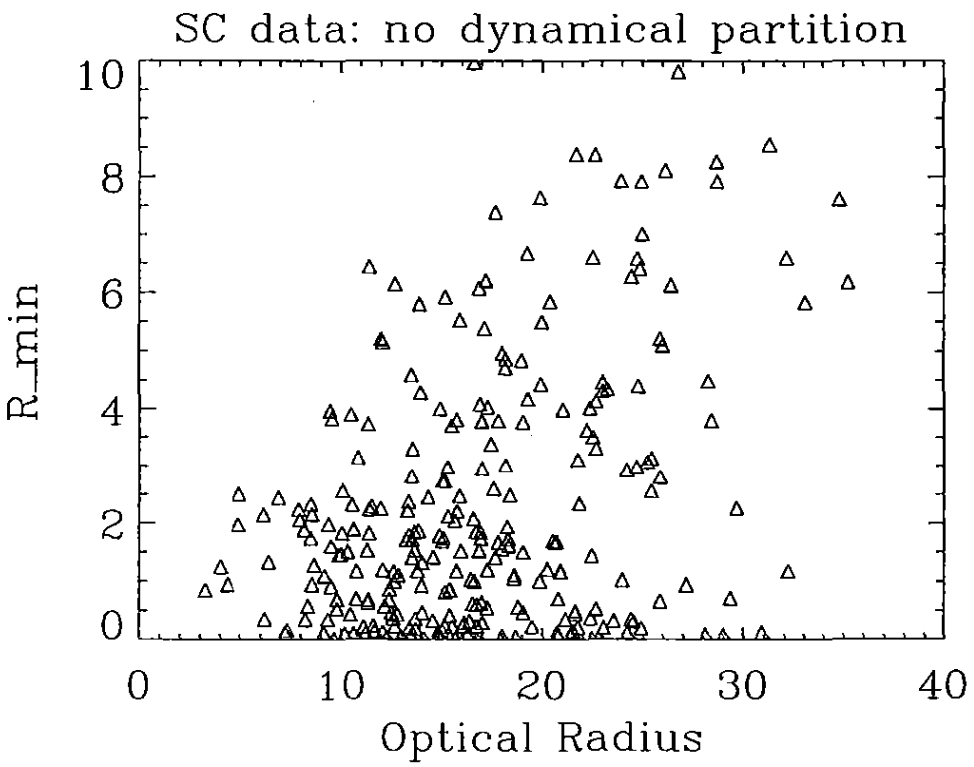


Figure 6: Scatter plot of R_{min} : R_{opt} for SC data without dynamical partitioning

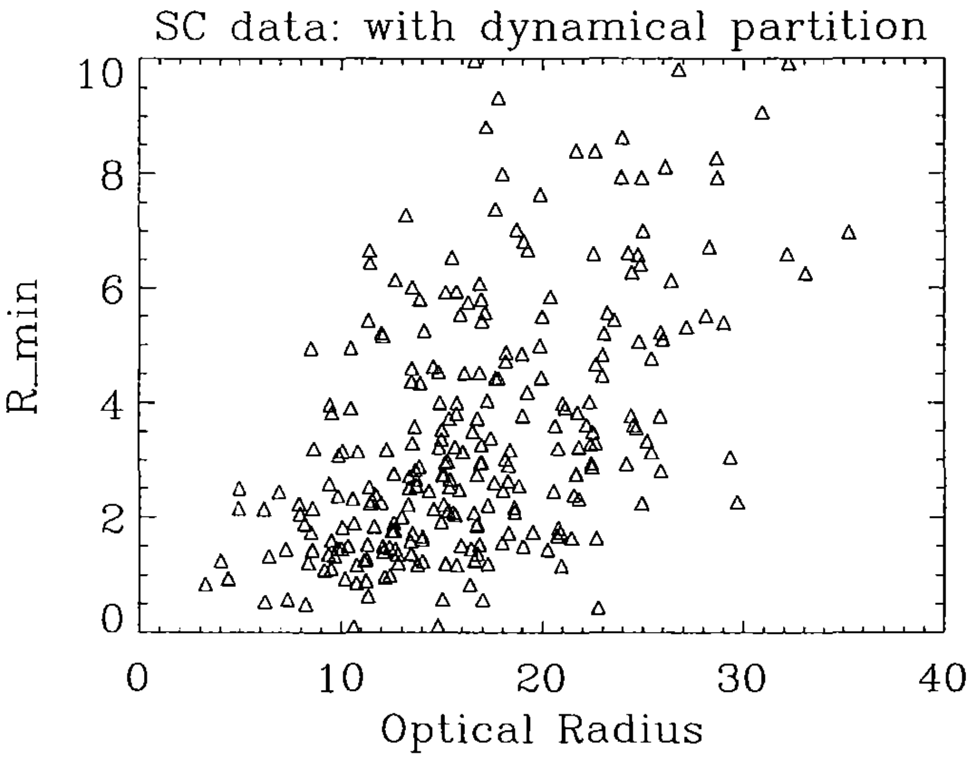


Figure 7: Scatter plot of R_{min} : R_{opt} for SC data with dynamical partitioning

samples lie within (Mathewson et al's (1992) definition of) the GA region, $260^o < l < 360^o, -40^o < b < 45^o$. In their Figure 12, Mathewson et al (1992) use their Tully-Fisher calibration to show that inside the GA region, the data exhibits a clear bias consistent with some form of large-scale flow whilst, outside of the GA region, there is no such bias.

It transpires that these considerations are of considerable significance for the analysis of the Mathewson & Ford (1996) sample and so, to facillitate the various comparisons that we make within the analysis, we adopt the following modus-operandi: For ideal data, for which no systematic bias of any kind exists, we should find a statistical equality between Hubble magnitudes and Tully-Fisher magnitudes; that is, we should find $M_{TF} \approx M_{Hubble}$ over the magnitude range of the sample. For present purposes, the total of Mathewson et al (1992) and Mathewson & Ford (1996) data can be considered to consist of four partitions, these being the non-GA and GA partitions of Mathewson et al (1992) data, and the non-GA and GA partitions of Mathewson & Ford (1996) data. To judge the effect of (say) the Mathewson et al (1992) Tully-Fisher calibration on (say) the non-GA Mathewson et al (1992) partition, we compute the regression model $M_{TF} = AM_{Hubble} + B$, together with the 2σ limits on the range of Hubble magnitudes in the sample - say $(M_{min}, M_{max})_{Hubble}$ - which contain about 95% of the sample, and then use the regression model to compute the magnitude mapping

$$(M_{min}, M_{max})_{Hubble} \rightarrow (M_{min}, M_{max})_{TF}.$$

This mapping then allows us to make direct judgements about the existence of biases somewhere in the system without reference to the details of the regression model. However, because the 2σ limits on the range of Hubble magnitudes differs between the four partitions, direct comparisons are made difficult. To circumvent this, we adopt as a standard reference range the 2σ limits on the range of Hubble magnitudes in the non-GA Mathewson et al (1992) sample - actually given by $(-23.3, -18.2)$ - and consider how this maps to Tully-Fisher magnitudes in the various circumstances considered.

7.1 The Mathewson et al calibration for MFB data

As already noted, Mathewson et al (1992) demonstrate clear evidence (in their Figure 12) for the existence of some form of large-scale bulk flow for that part of their sample in the GA region. For this reason, all calibration discussions in this paper relating to Mathewson et al (1992) and Mathewson & Ford (1996) data are restricted to the non-GA part of the samples. So, for purposes of later reference, we present the effects of Mathewson et al's (1992) own Tully-Fisher calibration on their own non-GA sample, and comment briefly.

Mathewson et al (1992) calibrated their Tully-Fisher relation against the Fornax cluster (for which there is a very narrow redshift dispersion) on the basis of the assumption that Fornax is at 1340km/sec (using $H = 85$km/sec/Mpc), to obtain

$$M = -8.18 \log V_{rot} - 2.86, \tag{1}$$

as their Malmquist bias corrected form. With this Tully-Fisher calibration, we find that, for the subsample *exterior* to the GA region, the 2σ limits on the range of Hubble magnitudes (covering about 95% of the sample) map into Tully-Fisher magnitudes according to

$$(-23.3, -18.2)_{Hubble} \rightarrow (-23.0, -18.1)_{TF} \tag{2}$$

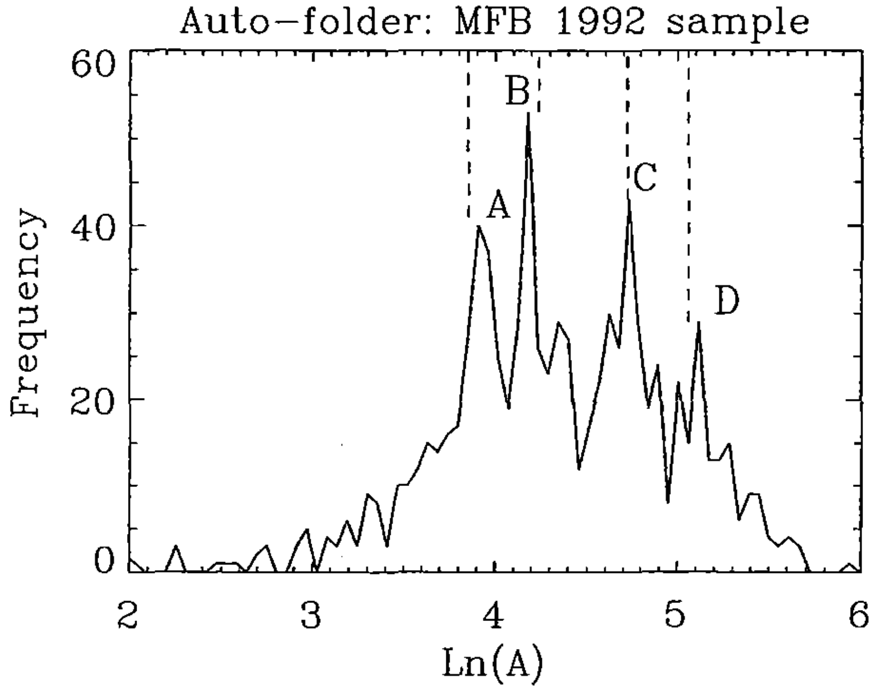


Figure 8: $\ln A$ distribution for the Mathewson et al (1992) sample with auto-folding and dynamic partitioning; Vertical dotted lines indicate peak centres of Persic & Salucci solution

which indicates that the bright end of Hubble magnitudes are too bright in the non-GA region. This could indicate either a problem with the Tully-Fisher calibration at the bright end of objects, or a problem with Mathewson et al's (1992) linewidth estimates at the bright end of objects, or that the non-GA sample is not quiet at the bright end of objects. However, since (1) has been calibrated, independently of the current sample, against the Fornax cluster there exists an objective rationale for accepting the calibration as it stands, and assuming the problem lies elsewhere - at least for now.

7.2 The auto-folder analysis

The original accurate folding of the Mathewson et al (1992) sample was performed by Persic & Salucci (1995) and their solution for the sample can be considered represented by Figure 1. We have already noted that an extensive Monte-Carlo simulation allowed a conservative estimate of the probability of the peaks in this latter figure occurring *by chance alone*, given the prior hypothesis raised on the Rubin et al (1980) data, to be less than 10^{-7}. Figure 8 shows the auto-folder solution for the same sample. The vertical dotted lines in this latter figure mark the positions of the peaks A, B, C, D of the Persic & Salucci (1995) solution, in Figure 1; it can therefore be seen that the peak structure revealed by the Persic & Salucci (1995) folding method is not an artifact of their method, but is an objective feature on the Mathewson at al (1992) sample. It is also worth noting that the auto-folder provides a very significantly sharpened signal.

7.3 The Refined hypothesis

The original hypothesis, based on the analysis of just 12 Rubin et al objects, was that the $\ln A$ distribution, computed according to the prescription of §5, would manifest distinct peaks at $(3.81, 4.22, 4.63, 5.0$ (cf Table 2), and the odds of the actual peaks occurring as they did on the Persic & Salucci (1995) folding of Mathewson et al (1992) data by chance alone given this original hypothesis, were computed in Roscoe (1999a) at about 10^{-7}. However, the analysis of the previous section, based on 900 objects rather than just twelve, has allowed an appropriate refinement of this original hypothesis so that it can now be stated as: *The distribution of* $\ln A$, *computed for folded ORCs according to the prescription of* §5, *will show significant peak structure with peaks centred on the* $\ln A$ *values* $(3.91, 4.18, 4.73, 5.12)$, *where 99.5% confidence limits on the positions of these peaks are given by* $(3.82, 3.99)$, $(4.10, 4.21)$, $(4.56, 4.81)$ *and* $(4.95, 5.17)$ *respectively.*

8 The Analysis of the Mathewson & Ford 1996 sample

The Mathewson & Ford (1996) sample, like the Mathewson et al (1992) sample, is also drawn from an area of the sky which contains the, so-called, GA and approximately one half of the Mathewson & Ford (1996) sample lies within the GA region, $260^o < l < 360^o, -40^o < b < 45^o$ but, as reference to Table 3 shows, is an average of 70% more distant than the Mathewson et al (1992) sample and is therefore considerably less bright. We should consequently be alive to the possibility that systematic measurement bias might exist in the Mathewson & Ford (1996) sample which does not exist in the Mathewson et al (1992) sample. Since we have made the working assumption that Mathewson et al's (1992) Tully-Fisher calibration is good, it can be used to test for the possible existence of any such bias - provided that we are careful to use only Mathewson & Ford (1996) objects *exterior* to the GA region for such a test.

8.1 The Mathewson et al calibration for MF data

For the subsample of Mathewson & Ford (1996) data which is exterior to the GA region (roughly half the sample), we find that the reference range of Hubble magnitudes maps into Tully-Fisher magnitudes according to

$$(-23.3, -18.2)_{Hubble} \rightarrow (-23.1, -17.5)_{TF}. \tag{3}$$

A comparison of this with the mapping of (2) for Mathewson et al's (1992) non-GA objects show a virtually identical perfomance at the bright end of objects, but also reveals the existence of a very strong systematic bias towards the dim end in the Mathewson & Ford (1996) sample of non-GA objects which does not exist in the Mathewson et al (1992) sample. Given that the non-GA objects are not believed to be participating in any large-scale flow, there are two basic possibilities for explaining the mapping (3) which can be listed as

- The MF Hubble luminosities are very much overestimated at the dim end;
- The MF Tully-Fisher luminosities are very much underestimated at the dim end.

The first possibility seems unlikely since Mathewson & Ford (1996) photometry is in the I band for which the internal and external extinction mechanisms are well understood, and for which well-tested correction techniques exist and have been applied by Mathewson & Ford (1996). The second possibility would necessarily have its source in the systematic underestimation of optical linewidths. Since Mathewson & Ford (1996) (and Mathewson et al 1992) used a subjective 'eyeball' technique for linewidth estimation (private communication), it seems that a systematic underestimation of dim-end Tully-Fisher luminosities is the most likely explanation for the systematic bias which we have shown to be strongly indicated for Mathewson & Ford's (1996) non-GA subsample.

8.2 A detailed investigation of MF bias

Where a systematic linewidth bias exists, the biased linewidths will be perfectly good tracers for the true linewidths; consequently, corresponding to a Tully-Fisher relationship calibrated for correctly estimated linewidths there will be an equally applicable Tully-Fisher relationship calibrated for systematically biased linewidths.

Accordingly, the basic rationale underlying the following is that any systematic underestimation of optical linewidths in the Mathewson & Ford (1996) data can be allowed for, in a broadbrush fashion, by a recalibration of the Tully-Fisher relationship according to the criteria that, for non-GA objects, $M_{TF} \approx M_{Hubble}$ over the magnitude range of the sample after any such recalibration.

A comparison of (3) with (2) shows that, in fact, the original Mathewson et al (1992) calibration of Tully-Fisher performs virtually identically at the bright ends of the Mathewson et al (1992) and Mathewson & Ford (1996) non-GA samples, but that its performance at the dim end of the Mathewson & Ford (1996) non-GA sample strongly suggests the need for a progressive recalibration towards the dim end of objects in this subsample. Temporarily ignoring the progressive nature of this inferred required recalibration, and simply assuming that the inferred linewidth bias is constant over the whole magnitude range, we found that recalibrating the Tully-Fisher relationship from (1) to

$$M = -7.50 \log V_{rot} - 4.68 \tag{4}$$

maps the reference range of Hubble magnitudes into Tully-Fisher magnitudes according to

$$(-23.3, -18.2)_{Hubble} \rightarrow (-23.2, -18.1)_{TF},$$

which indicates $M_{TF} \approx M_{Hubble}$ to an extremely good approximation over the magnitude range of the sample. The $\ln A$ frequency diagram corresponding to this recalibration is given in Figure 9, where the vertical dotted lines mark the positions of the peak-centres in the corresponding Mathewson et al (1992) diagram, Figure 8. We immediately see that the peaks A and B are perfectly reproduced, whilst the peak C is attenuated/displaced and peak D is non-existent. Since the figure arises from the application of (4) to the Mathewson & Ford (1996) sample, the foregoing circumstances suggest that the linewidth bias which this latter recalibration attempts to correct is strongly present for the slow rotators (low luminosity objects with $\ln A < 4.4$), is moderately present in the medium rotators, $4.4 \leq \ln A \leq 4.8$, and is absent in the fast rotators (high luminosity objects with $\ln A > 4.8$). This conclusion is entirely consistent with the already noted progressive nature of the bias effect apparent in the mapping (3).

The existence of differential bias in linewidth estimates throughout the Mathewson & Ford (1996) sample indicates the need for a differentially calibrated Tully-Fisher relationship for this sample; a simple composite calibration based upon (1) and (4) is suggested as follows:

$$\begin{aligned} M_{TF} &= -7.50\log V_{rot} - 4.68, & \ln A < 4.4 \\ M_{TF} &= -7.84\log V_{rot} - 3.77, & 4.4 \leq \ln A \leq 4.8 \\ M_{TF} &= -8.18\log V_{rot} - 2.86, & 4.8 < \ln A \end{aligned}$$

where the middle-range calibration is a simple average of the two extreme-range calibrations. The application of this composite calibration gives the magnitude map

$$(-23.3, -18.2)_{Hubble} \rightarrow (-23.0, -18.2)_{TF},$$

which is virtually identical to the mapping (2), provided by Mathewson et al's (1992) original Tully-Fisher calibration. The $\ln A$ distribution arising from the composite Tully-Fisher calibration is given in Figure 10, and we now see that each of the A, B, C, D peaks is strongly present at the exact positions of the corresponding Mathewson et al (1992) peaks in Figure 8.

8.3 The combined MF and MFB distribution

Finally, for completeness, Figure 11 shows the $\ln A$ distribution for the combined Mathewson et al (1992) and Mathewson & Ford (1996) samples consisting of a total of 1,954 foldable rotation curves. As before, the vertical dotted lines indicate the centres of the corresponding Mathewson et al (1992) peaks and we see that each of the four peaks is approximately doubled in strength.

9 The analysis of Courteau 1997

From our point of view, the value of Courteau's work lies in its provision of a rotation curve sample which is completely independent in all of its aspects of the Mathewson et al samples. More particularly, the Courteau (1997) analysis was primarily designed to address the problem of linewidth estimations, with a view to obtaining a standardized objectively defined 'black-box' mechanism for this purpose. This general approach is to be compared with the Mathewson et al method of linewidth estimation, which was simply a subjective case-by-case eye-ball method. Courteau considered several possibilities for linewidth definitions, and we present results using his V_{max} and $V_{2.2}$ definitions (his estimated 'worst' and 'best' respectively).

As for the analysis of the Mathewson et al (1992) sample, it was found necessary to filter out the least accurate of the velocity data; this was made possible for the Courteau (1997) data by the availability of absolute error estimates, given for each velocity measurement - Mathewson et al (1992) gave a parameter which quantified an estimate of relative error for each velocity measurement. The $\ln A$ profiles for Courteau data presented here are calculated by rejecting all measurements with an estimated absoluted error $\geq 5\%$. This data rejection policy had the effect that, of the 305 ORCs in the original Courteau sample, only 283 remained with sufficient data points allowing reliable folding by our folding process.

9.1 A problem with the Courteau calibrations

Firstly, a detail: all of the Tully-Fisher relationships quoted in Courteau (1997) have been derived of V_{rot} values which are twice the actual rotation velocities of the galaxies in his sample. This has

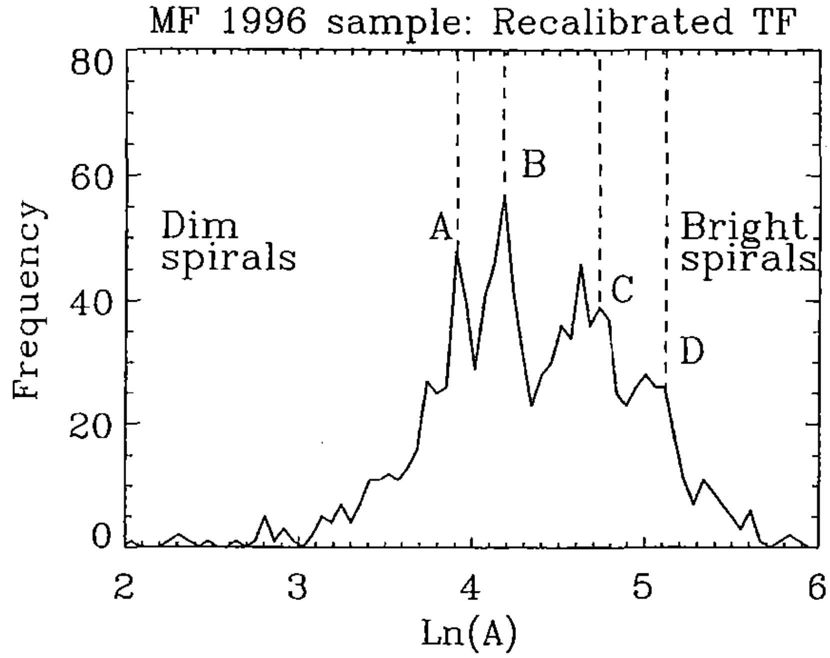


Figure 9: The $\ln A$ distributions for the Mathewson & Ford (1996) sample with auto-folding using the simple TF recalibration; Vertical dotted lines indicate predicted peak centres of the refined hypothesis.

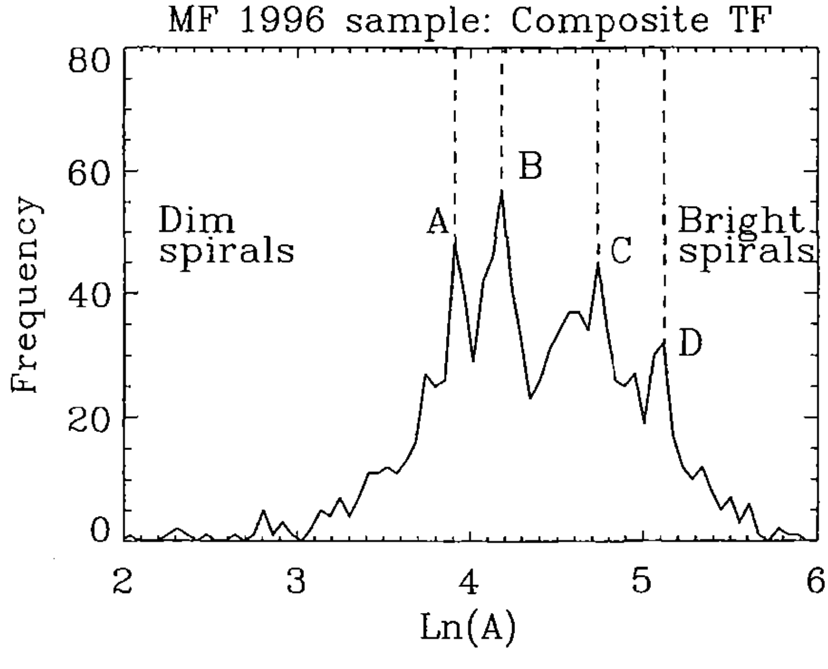


Figure 10: The $\ln A$ distributions for the Mathewson & Ford (1996) sample with auto-folding using the composite TF calibration; Vertical dotted lines indicate predicted peak centres of the refined hypothesis.

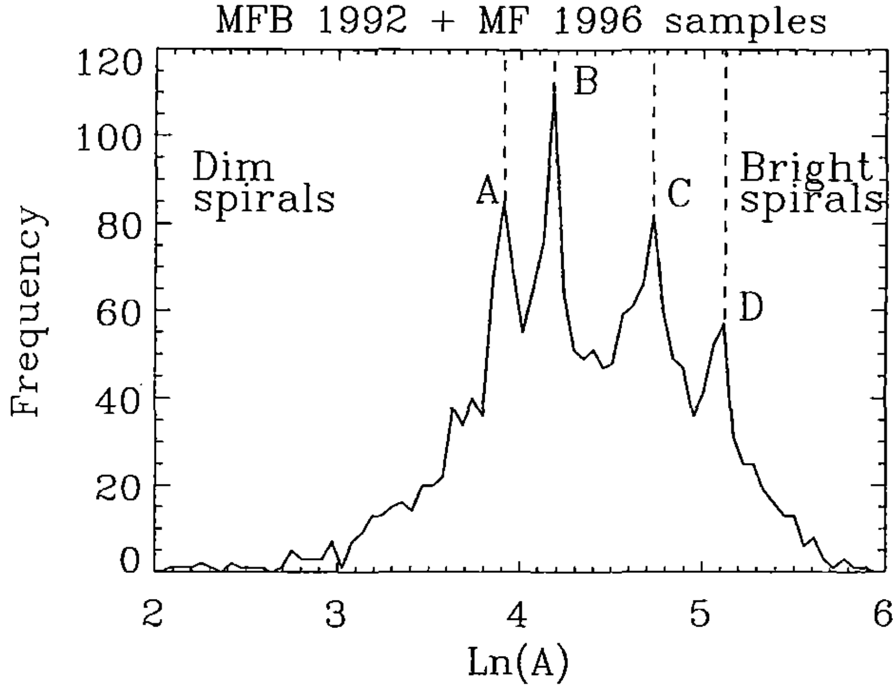


Figure 11: The $\ln A$ distributions for the combined Mathewson et al (1992)+Mathewson & Ford (1996) sample, both with auto-folder; Vertical dotted lines indicate predicted peak centres of the refined hypothesis.

the effect that the zero point values in his calibrations ≈ -4 whereas our are ≈ -6. The use of the $2\times$ factor is made clear in the 'read.me' files which accompany Courteau's data, but not in the actual 1997 paper.

The Mathewson et al Tully-Fisher relation was calibrated against the Fornax cluster within which there is a very narrow redshift dispersion. For this reason, there exists an objective rationale for believing the calibration is good for I-band photometry.

By contrast, the Courteau calibrations - given for each of his linewidth definitions - have been performed over a class of objects with a redshift dispersion in excess of 10,000km/sec, and so there is an underlying assumption that this class of objects is 'quiet' in an overall sense. However, as we show below, there is some problem with Courteau's calibrations, and it seems likely that this problem has its roots in the status of this latter underlying assumption.

We illustrate the problem by reference to Courteau's Tully-Fisher calibration obtained using the $V_{2.2}$ linewidth definition (judged by Courteau to be the best). After our use of $H = 85$km/sec/Mpc in place of Courteau's $H = 70$km/sec/Mpc, this calibration is given by

$$M = -6.36 \log_{10} V_{rot} - 6.36.$$

As we have already pointed out, for quiet data for which Hubble distances are reliable (to within H), the Tully-Fisher relationship should yield the model $M_{TF} \approx M_{Hubble}$. If we now consider the effect of the above Courteau calibration on our standard range of Hubble magnitudes, we find the mapping

$$(-23.3, -18.2)_{Hubble} \rightarrow (-22.4, -18.8)_{TF}$$

which compares extremely unfavourably with the magnitude mapping (2) which arises from the Mathewson et al Tully-Fisher calibration. In quantitative terms, whereas the Tully-Fisher magnitudes in (2) have a range which is 96% of the Hubble magnitude range, the corresponding figure above is 71% - so that the Courteau calibration acts to compress the Hubble magnitude range by 29%, compared to 4% for the Mathewson et al calibration.

An alternative way of viewing the problem is to calibrate the Tully-Fisher relationship to *ensure* that $M_{TF} \approx M_{Hubble}$, to within acceptable tolerance: from this point of view, we find that the Tully-Fisher calibration

$$M = -8.5 \log_{10} V_{rot} - 1.6$$

provides the magnitude mapping

$$(-23.3, -18.2)_{Hubble} \rightarrow (-23.3, -18.2)_{TF},$$

which is exact to within round-off. However, modern 'best practice' calibrations suggest that, typically, Tully-Fisher gradients (the crucial characteristic parameter) for R-band photometry should be about -6.5. Judged against this standard, we see that the gradient of this latter calibration must be classified as extreme and well outside of the expected envelope of values.

The only obvious explanations are that either the Courteau linewidth definitions are unreliable, or that the Courteau data is not quiet ideal data, but is kinematically very active. But, as already mentioned, the Courteau paper is primarily a study of linewidth definitions, and he estimates that his $V_{2.2}$ linewidth definition is comparable with the best H_I linewidth definitions - which are considered to be the 'industry standard'. This means that, almost certainly, Courteau's sample is kinematically very active. In this latter case, calibrating Tully-Fisher directly on the sample (as is done by Courteau) is likely to be an uncertain process.

9.2 Inverting the problem

In view of the manifest problems associated with calibrating the Tully-Fisher relationship on the Courteau data, we turn the problem around to ask the question: *Is it possible to use the discrete states hypothesis to calibrate the Tully-Fisher relation so that any resulting calibration is within the acceptable envelope for R-band photometry?* This question has a positive answer, yielding the calibration

$$M_{TF} = -6.64 \log V_{rot} - 6.67,$$

using Courteau's $V_{2.2}$ linewidth definition - which he judges to be the best of those considered. The effects of this calibration are shown in Figure 12, calculated using $V_{rot} \equiv V_{2.2}$ linewidths, and in Figure 13, calculated using $V_{rot} \equiv V_{max}$ which Courteau judges to be the worst of the definitions considered. In both cases, the reproduction of the predicted peaks is excellent

The important characteristic parameter is the gradient value, and our value of -6.64 (which is to be compared with those given in Courteau's various calibrations which range from -5.77 to -6.99 with a mean value of -6.38) lies well within the envelope of acceptable gradients for R-band photometry. Bearing in mind that there are only two free parameters in the calibration, and the degree of predicted coherence which the variation of those two parameters has induced in the $\ln A$ distribution of Figure 12, we can conclude that the Courteau rotation curve sample lends further powerful support to the discrete states hypothesis.

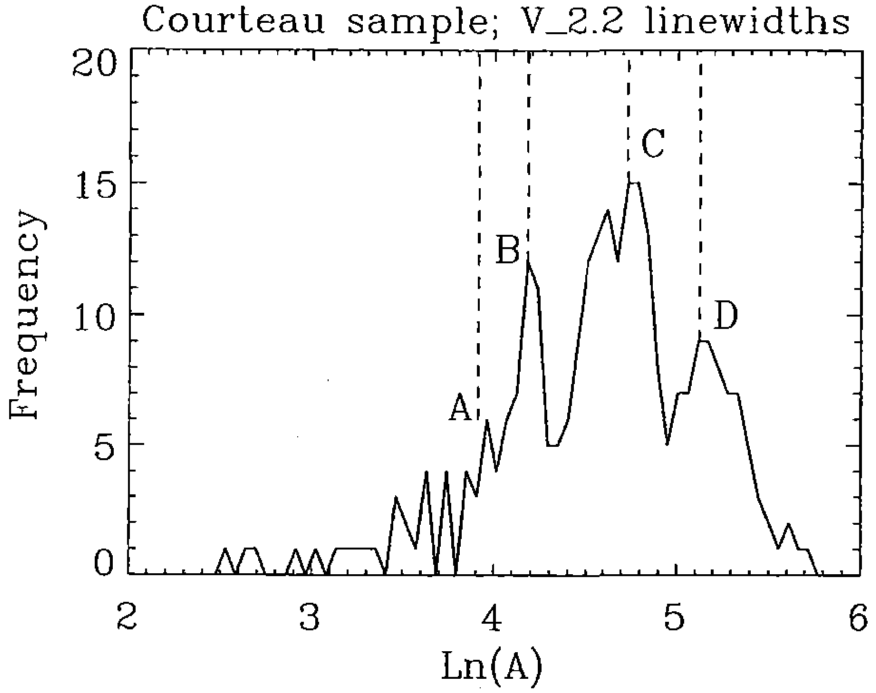


Figure 12: The $\ln A$ distributions for the Courteau (1997) sample using his $V_{2.2}$ linewidth definition. Vertical dotted lines indicate predicted peak centres of the refined hypothesis.

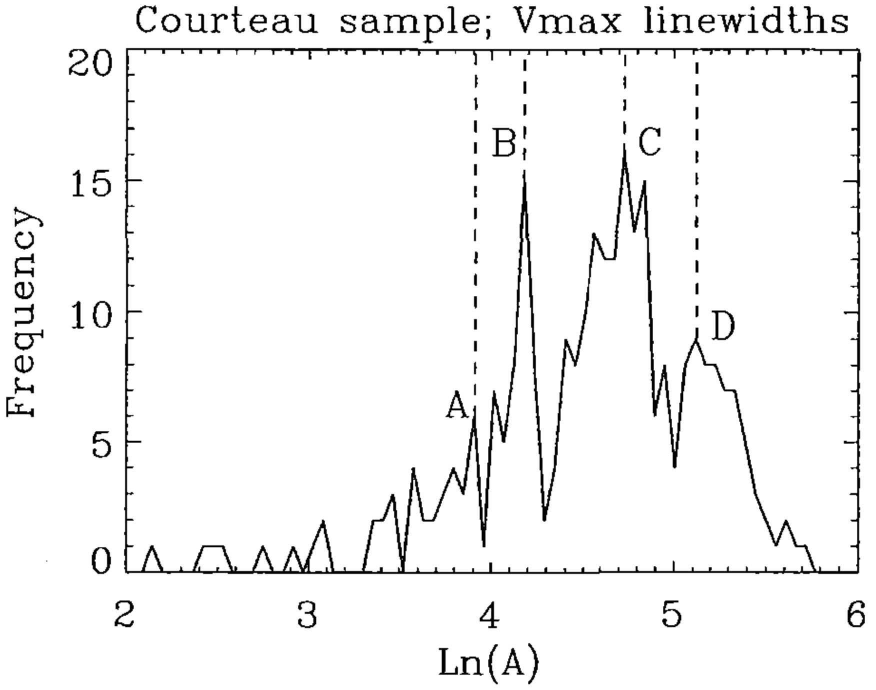


Figure 13: The $\ln A$ distributions for the Courteau (1997) sample using his V_{max} linewidth definition. Vertical dotted lines indicate predicted peak centres of the refined hypothesis.

Table 4: Comparison of peak positions in $\ln A$ frequency diagrams for MFB, MF & SC data

Sample	A	B	C	D
MFB	3.91	4.18	4.73	5.12
MF	3.91	4.18	4.73	5.09*
$SC_{2.2}$	-	4.18	4.76*	5.14*
$SC_{V_{max}}$	-	4.18	4.73	5.12

* indicates average over two points

10 Statistical significance of the new analyses

The results of the foregoing analyses, already encapsulated in Figures 8, 10, 12 and 13, are collected together for convenient comparison in Table 4. The refined specific hypothesis being tested states, briefly, that strong peaks in $\ln A$ frequency diagrams of rotation curve data processed in the way described, should occur coincidently with the peaks A, B, C & D in Figure 8, derived from our autofolder analysis of Mathewson et al (1992) data. The exact positions of these latter peaks are given in the first row of Table 4, whilst the second row gives the position of the corresponding peaks derived from Mathewson & Ford (1996), and the third & fourth rows give these peaks for Courteau (1997) data using his $V_{2.2}$ and V_{max} linewidth estimates respectively. Except for the A-peaks in Courteau data A, which were excluded because of small statistics, it is clear that the peak positions are essentially identical across the three samples.

For the purpose of a formal assessment of the statistical significance of these results, we restrict ourselves to a formal estimate of the significance of the results of the Mathewson & Ford (1996) analysis only, as exhibited in Figure 10. Specifically, we ask what is the probability that the A, B, C, D peaks in Figure 10 occur within 99.5% confidence limits about their observed positions with their observed magnitudes by chance alone.

Extensive Monte-carlo trials give a joint probability of 3.6×10^{-8} of the four peaks occurring simultaneously in the 99.5% confidence intervals at the requisite strengths (or greater) by chance alone. Of course, we could have noted that, in fact, the A, B, C, D peaks in Figure 10 coincide *exactly* with the A, B, C, D peaks of Figure 8, and a formal computation of the probability of this being a chance event gives a value of 1.2×10^{-13}, making it five orders of magnitude less likely.

11 Implications For Galactic Evolution

In dimensionless form, the power-law $V = AR^{\alpha}$ can be expressed as

$$\frac{V}{V_0} = \left(\frac{R}{R_0}\right)^{\alpha}$$

so that $A \equiv V_0/R_0^{\alpha}$. The detailed analysis of the correlation between α and $\ln A$, given in Roscoe (1999b), shows that V_0 and R_0 are each very strongly correlated with luminosity properties. Consequently, for absolute magnitude M and surface brightness S, we can write

$$A \equiv F(M, S, \alpha)$$

which, since A appears to assume discrete values $k_1, k_2, ...$, implies

$$F(M, S, \alpha) = k_1, k_2, ...$$

Thus it appears that spiral galaxies are constrained to exist one of a set of discrete state planes in the three-dimensional (M, S, α) space.

This then gives rise to one of two possibilities: either a spiral galaxy is 'born' on one of these planes, and remains on this plane over its whole evolution; or a spiral galaxy remains on one of these planes for very long periods, with the possibility of transiting to other planes in very short periods of time. There is currently no way of distinguishing between these possibilities.

12 Summary and Conclusions

It is crucial to remember that the phenomenon being discussed was discovered as a result of a specific hypothesis raised on an analysis of those twelve spirals in the Rubin et al (1980) sample of 21 spirals which had monotonic ORCs, and which was tested with extremely positive results against the 900 ORCs of Mathewson et al (1992) which were folded by Persic & Salucci (1995). This was reported in Roscoe (1999a).

Subsequently, we developed an automatic routine for the folding of further samples, and this has enabled us to give an independent re-analysis of the Mathewson et al (1992) sample together with the analyses of two new large samples published by Mathewson & Ford (1996) and Courteau (1997). The re-analysis of the earlier sample allowed a minor refinement of the discrete states hypothesis, and this was confirmed in fine detail on the two new samples.

After taking into account the results of Roscoe (1999b), we concluded that the immediate significance of the results was that any given spiral galaxy appears to be constrained to evolve over one of a discretely defined set of state planes, existing in a three-dimensional (M, S, α) space where M is absolute magnitude, S is surface brightness and α is a parameter computed for each galaxy from its rotation curve.

From the point of view of finding a theoretical perspective, one is tempted to consider the phenomenon in the light of the considerable dynamical evidence that the centres of spiral galaxies are occupied by super-massive objects. Given this picture, it seems reasonable to suppose that dynamics in the disc would carry some signature of the properties of the massive central object (say through boundary conditions set in the transition region between core-dominated and disc-dominated dynamics), in which case it seems plausible to hypothesise that the discrete state phenomenon identified in the discs is actually a reflection of a similar phenomenon existing in the massive central objects. Such a view is, of course, not consistent with the interpretation of these central objects as classical black-hole singularities since such singularities are considered to be structureless. In the absence of 'singularity physics' the super-massive central objects must necessarily be super-dense finite objects of a kind new to physics; the 'discrete state in discs' phenomenon might then be considered as evidence for the idea that these super-dense central objects exist in states of discrete massive coherence - or massively coherent quantum-like states.

Whatever the truth of the matter, it seems certain that the existence of this phenomenon poses very difficult questions for the standard galaxy formation theories, and will have a profound affect on our developing understanding of galactic dynamics and evolution in particular, and the cosmos in general.

Acknowledgements
I am grateful to M. Persic and P. Salucci, of SISSA Italy, V. Ford of Mt Stromlo Observatory ANU, and S. Courteau of the Herzberg Institute of Astrophysics Canada, for making available their data, and for patiently answering all queries.

A Dynamical partitioning: The Minimization of core-effects on disc dynamics and the computation of $\ln A$

A.1 Unusual points

The process to be described uses the techniques of linear regression and, following conventional definitions, an observation is reckoned to be *unusual* if the predictor is unusual, or if the response is unusual.

For a p-parameter model, a predictor is commonly defined to be unusual if its *leverage* $> 3p/N$, when there are N observations; in the present case, we have a two-parameter model so that $p = 2$. Similarly, the *response* is commonly defined to be unusual if its standardized residual > 2.

A.2 The algorithm

The basic assumption is that a rotation curve has (at least) three distinct dynamical segments: (a) the interior core-dominated segment, (b) the middle optical disk segment and (c) the flat outer segment which is primarily observed in H_I. The analysis of Roscoe (1999b) was then predicated on the hypothesis that dynamics in the middle optical disk segment (for an idealized case which discounts the inevitable disc-irregularities) could be described by the power-law $V_{rot} = AR^{\alpha}$, so that

$$\ln V_{rot} = \ln A + \alpha \ln R.$$

The computation of $(\alpha, \ln A)$, using dynamical partitioning, for any given rotation curve can now be described as follows:

1. Eliminate poor data over the whole ORC according to the data-rejection policy described in §4;
2. Form an estimate of the parameter-pair $(\alpha, \ln A)$ by regressing the $\ln V_{rot}$ data on the $\ln R$ data;
3. Determine if the *innermost* data point only (that is, the single point most likely to be affected by the core) is an *unusual* observation in the sense defined above.
4. If the innermost data point is unusual, then exclude it from the computation and repeat the process from (ii) above on the reduced data-set.
5. If the innermost data point is *not* unusual, then no further computation is required - the current values of $(\alpha, \ln A)$ are considered as final.

When applied to the 900 ORCs resulting from the Persic & Salucci (1995) fold of the Mathewson et al (1992) sample, for example, this process leads to the rejection of 11.8% of the total number of individual velocity determinations (2264 out of 19183).

References

[1] Courteau S., 1997, AJ, 114, 6, 2402-2427

[2] Lynden-Bell, D., Lahav, O., Burstein, D., 1989, MNRAS, 241, 325

[3] Mathewson D.S., Ford V.L. 1996, ApJS 107 97

[4] Mathewson D.S., Ford V.L., Buchhorn M. 1992, ApJS 81 413

[5] Persic M., Salucci P., 1995, ApJS 99 501

[6] Roscoe D.F., 1999a, A&A,343, 697-704

[7] Roscoe D.F., 1999b, A&A,343, 788-800

[8] Roscoe D.F., 1999c, A&AS,140, 247-260

[9] Rubin V.C., Ford W.K., Thonnard N. 1980, ApJ 238 471

Proceedings of ANPA West: A Catalogue

Keith Bowden

TPRU, Birkbeck College

Three projects that have been concerning me for some time are the (commercial) publication of a ***Best of ANPA***, probably aimed largely, but not solely, at the Combinatorial Hierarchy, the publication of old pre-ANPA papers on the Combinatorial Hierarchy, and the production of a ***Best of ANPA West*** in this (*Proc. ANPA*) format.

Some progress has been made on the first front, a number of pre-ANPA papers, particularly from the Cambridge Language Research Unit era have appeared in this Proceedings, and as part of my work on the latter project I have produced the following catalogue of the contents of the ANPA West Proceedings which I thought members might find interesting.

ANPA West ran from 1985 to 1997. Fred Young and Tom Etter were the Editors of the Proceedings and the meetings were held annually at Cordura Hall at Stanford University. The following catalogue was taken directly from a set of Proceedings supplied to me by Pierre Noyes last year. The contents lists provided do not always agree with the actual contents of the Proceedings. There were no page numbers in the Proceedings. The numbers in brackets refer to the number of pages in each paper. If anyone has any other copies of Proceedings (particularly Volume 9), not included below, I would much appreciate access to a copy.

Proceedings of ANPA West 6 (1990)
CAUSAL STRUCTURE AND THE QUANTUM

Lou Kauffman, A Recollection of Anne Dale (2)
Fred Young, Universality in Chaos and the Combinatorial Hierarchy (19)
G J Chew, Coherent-State Model of Expanding Universe (15)
David McGoveran, Advances in the Foundations (19)
H Pierre Noyes, A Finite Zitterbewegung Model for Relativistic Quantm Mechanics (19)

Herb Doughty, AA*-A*A=I over GF(p^2) (17)
Mike Manthey, Synchronisation: The Mechanism of Conservation Laws (11)
Tom Etter, Pre-Logic (13)
Eddie Oshins, Why Brown's *Laws of Form* and Pribram's "Hologram Hypothesis" are "...just what is not relevant..." (15)
Peter Marcer, The Grand Unification, the Dream of 20th Century Science has been realised; a Synthesis of the Facts to Substantiate this Assertion (17 A5)
E Seigal, Discreteness Precedes Finiteness (1)

Proceedings of ANPA West 7 (1991)
INTERDISCIPLINARY MODELS IN SCIENCE

Juan de Lorenzana, Informal Comments on Indistinguishability, the Combinatorial Hierarchy and Evolutionary Systems Framework (7)
H Pierre Noyes, Discrete Antigravity (22 A5)
Fred Young, Pattern, Form, Chaos (35)
Vaughan Pratt, Event Spaces and Their Linear Logic (12)
David McGoveran, Justifying the Combinatorial Hierarchy (11)
Tom Etter, The Theory of Amplitude (11)
Bill Honig, The Logical Structure of STR (& QM) using Imaginary and Transfinite Number Forms (10)
Herb Doughty, Blobs in Galois Fields (20)
Jean Burns, What is the Nature of an Object Described by Logic and Prelogic? (12)

Proceedings of ANPA West 8 (1992)
RECONSTRUCTING THE OBJECT

Fred Young, Maps, Codes and Strings (8)
Irv Stein, The Solution to the Problem (5)
Tom Etter, Minds, Bodies and Bits (4)
Deiderik Aerts, In Search of the Meaning of Quantum Structures (17)
E Thomsen, Adding to Space and Time (53)
H Pierre Noyes, The RQM Triangle (20)
Eddie Oshins, The Search for Classical Psychospinors (11)
Eddie Oshins, Models and Muddles (23)
Edward Seigal, Hawking Unruh Effect (25)
Gough and Shacklett, The Unification of Mind and Matter (19)
Herb Doughty, 3D Lucas Zooms (24)
Juan de Lorenzano, The Generation of the Combinatorial Hierarchy (17)

Tom Phipps, An Experiment to show the Acausality of Photon Behavious (5)

Proceedings of ANPA West 10 (1994)
MIND-BODY PROBLEM AND THE QUANTUM

H Pierre Noyes, Stapp's Quantum Dualism (17)
Herb Doughty, What if the Universe is Elegant? (3)
R H Davis, A Simple Proof of a Simple Bell-Type Inequality (2)
Tom Etter, Untitled
Fred Young, From Cellular Automata to Quantum Cellular Automata (11)
Fritz Lehmann, Combining Ontological Hierarchies, As Anticipated by Frederick Parker-Rhodes (24)
Mike Manthey, A Vector Semantics for Actions (27)
Eddie Oshins, Technical Comments on Quantum Psychology and the Metalogic of Second Order Change (11)
Eddie Oshins, Quantum Psychology and the Metalogic of Second Order Change (21)

Proceedings of ANPA West 11 (1995)

Lou Kauffman, Discrete Order Calculus and Formal Derivations of Physical Laws (22)
Keith Bowden, On the Emergence of Physical Structure from Information Theory or The Borders of Reality (10)
Keith Bowden and Clive Kilmister, An Extension to Eddington's Fundamental Theory or Eddington Off Limits (4)
H Pierre Noyes, On the Lorentz-Invariance of Bit-String Geometry (28 A5)
Vaughan Pratt, Rational Mechanics and Natural Mathematics (13)

Proceedings of ANPA West 12 (1996)

Jack Engstrom, Laws of Form's Distributive Law Integrates Relations between Addition, Multiplication, and Exponentiation of Natural Numbers (9)
E R Floyd, Overview of the Trajectory Representation of Quantum Mechanics (15)
Lou Kauffman, Arithmetic in the Boundary (11)
Lou Kauffman and H Pierre Noyes, Discrete Physics and the Dirac Equation (17)
H Pierre Noyes, Are Partons Confined Tachyons? (20)

Proceedings of ANPA West 13 (1997)

Juan de Lorenzana, Some Issues on the Construction of Order in Self-Organising Systems (16)
Keith Bowden, Classical Computation Can Be Counterfactual (or Can Schrodinger's Cat Collapse the Wavefunction?) (7)
Herb Doughty, Musings of a Mathematical Realist -1997 Version (7)
Tom Etter, Shape Theory: A New Operating System for Mathematics (9)
Edward Floyd, Differences In Predictions Between the Trajectory and Copenhagen Representations of Quantum Mechanics (13)
Louis H. Kauffman, (four papers)
Quantum Networks and Topology (5)
Virtual Logic - Boolean Algebra, Computer Proofs and Human Proofs (26)
Quantum Electrodynamic Birdtracks (8)
Sign and Space (5)
Thomas J. McFarlane, Integral Science (16)
P. D. Mountcastle, Signal Zoology and Fourier Botany (11)
H. Pierre Noyes, On the Distinction Between Classical and Quantum Probabilities (17)
Richard Shoup, Much Ado About Nothing or Nothing Up My Sleeve (4=20 slides)_
Fred Young, What Is Spacetime Foam? (15)
Milo Wolff, Exploring the Universe and the Origins of Its Laws (16)

ANPA West Journal Catalogue

Finally, in the following pages, is Tom Etter's catalogue to the ANPA West Journal, a separate publication edited by Tom Etter. I have one later issue *Vol. 7, No. 1*, which I suspect was the last. Its contents comprise

Geoffrey Chew, Reflections on Reality Inspired by an Event Model of Evolution
Tom Etter, Theories of Psi
Arleta Griffor, Mind and its Wholeness
Pierre Noyes, A Closing Note

Back Issues

The following articles appear in back issues of the ANPA WEST Journal which are available on request.

Volume1 # 1.

A Conversation with Pierre Noyes About ANPA History.

Pierre Noyes, a theoretical physicist at SLAC, founded ANPA in 1979.

Why Discrete Physics? *by Pierre Noyes*

A brief statement of the vision that has animated Pierre Noyes' work over the years.

An Introduction to the Ordering Operator Calculus. *by Dave McGoveran*

The author, a mathematician who has collaborated with Pierre Noyes, introduces us to his context-sensitive process-oriented "modelling methodology".

Science Without Logic? or Very Elementary Physics. *by Tom Etter.*

Quantum logic is an alternative to Boolean logic. This article introduces the concept of pre-logic, which is a weakening of Boolean logic.

Volume 1 #2

Articles in this issue have mostly to do with language and writing.

Radio Shack Meets Star Trek: A New Electronic Awareness Module.
by Nick Herbert.

Those of you who have read Nick Herbert's book "Quantum Realities", an admirably lucid introduction to the problem of interpreting quantum mechanics, will recognize his inimitable light touch in our lead article, which is about the kind of "writing" that might come from beyond the quantum veil.

What is Language that Egos May Share it? *by Helga Wild and Niklas Damiris*

The second article, a collaboration between brain scientist Wild and physicist-philosopher Damiris, is a brief headlong plunge into the dense philosophical thicket surrounding language, knowledge, world and self.

On Etter's Flying Bananas. *by Herman Mueller.*

The graceful prose of our third article quickly reveals its author to be a professional wordsmith. Ironically, his point of departure is to take your editor to task for making too much out of words. He goes on to share with us his own explorations of what lies behind words, touching on philosophical issues that are of great importance for the kind of science that some of us in ANPA are trying to create.

On Grounding the Bananas. *by Tom Etter.*

This article is your editor's response to Mr. Mueller's criticism. It tries to fill in some of the background that was presupposed by the essay he was criticizing, and in keeping with our theme, sketches some connections between writing and quantum measurement.

On to QED. *by H. Pierre Noyes.*

Our last article is in a rather different vein, being a progress report on some promising new turns in Noyes-McGoveran bit-string physics. Some of us non-physicists have a big stake in all this, since if it works out, it may prove to be the royal highway connecting our areas of interest to mainline experimental physics.

Volume 1 #3

A Theory of Shadows. *by Wayne Blizard.*

Wayne Blizard is a mathematical logician who was the first to give a rigorous axiomatic formulation of the theory of multi-sets. He is also a poet. In this article he wears both hats, which seems appropriate for someone who is introducing the `radical alternative to non-locality.

How to Be in Two Places at One Time. *by Nick Herbert*

Before we turn to what Bell's theorem means, we must be clear about what it says. This article, which is reprinted from `The `New `Scientist 21, `August `1986, is probably the best short introduction to Bell's theorem around, complete with a proof which sets some kind of record for the ease with which it conveys a subtle abstract idea.

How to Be Your Own Grandfather. *by Tom Etter*

You won't learn this from Star Trek, but Einstein's theory of relativity says that "warp speed" can take you into your own past. There are some fallacious proofs of this around, and at least one fallacious disproof has found its way into a respectable journal; here is a brief and painless proof which is actually valid.

How Do Scientists Work? *by Karel Pstruzina*

As science enters its post-classical phase, we must expect it to stray a bit from its classical methods. How about doing science in your sleep? Sounds crazy? Ah, but to quote Neils Bohr, ".. is it crazy enough to have any chance of being true?" Read this article and judge for yourself.

Karel Pstruzina is head of the philosophy department at the Prague School of Economics.

Dream Logic. *by Tom Etter*

Pstruzina's method may work because the sleeping brain, freed from its practical chores, becomes some kind of super- computer. But then again, maybe it works for a very different reason: maybe the waking world, and especially the quantum part of it, is more of a piece with the dream world than we realize. Perhaps, just as the molecular bonds in a gas are too unstable to hold together solid material bodies, the logical bonds in whatever underlies the quantum are too unstable to hold together solid material facts.

Etter sketches here a mathematical theory of "logical gas" which he calls pre-logic, and indicates how it necessarily leads to non-locality in a way that gives meaning to negative set membership.

A Chimeric Kind of Causality. *by Dr Jabir 'abd al-Khaliq FRS*

Dr. Jabir is a nomadic monad, quartz crystal sexer, quantum connection specialist, and foreign correspondent for the Fezziah Research Sector, Fez, Morocco. Here he gives us an illuminating refutation of certain claims about so-called symmetrical causality made by Nick Herbert in his book "Faster Than Light".

... I'll Take the High Road. *by David McGoveran*

Everyone who has worked with quantum theory knows there is something a bit "off" about it. As one physicist put it, "There are things in quantum mechanics which make me very uncomfortable, but I can't quite say what my discomfort is about."

Etter's diagnosis of what this discomfort is about points to logic. McGoveran, on the other hand, locates the malady, and its cure, in the realm of causality. In this paper, he proposes a theory of relativity of cause and effect. He postulates a multiplicity of causal frameworks, analogous to the multiple inertial frames of Einsteinian relativity, but more radically diverse in that their arrows of cause and effect may be quite inconsistent. Each of these frameworks is alike in lying within its own Lorentz-invariant space-time, which its causal arrows actually define, albeit inconsistently. Since all frameworks are in some way involved in the course of events that we perceive and measure, non-local correlations are to be expected.

Advances on Two Fronts. *by H. Pierre Noyes.*

We wrap things up with a brief summary of the current status of Noyes-McGoveran bit string physics, with bibliography.

Volume 2 #1

The Hamlet Problem - is Definiteness Real? *by Alex Comfort.*

ANPA West is trying to maintain a balance among articles presenting original work, background, and commentary. This article is a brief commentary on some of the points made in the last issue by Blizard, Pstruzina and Etter.

Member Theory. *by Tom Etter*

Wayne Blizard's 'A Theory of Shadows' in our last issue introduced us to multi-sets. Our final article is a follow-up to Blizard's article, providing more background on multi-set theory and how it differs from set theory. Finite set theory, finite MST, and finite MSTZ are each shown to correspond to a different simple notation for numbers, and the old question "What is a number?" is raised in a new form. Multi-set theory looks like it will be a useful tool in the new approaches to quantum theory, and we plan to publish more articles exploring its connections with probability theory, amplitudes etc.

Abolish Infrared Slavery *by Pierre Noyes.*

Next we have a progress report on Noyes-McGoveran bit string physics. McGoveran's approach to calculating the binding energy of the hydrogen atom has been extended by Noyes to the strong interactions. Unlike perturbation theory, this approach does not require that the coupling constants be small for the calculations to yield sensible results, which may be a step towards ending the so-called "color confinement" of quarks.

Don't Worry, Be Happy *by Niklas Damiris.*

There is a story about Metternich and Talleyrand, those great arch-rivals of devious diplomacy, that when one of then, I think it was Talleyrand, heard the other had died, his first reaction was "Now, I wonder why he did that?". Primitive religion, like Talleyrand in his moment of absent-mindedness, naively fills the course of natural events with human purposes. Primitive science, which is what the next generation (if there is one) will call what we practice today, commits the opposite folly, naively stripping the world of purpose altogether. This strange folly, one might almost say this dementia, leads us to treat the world, including even ourselves, as if it were an instrument of our purposes and nothing else. We've heard a good deal lately about the deadly threat that such an attitude poses to our environment. This article addresses the more insidious and perhaps even more deadly threat that it poses to our humanity. Damiris shows that the mechanistic folly is much harder to escape than most "new-age" people imagine. Those of us who like to think we are creating a non-mechanistic science had better pay close attention.

Boolean Fact Sheets. *by Tom Etter.*

Comfort's article raises several questions about logic: does it apply to the world itself?, is there an "old" logic that must be replaced by a better "new" logic? etc. Logic has been and will probably continue to be a recurrent theme here, so we have included next some elementary mathematical background old-fashioned Boolean logic. Many people don't realize how intimately and rigidly the Boolean operators AND, OR, NOT etc. are bound together. You can't make a small change in logic; changing anything changes everything. The main facts of how the logical operators define each other are here briefly summarized, along with the relationship of Boolean algebra to the bit string space of the combinatorial hierarchy.

Volume 2 #2

Schrodinger's Cat and the Cheshire Cat - Quantum Mechanics and Laws of Form.
by Louis H. Kauffman

The new science now on the horizon will undoubtedly count the English logician G. Spencer Brown as one of its early pioneers. Lou Kauffman, a mathematician best known for his work in knot theory, here gives us an account in words and pictures of Spencer Brown's main ideas, and perhaps more important, of his pioneering style of thinking, in the context of the revolutionary changes that quantum theory is making in our conception of logic.

Inertia and Tao *by Tom Etter.*

Quantum mechanics is like the faint but persistent note of an alarm clock trying to rouse us from what Blake called Newton's sleep. Actually, Newton got a bum rap here - Newtonian mechanics was itself a fainter version of the same alarm clock, trying to wake us from what really should be called Aristotle's sleep. It didn't succeed at all; we skillfully managed to blend its quiet and harmonious note into our Aristotelian dream. This article is an attempt to turn up the volume.

Basic Issues Concerning the Relationship Between Consciousness and the Physical World. *by Jean Burns*

Jean Burns is a physicist currently doing research on consciousness. Together with computer scientist Ravi Gomatam she has recently founded an ongoing Consciousness and Science Discussion Group which holds regular public meetings at the Langly Porter Institute that could be of interest to our bay area readers (See News and Events). In this article she gives us some background on four of the major issues that dominate current thinking about the mind-body problem.

Volume 2 #3

The Primary Algebra of Spencer-Brown is Non-Boolean *by Louis H. Kauffman*

An excellent brief introduction to Laws of Form, with a valuable discussion of its relationship to exclusive OR. Be sure to read the addendum on expression algebra.

Are the Laws of Form Non-Boolean? *by Tom Etter.*

Spencer-Brown's primary algebra is developed in a slightly different way; one is lead to a non-Boolean logic almost identical to quantum logic, except that the binary scalar field replaces the complex scalar field, and a negation operator on the vectors replaces the inner product. The essential message here is that quantum logic and its variants are incomplete stages in the growth of Boolean logic out of bare negation.

Construction of the Dirac Equation *by Pierre Noyes.*

A summary of a very significant development in linking the "primary level" of Spencer-Brown theory to the advanced level of particle physics.

A Comment on the Combinatorial Hierarchy *by David McGoveran.*

McGoveran is continually amazed by the abysmal state of modern science. Others of us who are intermittently so amazed may derive courage from this brave polemic which he delivered to the Light Hearted Philosophy Group at Stanford.

President's Message *by Fred Young.*

Fred Young is the current president of ANPA West and the president elect of ANPA international. Here he surveys the diverse activities of our membership and reminds us of the common themes in our various approaches that might facilitate our intercommunication.

Volume 3 #1

Anti-Gravity. *by Pierre Noyes*

What are the consequences if currently planned experiments show that the anti-proton falls up? Mega-technology and the meaning of quantum mechanics come together in this fascinating glimpse of our possible future in space.

Zeno Ball. *by Tom Etter.*

A short note on Zeno's paradox in a perfect Newtonian world.

Strict Finitism Meant to Please the Anti-finitist

by Jean Paul Van Bendegem

Finitary mathematics is an important topic for ANPA members, since finite methods have replaced continuum mathematics in the ANPA bit-string reconstruction of physics. What are the limits of this kind of replacement? Does science have any need at all for the mathematics of infinity? Does mathematics itself even need infinity? The answers offered in this imaginative article will surprise you - to say more would give away the plot.

Racter Report #1: Acausality. *by Tom Etter*

Are people computers? Is logic relative? Can causality go backward in time? What's beyond quantum mechanics? These themes are woven together in this first report on a new version of quantum logic that is closely related to bit-string physics. The Racter here is also the Racter who will preside over the interactive journal.

Volume 3 # 2

What Are We Willing to Take for Granted?

by John Dobson.

John Dobson is the founder of an unusual group known as the San Francisco Sidewalk Astronomers whose mission is to give ordinary people everywhere a chance to see the heavens through high-quality telescopes. He himself started building telescopes while a novice at a Vedanta monastery, and became well-known for his twelve inchers with their so-called Dobson mounts. For a while he and one of his twelve-inchers became a fixture at the intersection of Jackson and Broderick streets in San Francisco, through which anyone could look on a clear night. Later he and his associates built some larger telescopes which they carried around the country to national parks to avoid the glow of city lights; by now over a million people in the US and Canada have looked through them. Some of you may have heard John's talk at the last ANPA West meeting; this paper is in the same vein.

Remote Citizens of the Moon, Part I
by Tom Etter.

As mentioned in the last issue, we want to branch out in new directions. This is the first of a two-part series on a plan for colonizing the moon and beyond.

Volume 4 #1

The Dialectics of Freedom. *by Pierre Noyes*

Does quantum mechanics contradict classical determinism? This question has been argued since the 1920s; in this article Pierre Noyes tackles it from a contemporary vantage point and stands the usual answer on its head.

Remote Citizens of the Moon, Part II: Moon City. *by Tom Etter.*

Continuing his thoughts on a future in space, Tom Etter proposes a novel scenario for our first space colony. This is part two of what has been expanded into a three part article.

On Haley's Comet Coming Back (a poem). *by Herman Mueller.*

What is ANPA? *by Fred Young.*

Volume 5 #1

Towards Solving the Mind-Body Problem. *by Emmanuel Ransford*

This essay presents a dualist account of the mind-body relationship that relies heavily on quantum mechanics for its guiding ideas. Unlike many current quantum theorists of consciousness, Ransford is not afraid to follow consciousness all the way down to the elementary particles.

Empathy *by Tom Etter.*

One normally classifies mind-body theories as either dualist or monist. This essay presents mind-body as a polarity like up and down, very asymmetrical in everyday life but symmetrical in the larger picture. Instead of minds and bodies there is a mind-body field whose mathematical structure involves the "curvature" of logic. Despite this very different starting point, there are some significant points of agreement with Ransford.

Knowledge, Love and Happiness. *by Niklas Damiris*

The pioneers of modern psychology patterned their new science of mind on physical science, hoping to duplicate its successes by adopting its methods. There is a growing body of thought today, especially in Europe, that sees this as a great mistake. Psychology is about us, and to divorce it from the big philosophical questions about who we are and how we should conduct our lives is to trivialize its subject matter. Damiris here tackles the big questions. Taking his point of departure from Foucault's reflections on individuality and society and Lacan's reinterpretation of Freud, he challenges us, all of us, to reinvent ethics. This and the previous essay, though their subject matters may seem quite different, are actually much of a piece.

Modern Metaphors For a New Millennium. *by Herman Mueller.*

Another strand of thought that has recently entered our journal is our future in space. Here's a poet's reflection on what we are up to.

Volume 5 #2

Objectivity and Huge but Finite Numbers by *G.F. Chew*

Berkeley physicist Chew, whose work on S matrices became known to the public as Bootstrap Theory, offers us a glimpse of his latest thinking about the deep problem of bringing together our theories of the very large and the very small into a coherent whole.

Peeking at the Conscious Brain: New Clues, New Challenges by *Emmanuel Ransford*

This and the following paper are sequels to the authors' contributions in our last issue.

Looking At, Through, and With: Comments on Ransford by *Tom Etter*

Book Review: **Eddington's Search for a Fundamental Theory: A Key to the Universe** by C. W. Kilmister, Cambridge, 1994 *Reviewed By Pierre Noyes*

A letter to the Editor of ANPA WEST *from A. K. Kwasniewski.*

Volume 6 #1

Discrete Ordered Calculus by *Tom Etter & Louis H. Kauffman*

The purpose of this short paper is to give a quick introduction to the discrete ordered calculus devised by Louis Kauffman and Pierre Noyes in their paper on the derivation of electromagnetism from the formalism of quantum mechanics. In fact we improve on this original version of the discrete ordered calculus by introducing a fundamental time shifting operator that is distinct from the time shift associated with a derivative in the original calculus.

The Abandonment of Simultaneity by *Pierre Noyes*

Some recent investigations in linguistics, communication, and social organization have found that progress can be made only by abandoning the concept of simultaneity in favor of a multi-component hierarchical description of overlapping times. It is suggested that the same approach might offer a clue to the solution of the problem of joining relativity theory to quantum mechanics which has resisted more conventional approaches for forty years.

Poem: To Hell with Education A Denunciation of Scholasticism in Science

With Apologies to C.S. Lewis' 'The Screwtape Letters' By *Viv Pope*

Boolean Geometry and Non-boolean Change By *Tom Etter*

The following paper was the basis for my talk at ANPA 16. At the time I believed its basic idea to be original, a belief supported by a number of knowledgeable readers. However, I subsequently learned that this idea, which I called Boolean geometry, was actually around as early as the 1930's. Having never quite made it into the mainstream of logic, it was reinvented not only by me but by several others, including Gordon Pask and, I believe, Gian-Carlo Rota - the full history here remains to be uncovered. However, the connection I made to negative quantum amplitudes does appear to be new, so my plan became to revise the paper into a larger work in which this connection is developed in detail. This larger work has indeed become larger! What was to be only an introductory section on link states turned into my 90-page IJGS paper, followed by a series of shorter papers on the same topic, and I'm afraid the grand synthesis of link theory and Boolean geometry is still only a sketch. Thus it seems like a good idea to go ahead and release this paper in its present unfinished state. Like the DOC paper, it's a historical record, and also I believe that it's not a bad introduction to its subject, which could well be of interest to other investigators of "strange" logics.

Volume 6 #2

Integral Science by *Tom McFarlane*

Our first article is by someone who balances his practice of science with the practice of more traditional disciplines based on an inner approach to reality. The author makes an eloquent case for the possibility of achieving such a balance within a single discipline.

How To Take Apart a Wire, Part I by *Tom Etter*

This paper, which was prepared for the PhysComp96 conference on physics and computation, is an exposition of link theory for those more at home with computers than with quantum mechanics. Part I starts with Hilbert space quantum mechanics and ends with diagrams of quantum measurement that exactly match those from link theory. Part II will go the other way, developing quantum theory "from the bottom up."

Natural Interactivity by *Helgi-Jon Schweitzer*

The author is a psychologist who has for many years studied the synchronization and entrainment of biological rhythms, both in individuals and in groups. This work has led him to novel ideas about the nature of interactivity, which he sees as something fundamental that is irreducible to a back-and forth exchange of information or causal influence. After giving us a survey of many intriguing experimental results, he offers some speculations and warnings on the possibly disrupting effects of new media and channels of communication on the natural rhythms that maintain the social order.

Schrödinger's Cat and the Cheshire Cat

- Quantum Mechanics and Laws of Form

by Louis H. Kauffman

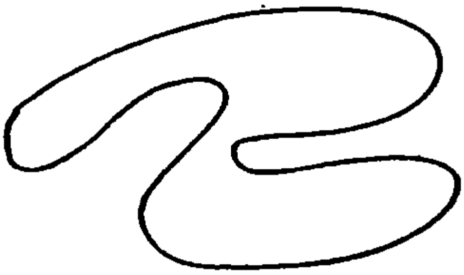
I. Introduction

It is the purpose of this note to point out a deep relationship between the quantum mechanical viewpoint and the fundamental context of Laws of Form [4]. In the course of this discussion we shall also see the issues of indistiguishables [3] and the combinatorial hierarchy [1] in a new light.

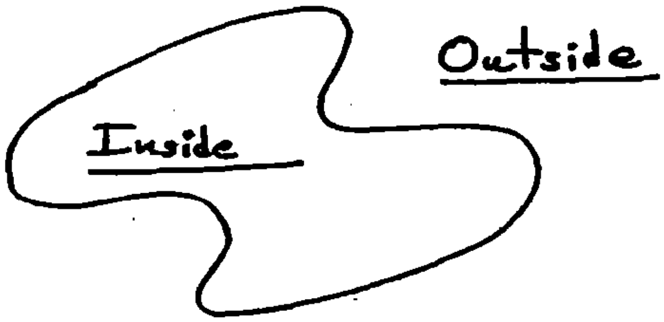
This note is only a sketch of these relations. It is a first attempt to articulate matters that dive below the surface formed through words. A raid on the inarticulate is a foray into the inchoate.

II. Descent into the Form

Consider a distinction.



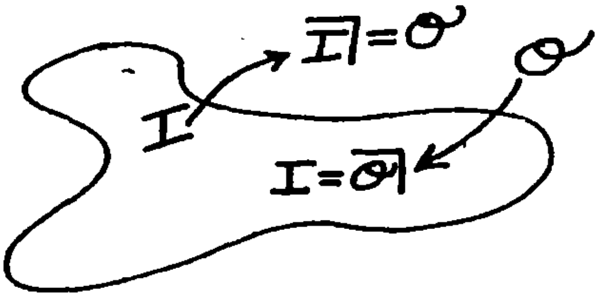
Let the distinction be drawn and
Let it be seen
To sever a space
Into
Inside and
Outside.



Let these states
of inside and outside
Be denoted
By I (for inside)
And O (for outside).

And let us agree upon
A transformation
$X \rightsquigarrow \overline{X}\rceil$
That takes
Inside to Outside
And
Outside to Inside.
And so we have
The equations
$\overline{I}\rceil = O$
$\overline{O}\rceil = I$.

Crossing from inside yields outside.
Crossing from outside yields inside.

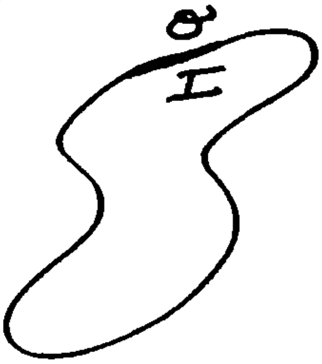


Here we have the
Language of
Not.
Not this, then that.
Not outside, then inside.

And O and I are surely names.
And ye could surely let us
Rest a bit
From the
Rigors of Infinite Repetition
By
Agreeing
That O being <u>identical with</u> O
(In the first place)
We could allow a bit (a wee bit)
of
<u>Condensation</u> with
$$\overline{O\,O} = O$$
and
$$\overline{I\,I} = \overline{I}\,.$$

This can do no harm,
But if we would be counting then
Thee must re-think these
Colapses from infinity.

Be that as it may,
Here we are
In the very simple language
Of a first distinction.



$$\overline{\theta}\rceil = I \qquad \theta\theta = \theta$$
$$\overline{I}\rceil = \theta \qquad II = I$$

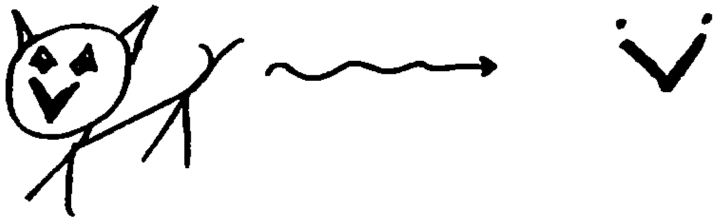
No quantum logic here.
This is the boolean realm.
Nor is this Laws of Form.
No, not yet.

But look here.
Ye had best worry and worry well
On these collapses of identicals.
What gives here?
Can $\theta\theta = \theta$ if θ and θ are
Yet the same.
Can one yield two and still be one.
Aye 'tis a giant step that
And I'd not be trusting these

Mathematicians
Who think that
All you have to do
To count
Is just write down your
Rows of identicals that
Are not identical.
/, //, ///, ////, /////, //////, ...
The whole world of definite
Solid
Clear
And perfect
Arithmetic
Is built on
A consideration of
A stroke
/
That
Is itself /
And
Is not itself //.
The very
Economy
Of our speech and reference)
Is indivisibly woven
In the fabric of condensation.
In the context of
Fusion of
Likes and Nearly likes and even Un likes.

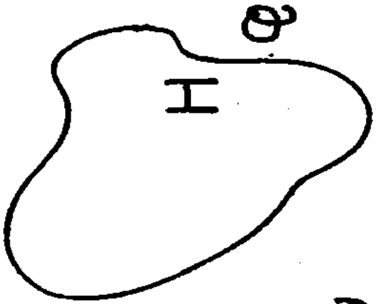
We walk above the
Shifting sands of
Chaos and Paradox.
And
Imagine
Solid Ground.

And yet there is something
Definite in this speech
And thee may suspect
That 'tis a creation
Of the definite reference
of Θ' and I.
Let us embark on the
Story of the
Cheshire Cat.
(A precursor to Schrödinger's cat)
That Cheshire cat was
Neither alive nor dead
When he became
Nothing but
His smile.



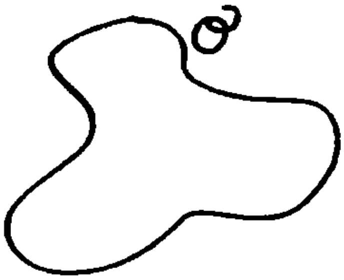
The slightest mark will do.

And so we shall
Eliminate the I.



$\overline{\vartheta\,|} = I$ $\qquad$ $\vartheta\vartheta = \vartheta$

$\overline{I\,|} = \vartheta$ $\qquad$ $II = I$



$\overline{\vartheta\,|} =$ $\qquad$ $\vartheta\vartheta = \vartheta$

$\overline{\quad|} = \vartheta$ $\qquad$ $=$

But this is
Half-way down.
Ye still can see ’his body
Through the fog.
The outside disappears as well
If inside you no longer tell.
But if the transform you will
Keep,
Then outside lives

Above the deep
As that which is
Not void.
And so we see
Writ large and clear
In these equations
For our ear

$\overline{\vartheta\,\rvert} =$: not out is void
$\overline{\rvert} = \vartheta$: not void is out
$\vartheta\vartheta = \vartheta$: out out is out
$=$: void is void

No point in keeping ϑ around.
We'll just allow him to sink
Into the ground
Of transformation
$\overline{\rvert}$
And we find

$\overline{\overline{\rvert}\rvert} =$
$\urcorner\urcorner = \urcorner$

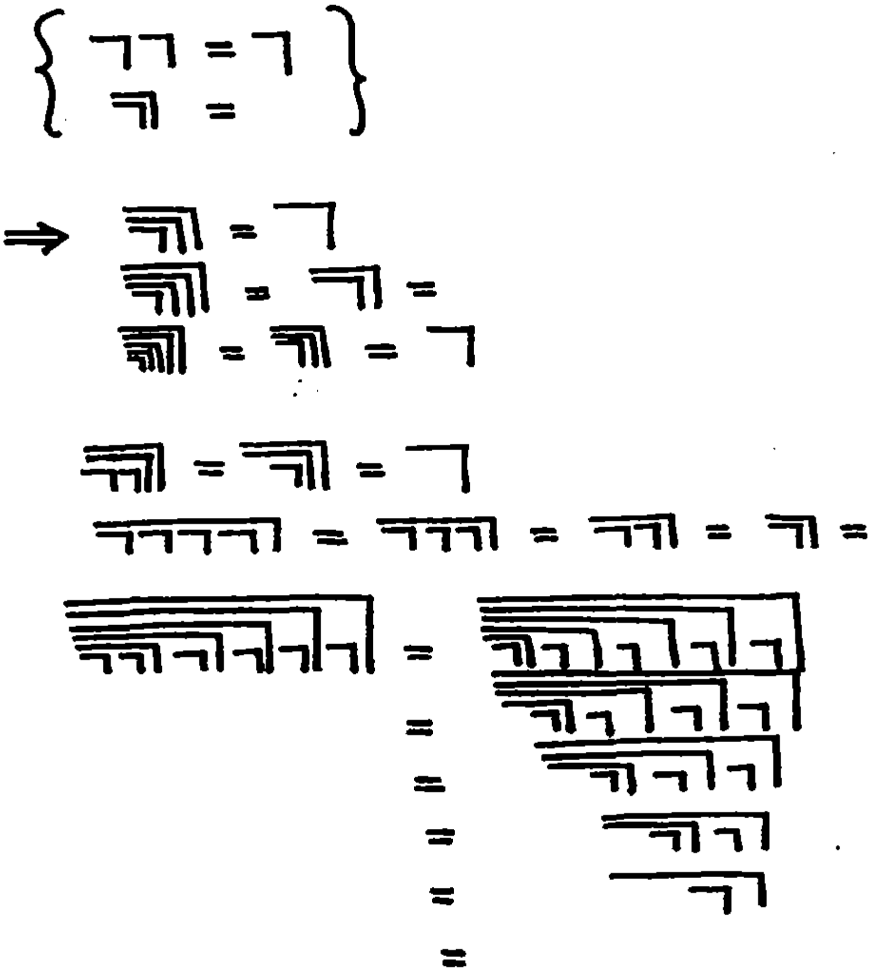
The primary forms of
Indication's kind.
A mark ($\urcorner$) to mark a side.
A mark ($\urcorner$) to indicate the passage from the void.
A harmony of language on

Arrival from
The other side.

$\overline{\quad}| = \overline{\quad}|$: cross from void equals marked
$\overline{\overline{\quad}|}| =$: cross from marked equals void.
$\overline{\quad}|\overline{\quad}| = \overline{\quad}|$: marked marked equals marked
(in the first place)

This completes the first descent —
Into the form.
"We take therefore the form of
distinction for the form."
It is not the mark ($\overline{\quad}|$) that —
Marks this place, but the
Process of descent.
The process of going down,
Down into the
-east saying
Before
Silence
Overwhelms
The urge
To speak.
"Whereof we cannot speak we
must [choose to] remain silent."

III. Abandon Form to Indication

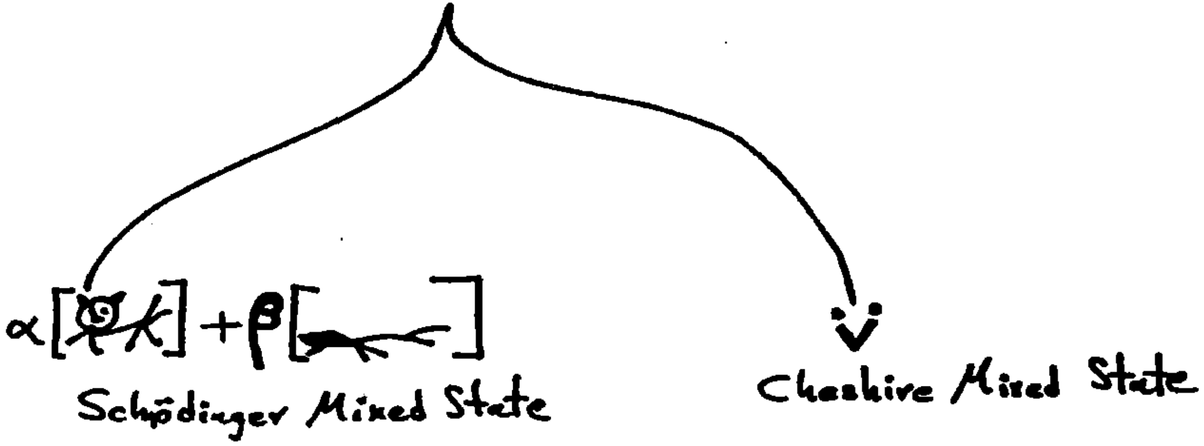


IV. Abandon Indication to Void

V. Schrödinger's Cat

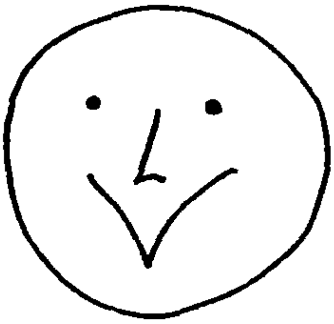
This famous cat has gone well beyond the strictures of boolean logic. Unobserved, this catus is in a state of possibility. Not alive, not dead, described by a wave function, not-pinned-down and yet ready to spring forth into an eigenstate at the behest of an observer who opens the box.

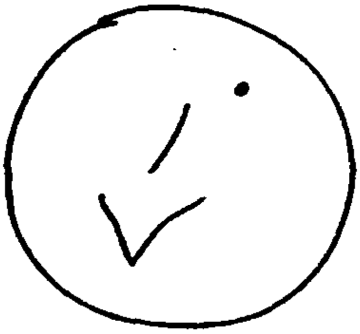
The place fundamentale of confluence between the quantum scenario of the cat Schrödinger and the cat Cheshire is in <u>ambiguity</u>.

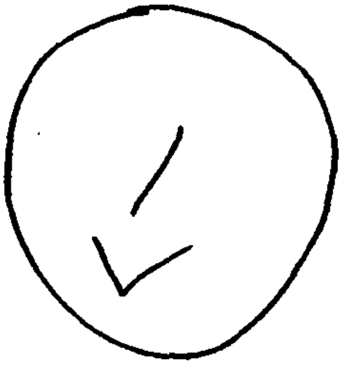


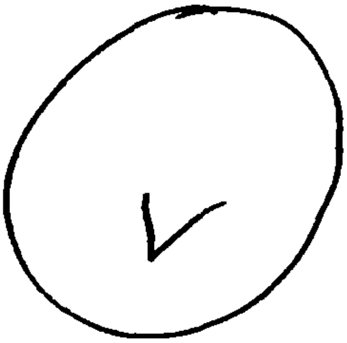
In the <u>Schrödinger mixed state</u> we live way <u>above</u> $\overline{I}/O$ and indicate the confusion of possibilities by a linear superposition in Hilbert Space: $\alpha I + \beta O$.

In the <u>Cheshire mixed state</u> we live way <u>below</u> $\overline{I\,\rceil}\,O$ and indicate the possibility of possibilities by a mark $\rceil$ regarded as the smile of the cat. $\vee$.









A Cheshire state unfolds into its multitude of "real" possibilities. The power of Laws of Form lies in its ability to evoke fundamental Cheshire Cat states near (as close as humanly possible) the first distinction.

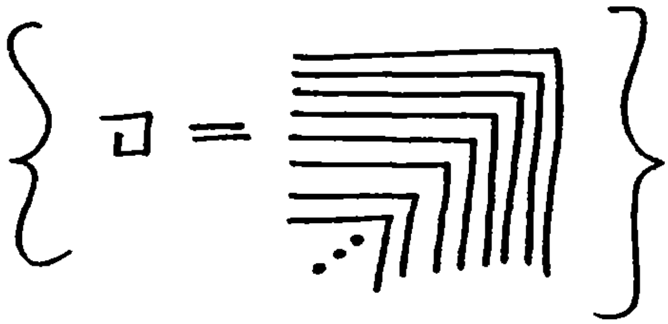
As we go back up into articulation from Cheshire state the cat becomes real once again, but the process of becoming real is organic unfoldment, not mechanical eigenprojection. As quantum mechanics matures, it will become apparent that there is a whole world of quantum mathematics and quantum description that intermediates between the pure Cheshire world and the pure Schrödinger world.

A simple example of this intermediation is the emergence of eigenforms, the simplest being,

$$J$$

with

$$\overline{J}| = J$$



Here J is a "vector" of eigenvalue one for the operator $\overline{}|$. But J is a vector in a Hilbert space of forms and ideas, the non-numerical explosion from the void occasioned at the cusp of paradox/time/self-reference.

VI. Quantum Logic

The first step in classical quantum logic is the orthogonality structure of subspaces of a vector space. Let V be a vector space (say $\mathbb{R}^n$) with a standard inner product, hence notion of perpendicularity. Let $a \cdot b = 0$ mean a perp to b, and write

$$a \perp b \iff a \cdot b = 0 .$$

If $W \subset V$ is a subspace, let $\overline{W|}$ denote the subspace perpendicular to W:

$$\overline{W|} = \{ v \mid v \cdot w = 0 \quad \forall w \in W \}.$$

Then

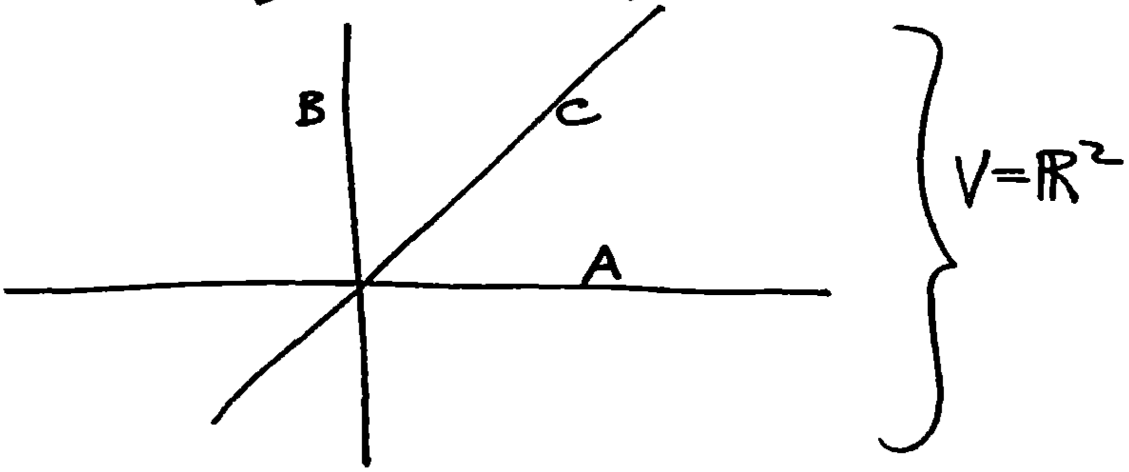
$$\overline{\overline{W|}|} = W$$

for any subspace $W \subset V$.

Let WW' denote the subspace spanned by W and W'.

Thus $WW' = \{w + w' \mid w \in W$ *and* $w' \in W'\}$. Also, let $W \cap W'$ denote the intersection of W and W' *and* (by analogy) let $W \cup W' = WW'$. Thus $\cup$ is *not* set theoretic union.

We then see that this algebra of subspaces is decidedly non-boolean. For example



let V be the plane, $\mathbb{R}^2$. Let A, B and C be the one-dimensional subspaces depicted above. Then $A \cup B = V$, $A \cap B = \phi$ (ϕ = the 0-subspace). But

$$(A \cup B) \cap C = V \cap C = C$$

$$(A \cap C) \cup (B \cap C) = \phi \cup \phi = \phi .$$

Thus *the distributive law does not hold*.

While $\overline{A}\rceil = B$, there are infinitely many intermediate states C. While each such state

C is <u>generated</u> by A and B (Since $A \cup B = V$) these intermediate states are nevertheless completely distinct from A or B taken seperately. It is only through the cooperation of A and B that these intermediate states arise.

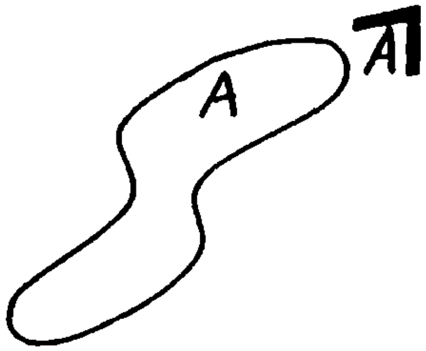
This is exactly the situation in the logic of quantum mechanics that gives rise to complementarity, the catus Schrödinger, non-locality and all manner of seemingly strange things.

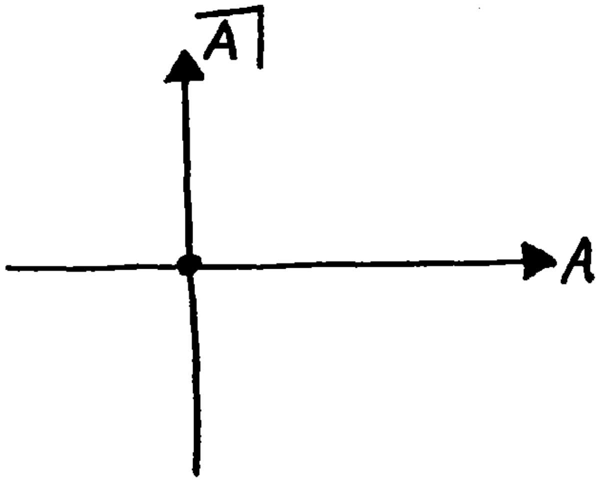
Nevertheless, we can understand this logic through the patterns of geometric orthogonality, and via the understanding that

A and $\overline{A|}$

denote fully complementary spaces in a wider space. That A and $\overline{A|}$ taken together generate a new space of even greater possibility than either taken seperately. This is the moral of negation in the quantum logic.

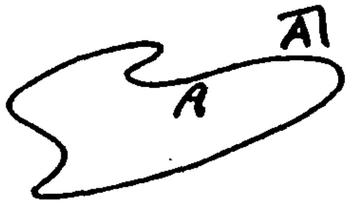
How are we to see this extra dimension in the picture of two sides?





All that space in-between,
Is it real
Or is it
Imaginary?

We must begin to
See the
Usual
Distinction/Form



As a
Cheshire Cat.
Apparent exclusivity
Is only the
Smile.
The
Whole is
Much Greater
Than the parts.

The models of
Orthogonal Subspaces
Or
Non-boolean lattices
Helps us
Follow a
Precise way
Into
Spaces of
Possibility.

Return now to
The Mathematics.

<u>Theorem</u>. $A, B \subset V$ subspaces.
Then

$$A \cap B = \overline{\overline{A|}\,\overline{B|}\,|}.$$

<u>Proof</u>.

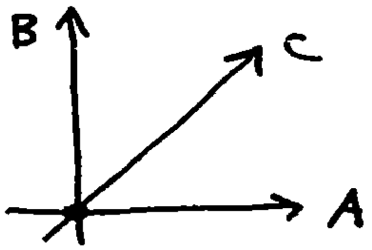
$$\begin{aligned} v \in A \cap B &\Leftrightarrow v \in A \text{ and } v \in B \\ &\Leftrightarrow v \in \overline{\overline{A|}\,|} \text{ and } v \in \overline{\overline{B|}\,|} \\ &\Leftrightarrow v \perp \overline{A|} \text{ and } v \perp \overline{B|} \\ &\Leftrightarrow v \perp \overline{A|}\,\overline{B|} \quad \text{(check it!)} \\ &\Leftrightarrow v \in \overline{\overline{A|}\,\overline{B|}\,|}. \end{aligned}$$

<u>QED</u>

Thus, we see that this
Algebra of Subspaces
Can be expressed in a
Formalism

Quite similer to the
Algebra of Laws of Form.
But the lawof transposition fails:

$$\overline{\overline{A}|\,\overline{B}|}|\,C \neq \overline{\overline{AC}|\,\overline{BC}|}|$$



Other laws hold, and this
Heralds an
Investigation of
Axiomatics.
(To be done elsewhere, but note
that $\overline{\overline{A}|B}|A = A$, $\overline{\overline{A}|A}| = \phi, \ldots$)

But now it is
Time for
Tea.
New cups,
New cups.
Move down,
Move down.

VII. Beneath Quantum Logic

Let the zero subspace ϕ
Be denoted by
Void.

Then $\overline{V}\rceil = \phi \qquad VV = V$

$\overline{\phi}\rceil = V \qquad \phi\phi = \phi$

Becomes

$\overline{V}\rceil = \qquad VV = V$

$\rceil = V \qquad =$

A familiar scenario.

But now we are
Invited
To
Identify the
Operation of <u>ortho-complementation</u> ($\rceil$)
With the
Whole space (V).

We are invited to identify the <u>operation</u> of complementation with the whole space. We are invited to find that the space of possibility is <u>itself</u> the transformation into the complementary point of view.

Space in and of itself—
Can no longer be regarded as
An object.
Space is dynamic.

Space is the transformer of the Void.
And,
(Unfolding the Cheshire Cat)
Space (Whole space)
Acts
Upon its (Subspaces)
To produce
The logic of orthocomplementation in
A
Non-boolean mode.

In the land of the
Cheshire Cat,
At the level of
The Smile,
There is only
Space/Transformation /$\overline{\;\;}|$
and
Void.
In this rarefied form of
Indication,
The distributive law holds:

$$\overline{\overline{\overline{\;\;}|\,\overline{\;\;}|}\,|}\,| = \overline{\overline{\;\;}|\,\overline{\;\;}|}\,|\,\overline{\;\;}|\,.$$

And the primary algebra of
Laws of Form is
Seen as
The smile of the cat,
As Schrödinger's cat
Dissolves to Void.

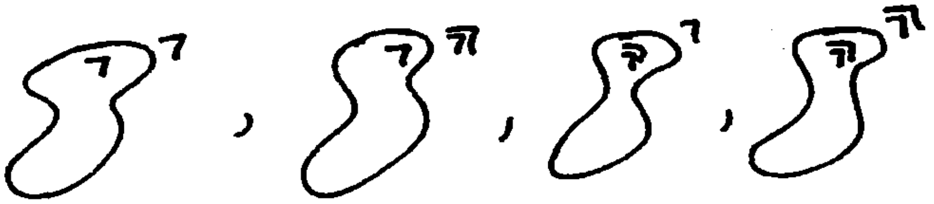
The space of quantum logic
Is a space distinguished by
An observer.
The observer and the space
Are,
In the Form,
Identical.
It is a fantasy of observers that
There is great importance to the
Distributive, boolean algebra of ⌉ alone.
We have seen that the
Importance
Lies (sic)
In the
Potent
Condensation
Of a
Smile.

III. Indistinguishables

⌉ and ⌉ are indistinguishable.
And yet,
⌉⌉ =
While
⌉⌉ = ⌉.
Distinction is relative in the
Realm of
Languages.

VII. Hierarchy

Here the program is to closely abide in the hierarchies arising from distinctions that arise from given realms of distinction. To use vectors and matrices (as in the Combinatorial Hierarchy [1]) is fine but needs to be slowed down so that we see the genesis of these very structures (vectors, matrix mappings) from the most primitive levels. I beg off doing this here, but point out that if we are given a distinction with evaluations $\urcorner$, $\overline{\urcorner}$ for its sides



then there are four possibilies (both sides marked, one marked, other marked, both unmarked) and transformations of these states arise. (Flip one side, Flip other side, Interchange, Leave alone) And this is a short step from the Pauli matrices, spinors, quaternions and the first levels of the Combinatorial Hierarchy. Matrix algebra unfolds from the smile of the Void.

References

1. Bastin T., Noyes H.P., Amson J., Kilmister C.W. Int. J. Theo. Physics 18 (1979).

2. Carroll L. *Through the Looking Glass* in *The Works of Lewis Carroll* (ed. by Roger Lancelyn Green) London: Hamlyn (1965

3. Parker-Rhodes, A.F. *The Theory of Indistinguishables* Reidel, Dordrecht (1981).

4. Spencer-Brown G. *Laws of Form* The Julian Press Inc., New York (1972).

Department of Mathematics, Statistics and Comp. Sci.
Univ. of Ill. at Chicago
Chicago, Ill. 60680

April 8, 1990

COMMENTS SOLICITED

Event Spaces and Their Linear Logic

DRAFT

PRELIM. VERSION OF PAPER SUBMITTED TO A CS CONFERENCE

Vaughan Pratt
Computer Sci. Dept., Stanford, CA 94305, USA
pratt@cs.stanford.edu

February 7, 1991

Abstract

An event space is a partially ordered set with a top (empty meet) and all nonempty finite and infinite joins. Among the many ways to equip posets with various flavors of meets and joins, event spaces attain a certain natural optimum with regard to both applications and algebraic structure.

In the motivating application they represent concurent behavior by coding concurrent events as points, the concurrent combination or conjunction of events as nontop joins, and their conflicting combination or disjunction as top joins. This is in contrast to a Boolean algebra which codes disjunction as join but conjunction as meet.

A striking algebraic property is that the dual of an event space is a similar event space of the same cardinality and having as its dual the original space, a double-duality principle akin to that of vector spaces but that also holds for spaces of infinite dimension. Like vector spaces, event spaces have both direct and tensor products, with similar applications. However whereas direct product and tensor product of vector spaces are each self-dual operations, event spaces have a direct sum and a tensor sum distinct from their corresponding products. This endows event spaces with a rich calculus that can serve as a parallel programming language with a built-in logic.

1 Introduction

In depicting as a graph a process such as the steps in manufacturing a part, the question arises as to whether to take the vertices of the graph to be the states or the events of the process.

Choosing states commits one to an automaton view in which the events then appear as the edges of the graph. Progress in an automaton can be indicated by a path traced along the edges of the graph starting at a designated initial state.

Choosing events yields the schedule or PERT chart view. In this case the edges denote temporal constraints between events. This might simply be event order, such as the rule that one may not have children before getting married. Or it might be more detailed delay information, such as the requirement that the data line not be read less than 3 nanoseconds or more than 10 nanoseconds after the strobe line has been asserted. Progress in a schedule can be indicated by checking off events as they take place.

Whereas computer scientists tend to be attracted to automata, project managers seem to prefer schedules. This raises such questions as which is preferable according to what criteria, and how much does one give up in making a choice either way? A prerequisite for answering these is a detailed understanding of the relationship between automata and schedules, see e.g. [NPW81, vG90, Pra91].

Reprinted from Proc ANPA West 7 (1991)

This paper describes a representation of behavior called event spaces. These are nominally for representing schedules. As with some other representations of schedules, e.g. partial orders [Gra81, Pra82], and their elaboration to event structures [NPW81], to every event space corresponds an equivalent automaton. Unlike these other representations however, the automaton or state space $a^{\perp}$ corresponding to an event space a is very similar to a. For the most part events are simply reinterpreted as states. However the final event of the event space is missing, and in its place is an initial state. And whereas the basic operation of an event space is concurrence of events, represented as join or least upper bound in the ordering of events, the basic operation of a state space is the dual notion of choice of states, represented as meet or greatest lower bound in the same ordering (modulo the switch of endpoints).

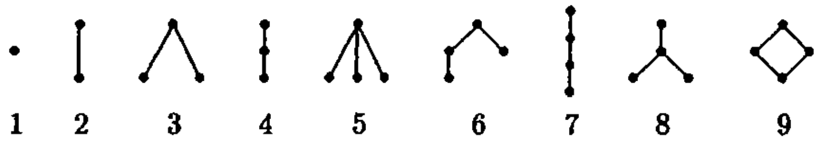
Event spaces simultaneously subsume, extend, and simplify the Nielsen-Plotkin-Winskel notion of event structure [NPW81] and hence its dual notion of prime algebraic domain as the corresponding automaton. And although event structures have a nice calculus of operations constituting a parallel programming language, event spaces add several useful operations to this language to extend it to a *logic* of concurrency.

Event spaces behave very much like vector spaces. In particular they admit linear transformations: the set $a \multimap b$ of linear transformations from event space a to event space b itself forms an event space. However event spaces improve on vector spaces in three ways. First, the duality of vector spaces by which they are isomorphic to their double dual holds only for finite dimensional vector spaces, but holds for all event spaces, whether of finite or infinite dimension. Second, vector spaces are rigid and require the addition of a norm to give them some flexibility, whereas event spaces are inherently flexible without the addition of a separate norm. Third, sum and product of vector spaces degenerate to the same operation, and similarly for tensor sum and tensor product, whereas for event spaces these are all distinct operations, respectively concurrence and choice, and cointeraction and interaction. These operations are among those of linear logic [?], for which event spaces constitute a particularly simple and natural model.

2 Event Spaces—Definition and Representation

An *event space* $(X, \bigvee, \infty)$ is a set X which can be partially ordered by a relation $\leq$ so as to make $\bigvee$ the join operator defined on all nonempty subsets of X, with ∞ the top element of X and so satisfying $x \leq \infty$ for all $x \in X$. We abbreviate $\bigvee\{x, y\}$ to $x \vee y$. $\bigvee$ uniquely determines $\leq$ via the relationship $x \vee y = y$ if and only if $x \leq y$.

The following Hasse diagrams[1] depict the nine event spaces having up to four points (events).



Event spaces are analogous to vector spaces. Just as every vector of a vector space is uniquely

---

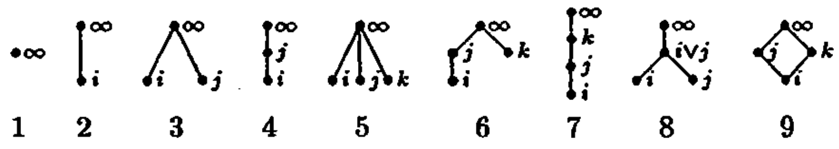
[1] A *Hasse diagram* depicts a partially ordered set as an undirected graph whose edges have an implicit upward orientation. Hence $x \leq y$ is represented by the existence of a path from point x leading upwards to point y. Since posets are reflexive there is an implicit self-loop at every vertex, and since they are transitive there is an implicit edge from x to z for every upward path $x \leq y \leq z$. Deleting y from the poset does not mean that $x \leq z$ no longer holds but rather that the implicit edge from x to z in the Hasse diagram now needs to be made explicit.

representable as a linear combination of basis vectors, so does every event of an event space have a representation.

However event spaces are less degenerate than vector spaces. One consequence is that a distinction emerges that did not need to be drawn for vector spaces, namely between the syntactic representation of an event as a term, and its semantic representation as a set.

In the term representation the top event is denoted by the constant ∞ (top), and all other events are each denoted by the join of the set of generators below that event. A *generator* is any event that is neither ∞ nor the join of two incomparable events.[2]

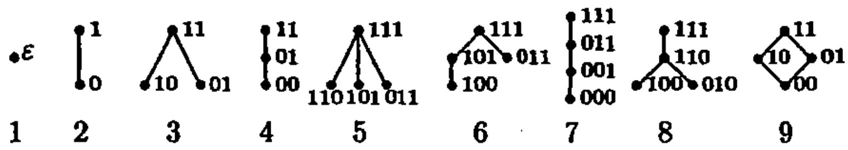
The event spaces depicted by the above diagrams have up to three generators. If we call these respectively i, j, k we may label events with terms as follows.



This representation is by no means the most succinct possible. For example if to space 8 we adjoin a third generator k sibling to i and j then our representation assigns to the event just below ∞ the term $i \vee j \vee k$. Yet this event could be represented more succinctly as the join of any two of these three generators.

In the set representation an event x is represented as the set of all *representatives* r for which $x \le r$ does not hold. A *representative* is any event that is neither ∞ nor the meet of two incomparable events. In this representation $x \vee y$ is represented by the union of the representations of x and y, and ∞ by the set of all representatives.

The event spaces depicted by the above diagrams have up to three representatives. In the following we label each event with a bit vector whose coordinates (bit positions) are representatives, with the presence of a given representative in a set denoted by a 1 in that coordinate; thus 111 denotes $\{i, j, k\}$. When there are no representatives, as in space 1, the bit vector is empty, denoted ε.



We shall see shortly that there is a dual relationship between these two representations.

3 A Calculus of Event Spaces

Just as numbers combine under addition $x+y$, multiplication xy, and exponentiation x^y, and propositions under disjunction $p \vee q$, conjunction $p \wedge q$, and implication $q \to p$, so do event spaces combine

---

[2] With this definition of generator it may be verified that distinct events are always denoted by distinct terms. Had we defined a generator to be any event x not representable as the join of some possibly infinite set of events other than x we would not always have enough generators to keep denotations distinct, e.g. for the event space consisting of the unit interval $[0,1]$ of reals, standardly ordered.

under *concurrence* $a\|b$ (a form of sum), *choice* $a\#b$ (a form of product), and *linear implication* $a \multimap b$. We will define these in a moment, but first let us look ahead to the rest of the calculus of event spaces that we shall define in this paper.

Common to numbers, propositions, and event spaces are the constants 0 and 1, units or identities for sum and product respectively. For the time being we shall take both 0 and 1 to denote the singleton event space, example 1 in the above diagrams. Later we shall introduce a "sign bit" separating these and making 1 a negative event space or positive state space.

However whereas numbers satisfy $(x^y)^z = x^{yz}$ and propositions $z \rightarrow (y \rightarrow x) = (y \wedge z) \rightarrow x$, the interpretations we give below of $a \multimap b$ and $a\#b$ turn out not to satisfy $c \multimap (b \multimap a) = (b\#c) \multimap a$. This is unfortunate because such a principle is extremely useful for dealing with implications one premise at a time. In the context of Church's λ-calculus this principle is sometimes called Currying, in recursion theory it is referred to as the *s-m-n* theorem, and in computer programming it corresponds to the idea of instantiating one of the arguments of a two-argument function with a constant to yield a one-argument function.

We therefore introduce an additional product $a \otimes b$ which satisfies $c \multimap (b \multimap a) = (b \otimes c) \multimap a$. The unit for this product is the doubleton event space, example 2, rather than the singleton, which we denote $\top$. Corresponding to the above equation for tensor product is the equation $a = \top \multimap a$ for the unit. The generic name for a product introduced in this way is *tensor product*, but in the context of event spaces we shall call it *interaction* because $a \otimes b$ represents the interaction or collision of event spaces a and b.

Negative $\top$ is $\bot$; for now $\bot = \top$ as with $0 = 1$. Although 0 can be viewed as logical false, so can $\bot$, which is "linear false" in that it is the more appropriate constant to use in conjunction with linear implication.

We define linear negation $a^\bot$ as $a \multimap \bot$. Using negation we may establish a number of relationships: $(a\#b)^\bot = a^\bot \| b^\bot$ and dually $(a\|b)^\bot = a^\bot \# b^\bot$; $a \otimes b = (a \multimap b^\bot)^\bot$; and $a \multimap b = (a \otimes b^\bot)^\bot$. By analogy to the first of these we may define a *tensor sum* $a \oplus b$ dual to tensor product, as $(a \oplus b)^\bot = a^\bot \| b^\bot$.

A concept with no arithmetic or (ordinary) logical analog is that of the free event space $!a$, a space formed from a by taking its underlying poset $U(a)$, i.e. "forgetting" that a has joins and a top, and adding additional elements to supply the apparently missing joins and top. Then the maps from $!a$ to b behave exactly like monotone functions from a to b without the requirement that they preserve the joins or top of a. This gives us a new implication $a \Rightarrow b = !a \multimap b$ denoting the event space consisting of such maps, a somewhat larger space than $a \multimap b$.

Free event spaces interact according to $!a \otimes !b = !(a\#b)$. Hence

$$\begin{aligned} c \Rightarrow (b \Rightarrow a) &= !c \multimap (!b \multimap a) \\ &= (!b \otimes !c) \multimap a \\ &= !(b\#c) \multimap a \\ &= (b\#c) \Rightarrow a \end{aligned}$$

This now gives us two complete sets of operations for sum, product, and implication, an ordinary one consisting of $a\|b$, $a\#b$, $a \Rightarrow b$, and constants 0 and 1; and a linear one consisting of $a \oplus b$, $a \otimes b$, $a \multimap b$, and constants $\bot$ and $\top$. Negation $a^\bot$ is defined linearly but is common to both as far as the De Morgan duality of sum and product are concerned.

The dual of $!a$ is $?a = (!a^\bot)^\bot$. If we view $!a$ as having some of the flavor of an existential quantifier or diamond modality then $?a$ is the corresponding universal quantifier or box modality.

We now specify in detail the interpretations of all of the above operations and constants. We distinguish primitive from derived operations. The four primitive operations are $a^\perp$, $a\#b$, $a \multimap b$, and $?a$. All four constants, $0, 1, \bot, \top$, are primitive. The five derived operations are $a\|b$, $a \otimes b$, $a \oplus b$, $!a$, and $a \Rightarrow b$. These are all specified via the equations we have already seen above.

Actually these are not really equations but rather isomorphisms. For example the equation $a\#b = b\#a$ says that product or choice computed in the opposite order gives only an isomorphic space, not the same space. The need for this distinction becomes apparent when considering the projection of these respective products on the first coordinate, which must give respectively an element of a and of b. This could not happen if the products were the exact same space.

The constants have already been specified in sufficient detail: 0 and 1 are (for now) the singleton event space while $\bot$ and $\top$ are the doubleton.

The first operation we shall define is negation, $a^\perp$. Although we could have made this a derived operation via $a \multimap \bot$, this operation is so simple yet fundamental that we shall treat it as a primitive.

An operational but accurate description of $a^\perp$ in terms of a is to forget the join structure and regard a just as a poset. Remove the element ∞, take the order dual of the remaining elements (i.e. turn the remainder upside down), and replace ∞ (still at the top). This yields a new poset $a^\perp$.

Theorem 1 . *The poset $a^\perp$ constructed as above is an event space.*

Proof: The construction automatically equips $a^\perp$ with a top, namely ∞. It remains to show that the join of any nonempty subset Y of $a^\perp$ exists in $a^\perp$. The following closely parallels the argument that a complete semilattice is a complete lattice.

If ∞ is in Y then $\bigvee Y = \infty$. Otherwise consider the set $\mathcal{L}Y$ of lower bounds on Y in a. If $\mathcal{L}Y$ is empty then the only upper bound on Y in $a^\perp$ will be ∞, which is then the least upper bound on Y, the desired join $\bigvee Y$. If $\mathcal{L}Y$ is nonempty then it has a join j in a. Now every element of Y is an upper bound on $\mathcal{L}Y$, whence j as the least upper bound must be a lower bound on Y. Hence it is the greatest lower bound on Y. But then it is the least upper bound on Y in $a^\perp$. ■

Applying this operation to the nine event spaces of cardinality at most 4, we see that all but the last two are self-dual, that is, left unchanged by this operation. The last two are duals of each other: inverting the Λ-shaped part of 8 turns it into the V-shaped part of 9, and vice versa, during which the top element remains fixed.

The next operation to define is $a\#b$. Given $a = (X_a, \bigvee_a, \infty_a)$ and $b = (X_b, \bigvee_b, \infty_b)$, we define $a\#b$ to be the poset whose underlying set is the cartesian product of the underlying sets of a and b, and which is ordered so that $(x, y) \leq (x', y')$ just when $x \leq x'$ and $y \leq y'$.

Theorem 2 *The poset $a\#b$ constructed as above is an event space.*

Proof: It can be seen that the element (∞_a, ∞_b) is top. Now consider any nonempty set Y of pairs in $a\#b$; take j to be $(\bigvee_a Y_a, \bigvee_b Y_b)$ where Y_a is the set of elements of a appearing as the first element of some pair in Y (the first projection of Y), nonempty because Y is, and similarly for Y_b. It is evident that j is an upper bound on Y. Any upper bound on Y must have its first component bound Y_a and its second Y_b. Hence j must be the least upper bound on Y. ■

We come now to the more delicate operation $a \multimap b$. For this we need the notion of an event map.

Given event spaces $a = (X_a, \bigvee_a, \infty_a)$ and $b = (X_b, \bigvee_b, \infty_b)$, an *event map* $f : a \to b$ is a homomorphism of event spaces. That is, it is a function $f : X_a \to X_b$ satisfying $f(\bigvee_a Y) = \bigvee_b f(Y)$ for all nonempty $Y \subseteq X_a$, and $f(\infty_a) = \infty_b$.

We take $a \multimap b$ to be the set of all event maps from a to b. We make it a poset by ordering it pointwise[3].

Theorem 3 *The poset $a \multimap b$ constructed as above is an event space.*

Proof: The top event map is of course just the constantly top map, satisfying $f(x) = \infty_b$ for all events $x \in a$.

Given any nonempty set F of event maps $f : a \to b$, take its join $\bigvee F$ to be the function $g : a \to b$ defined as $g(x) = \bigvee_{f \in F} f(x)$ for each event x in a. It remains to show that g is an event map.

Now $g(\infty_a) = \bigvee_{f \in F} f(\infty_a) = \bigvee \infty_b = \infty_b$, whence g preserves ∞. That g preserves nonempty joins follows thus.

$$\begin{aligned} g(\bigvee_{y \in Y} y) &= \bigvee_{f \in F} f(\bigvee_{y \in Y} y) \\ &= \bigvee_{f \in F} \bigvee_{y \in Y} f(y) \\ &= \bigvee_{y \in Y} \bigvee_{f \in F} f(y) \\ &= \bigvee_{y \in Y} g(y). \end{aligned}$$

∎

We conclude the specification of the primitive operations with that of $?a$. We take this to be the set of order ideals[4] of a, ordered by inclusion.

Theorem 4 *The poset $?a$ constructed as above is an event space.*

Proof: The top of $?a$ is X itself. Given any nonempty set of order ideals, their union is also an order ideal. ∎

Referring to the nine event spaces diagrammed above by their numbers, we have $?1 = 2$, $?2 = 4$, and $?3$ is 9 plus a fifth element above it.

We now derive in detail the remaining five operations via the equations discussed above.

<To be done.>

---

[3] Pointwise ordering of functions with the same domain and codomain means that $f \le g$ just when $f(x) \le g(x)$ for all x in their common domain.

[4] An order ideal or lower set of a poset a is a subset $Y \subset X$ such that if $y \in Y$ and $x \le y$ then $x \in Y$.

4 State Spaces

Thus far we have worked entirely within the universe of event spaces. We have taken the dual $a^\perp$ of an event space a to be the result of inverting all but the top of a. In fact it will become apparent that it is more natural to turn $a^\perp$ upside down. In that case $a^\perp$ has a simpler description: treating a as a poset, delete the top of a and adjoin a new bottom.

But now we need not have an event space. We call such posets state spaces. A state space is simply an inverted event space, having a bottom and nonempty finite and infinite meets.

Whereas the top event is the inaccessible event (heat death, apocalypse) of the schedule, the bottom state is the initial state of the automaton. And whereas the joins of an event space denote concurrent events, the meets of a state space denote decision states.

We may now extend our calculus to the larger universe containing both event spaces and state spaces. This requires the binary operations to be extended to all four combinations. Rather than treating the four combinations explicitly we may regard this universe as consisting of signed event spaces, event spaces equipped with an additional bit of information which is 0 or 1 (positive or negative) according to whether the event space is a real event space or one masquerading as a state space.

To combine signed event spaces under any of our operations we apply that operation independently to the two components, the event space and the sign bit. The operation behaves on the event space component as before. On the sign bit the operation is assigned its classical or Boolean interpretation. The operations $a\|b$ and $a \oplus b$ become disjunction $a \vee b$. The operations $a\#b$ and $a \otimes b$ are conjunction $a \wedge b$. Linear negation $a^\perp$ is just Boolean negation while $a \multimap b$ and $a \Rightarrow b$ are implication. The constants 0 and $\perp$ are 0 while 1 and $\top$ are 1. The operations $!a$ and $?a$ are the identity operation.

The only new distinctions so created between operations and constants are between constants: we no longer have the identities $0 = 1$ and $\perp = \top$.

The question now arises as to the meaning of certain operations applied to a mix of types, e.g. $a\|b$ when one is a schedule and the other an automaton. The formal meaning is not in doubt: the result in this example is an automaton because $0 \vee 1 = 1$. However the physical meaning is less clear.

In the case $a \multimap b$ with a a schedule and b an automaton, the result is $0 \rightarrow 1 = 1$, an automaton. There is a natural interpretation of this. We may view the schedule a as a program being executed on the machine b; this combination is then the programmed machine $a \multimap b$. Each state of this machine assigns to each event of a a state of b indicating the progress made thus far in executing a on b.

<More to come.>

5 Acknowledgments

Rob van Glabbeek turned my attention from the overly abstract n-dimensional complexes of [Pra91] to more concrete and plausible cubical sets, and a phone conversation with Bill Lawvere then further turned it from cubical sets to their dual objects bipointed sets. The connection between that duality and Stone duality then dawned on me, and the additional intuition conveyed by the self-duality of **CSLat** [Lat76, Joh78] led quickly to **Aut**, though the concrete interpretation of the phenomena in **Aut** in terms of unreachable events and initial states took time. That $!a$ even made sense in **Aut** was completely unobvious to me until I had digested the relevant parts of the papers of Barr [Bar91] and Seely [See89], after which I wondered why I did not see it right away. Thanks also to Michael

Barr for email correspondence demystifying monads (triples) to the point where I could understand that they were closer to the foundations of algebra than needed either for the audience or the results. And I have found very helpful the ideas on event spaces flowing from my Stanford class CS 204, Undergraduate Programming and Problem Solving Seminar, as well as its auditors Vineet Gupta, Pat Lincoln, and Bharat Shyam.

References

[Bar91] M. Barr. $*$-Autonomous categories and linear logic. *Math Structures in Comp. Sci.*, 1(2), 1991.

[Gra81] J. Grabowski. On partial languages. *Fundamenta Informaticae*, IV.2:427–498, 1981.

[Joh78] P.T. Johnstone. A note on complete semilattices. *Algebra Universalis*, 8:260–261, 1978.

[Lat76] D. Latch. Arbitrary products are coproducts in complete ($\vee$-)semilattices. *Algebra Universalis*, 6:97–98, 1976.

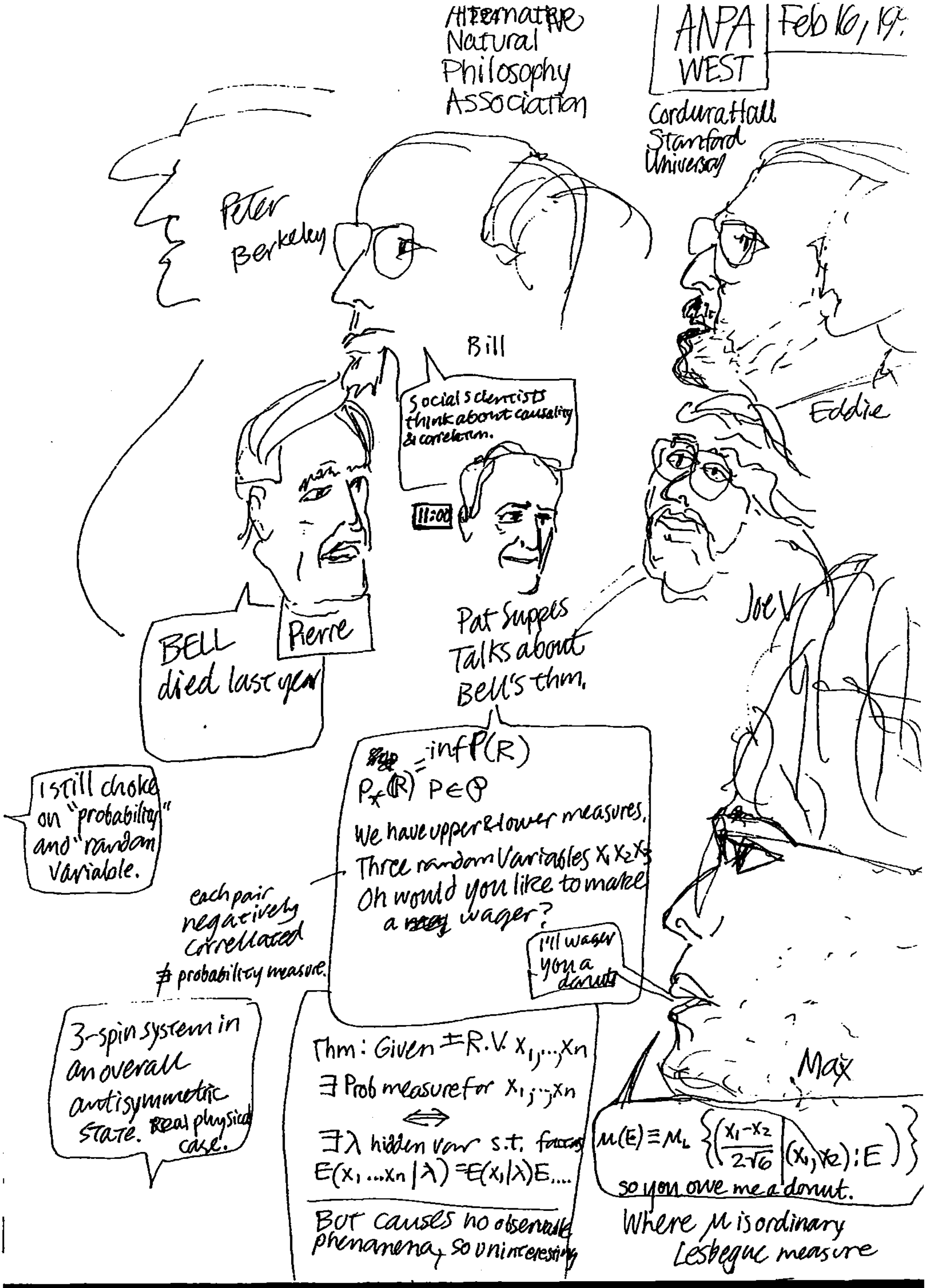
[NPW81] M. Nielsen, G. Plotkin, and G. Winskel. Petri nets, event structures, and domains, part I. *Theoretical Computer Science*, 13, 1981.

[Pra82] V.R. Pratt. On the composition of processes. In *Proceedings of the Ninth Annual ACM Symposium on Principles of Programming Languages*, January 1982.

[Pra91] V.R. Pratt. Modeling concurrency with geometry. In *Proc. 18th Ann. ACM Symposium on Principles of Programming Languages*, January 1991.

[See89] R.A.G Seely. Linear logic, $*$-autonomous categories and cofree algebras. In *Categories in Computer Science and Logic*, volume 92 of *Contemporary Mathematics*, pages 371–382, held June 1987, Boulder, Colorado, 1989.

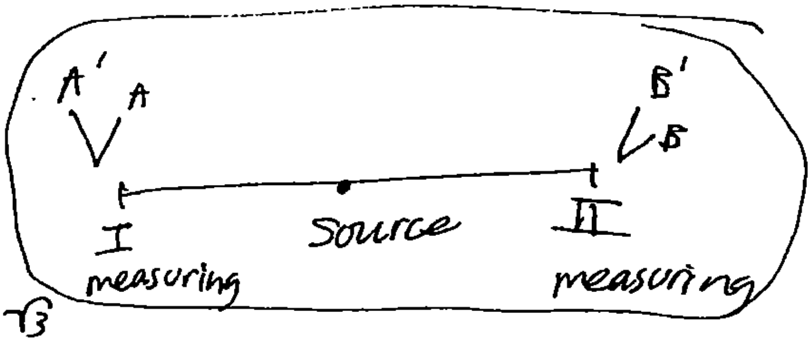
[vG90] R. van Glabbeek. *Comparative Concurrency Semantics and Refinement of Actions.* PhD thesis, Vrije Universiteit te Amsterdam, May 1990.



Reprinted from Proc ANPA West 7 (1991)

②

BELL CASE



$AB - AB' + A'B + A'B' < -2$

$AB = A'B' = -\cos 30° = -\frac{\sqrt{3}}{2}$

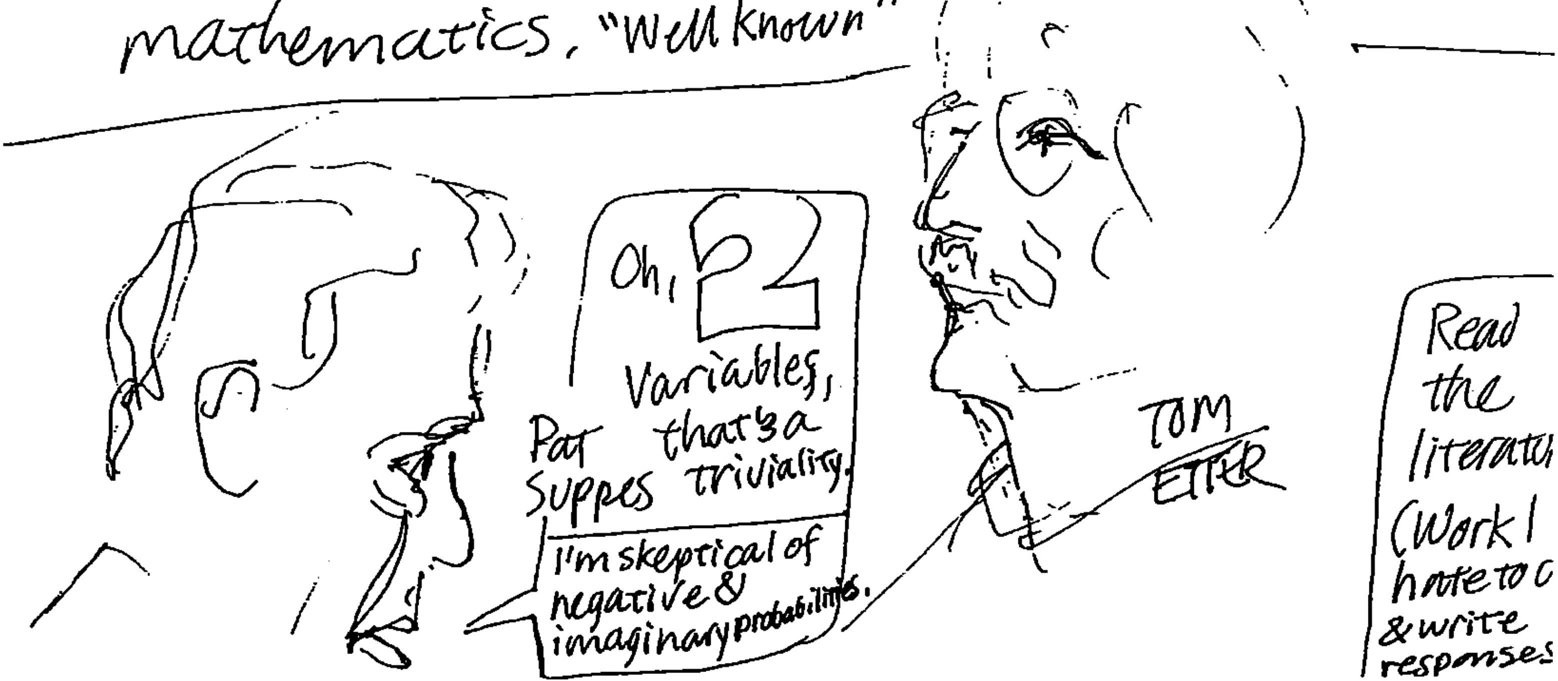
$AB' = -\cos 60° = -\frac{1}{2}$

$A'B = -\cos 0° = -1$

So $-\frac{\sqrt{3}}{2} + \frac{1}{2} - 1 - \frac{\sqrt{3}}{2} < -2$ So Bell inequality NOT satisfied.

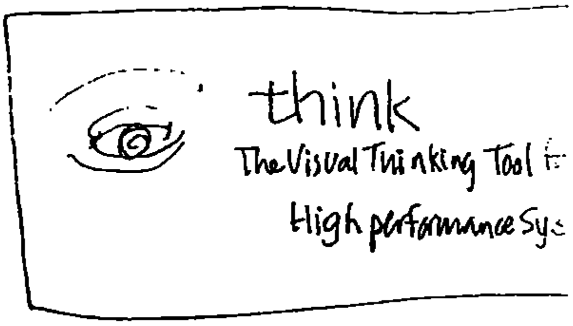
What is the correlation between A & A'?
Not predicted by quantum mechanics.

Reminds me of my frustrating days in probability class, barely grasping anything, being outraged by opaque language of speaker.
DEFINE new language! Not just technical terms, but funny phrases, like "for example $x_1, \ldots, x_n$" which introduces notation, NOT gives an example.
& Use it in a setting other than mathematics. "Well known"



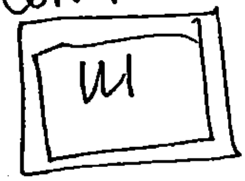
③

Vaughan
Pratt



think
The Visual Thinking Tool f
High performance Sy

COMPUTERS



Builtin
layers

pix
↑
word
↑
number

leads to
mis
matches.

Screen
Realities

Computer

What difference
Why?
Why not?

→ usefulness.

:00 PIERRE NOYES

"Anti-protons will fall up!" A prediction I'm staking my professional reputation on.

I'm sometimes accused of not having a sense of humor. I hope to preserve that.

I'm talking about Newtonian anti-gravity. Anti protons are the only ones we can measure now.

Start by questioning the equivalence principle... leads to antigravity. This is probably the most likely ~~then~~ formulation of discrete physics! Experimentally verifiable in a couple years.

Gravitational ≡ inertial mass.

Special case: $E=mc^2$

QM — Special Relativity

Newton   Special Relativity — General Relativity

Quantum Gravity

→ People haven't been able to reconcile with QM.
Perhaps have to deny equivalence principle.

CPT

Weak

Does gravity violate CPT?
In the conventional sense, yes!

BOHR $\frac{1}{137} = 3+7+127$

SOMMERFIELD $(1/137)^2$

ELECTROMAGNETIC

LIKE CHARGES REPEL ↓↓

UNLIKE CHARGES ATTRACT

QUANTIFIED NEWTON $1/1.7\times10^{38}$

CORRECTION NEGLIGIBLE (FOR PARTICLES)

Ultimately INERTIAL MASS and GRAVITATIONAL CHARGE are distinct.

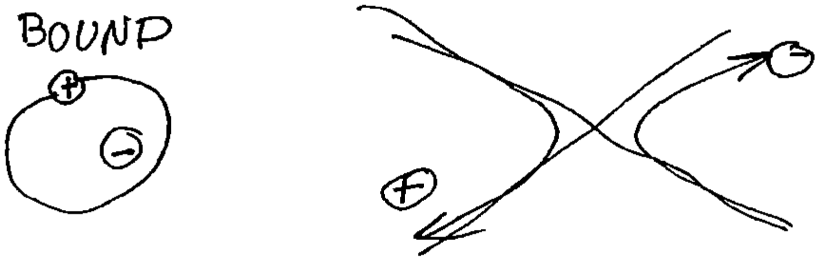
GRAVITATION

LIKE MASSES ATTRACT

UNLIKE MASSES REPEL

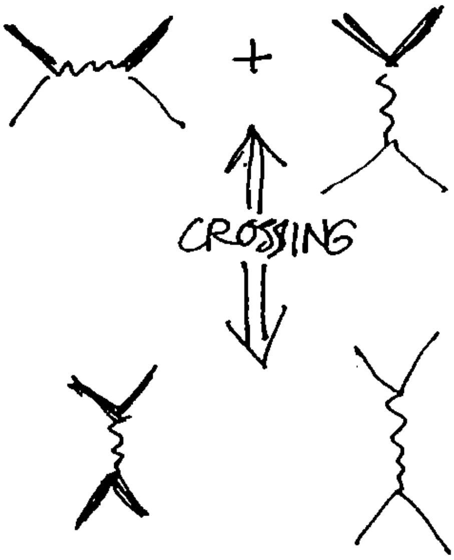
⑤

ATTRACTION AND REPULSION



CLASSICAL

BOUND STATES
NON-PERTURBATIVE



Particle and articles annihilate long before they could form a bound pair.

~~FEYNMAN~~

QUANTUM

CHANGE IN VEL ⇒ DEFINES FORCE

FORCE ON CHARGE ⇒ DEFINES FIELD

⇓ FEYNMAN'S PROOF

Maxwell's Equations.

$F = ma$
$E = mc^2$

STRING of 0's & 1's.

John S Chappell.

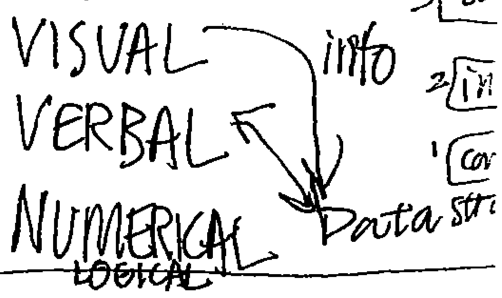
Discretize

VAUGHAN & Pixels.

⑥

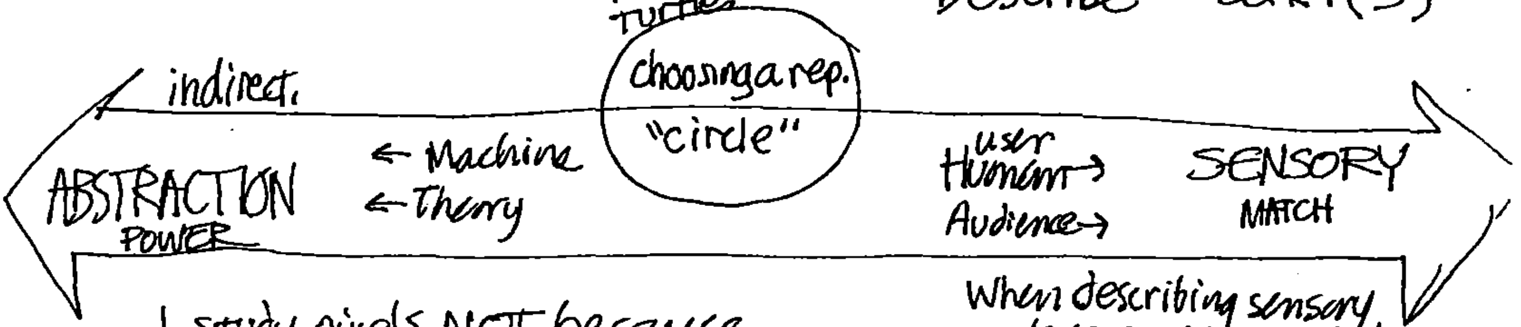
Discretizes loses info.
How do you subdivide forever?
How do you zoom?

Want the compute
TO use



Discretize 10010111
Describe SQRT(3)

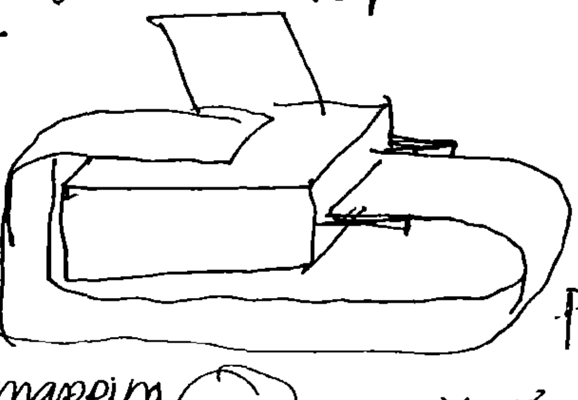
def
word
pixels
equation
iterative form.
TURTLE



I study pixels NOT because
I advocate them (though I do)
but as a counterexample
To elucidate a deeper issue.
Like visiting another country to become aware of door opening directi

When describing sensory
phenomena. Graphics prog
Spatial, simultaneous.
Spreadsheet.
LCD.

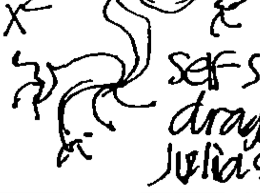
3:30 FRED YOUNG
Fractals



Peitgen:
A language.
Parameter space.

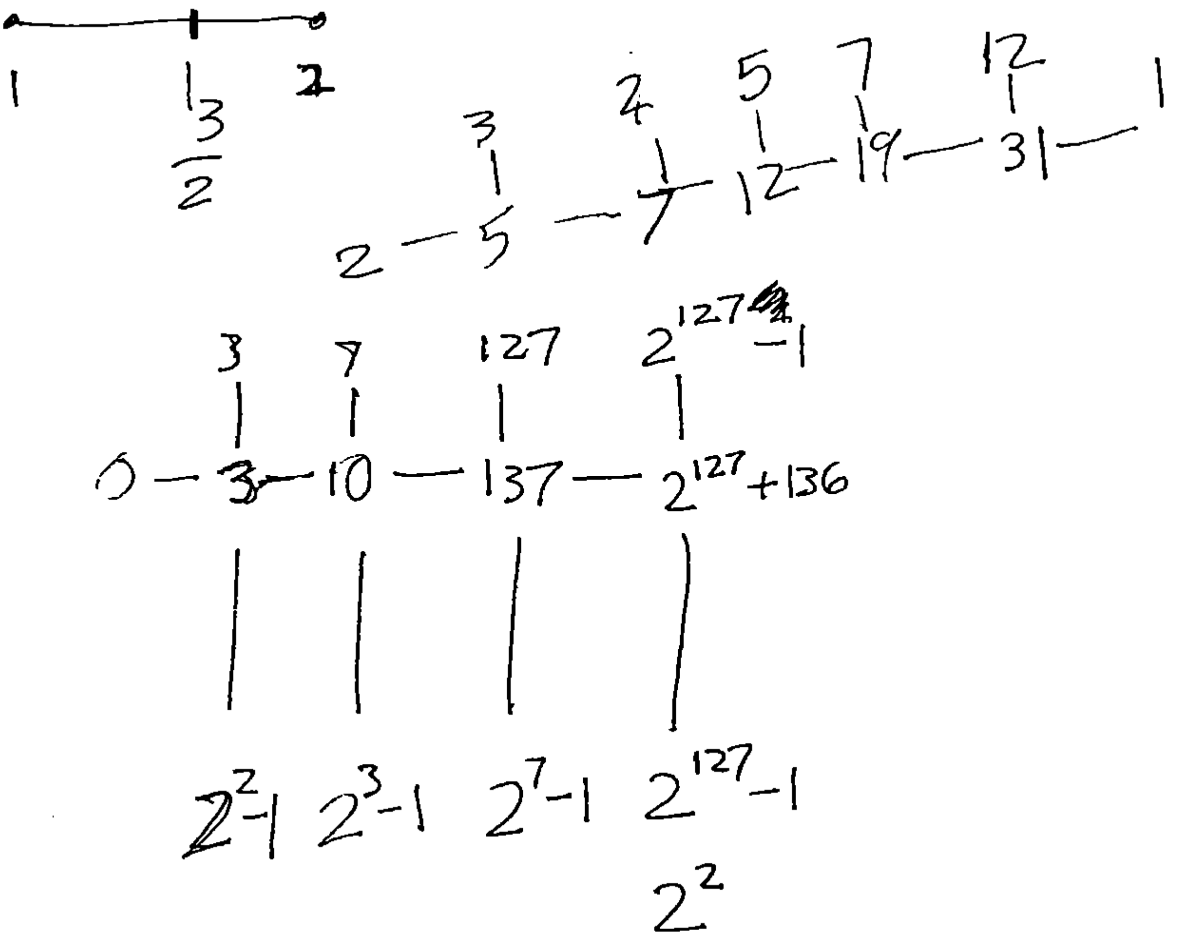
Self-mapping fractals.

$X \leftarrow X^2$



$X_{n+1} = X_n^2$

⑦



Barnsley — code any photo as a fractal set.

ARITHMETICS

Goal: model
semantics of
concurrency

$\boxed{5{:}00}$ $\boxed{\text{Vaughan Pratt}}$

(8)

NUMBERS $x+y$ xy x^y $\neg x = 0^x$

tensor products
of event spaces

PROPOSITIONS $p+q$ pq. $q \to p$ $\neg p = p \to 0$
(BOOLE 1847)

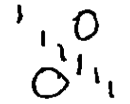
did he
mean
OR or XOR? Boole didn't know!
$p+p=0$.

electrified logic by
adding arith.

RELATIONS $R+S$ $R \cdot S$ $S \to R$ $\bar{R}$
(PEIRCE 1873) $S \dagger R$ $R;S$ $S \backslash R$ $\breve{R} = R \backslash 0'$

composition residuation

← taking the order
of the pair into
account.



iterated prod

POSETS $P+Q$ PQ P^Q
(Birkhoff 1937 42) $P \oplus Q$ $P \cdot Q$ Q_P

CARDINALS ignore order
ORDINALS preserve order

VECTOR SPACES $U+V = \cancel{U} \times V$ $U \oplus V = U \otimes V$ U^V !

direct ~~di~~product
tensor product

SETS
$V^* = \mathbb{R}^V$

LINEAR LOGIC $A \oplus B$ $A \& B$ $B \Rrightarrow A$ $A^\perp = A \multimap \perp$
(GIRARD 1987) $A ⅋ B$ $A \otimes B$ $B \multimap A$

He knows how to a
& multiply proc
sometimes trea
as a set of propo
with implication
sometimes ~~wit~~
respecting stru

EVENT SPACES $a+b$ ab $b \Rrightarrow a$ $a^\perp = a \multimap \perp$
(PRATT 1991) $a \oplus b$ $a \otimes b$ $b \multimap a$

CONSTANTS (units) 0 1 monotone maps "~"
$\perp$ $\top$ homomorphism "implies"

bottom top

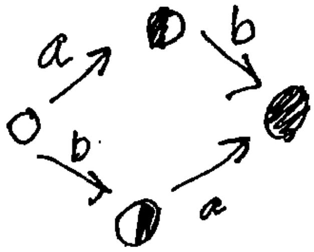
⑨

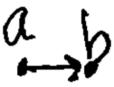
SCHEDULES

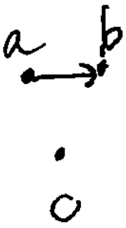
AUTOMATA

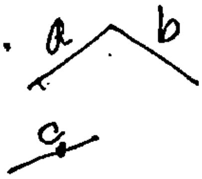
The powerset of a schedule

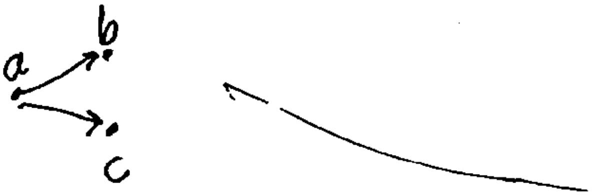
a b

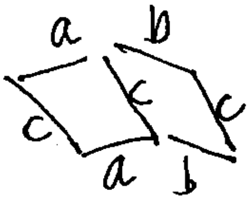


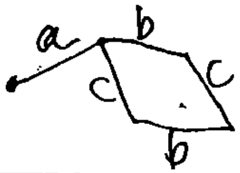










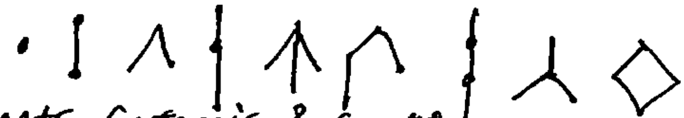


Can we do it so they're more similar?

Event spaces: POSET$(X, \vee, \infty)$ Top, like the horizon.

↳ $(2^X - \emptyset) \to X$ is nonempty join

E.g.



∞ concurrent events

Meaning: X Events (atomic & compound)

$\vee y$ concurrence

∞ never reached apocalypse, fixed pt of transforms

an upper semilattice

(semilattice w/o null s

V y is AND if $v y \neq \infty$

$a + b = (b^\perp a^\perp)^\perp$

$a \otimes b = (b \multimap a^\perp)^\perp$ collision of 2 event spac

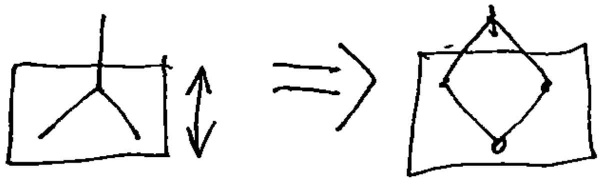
$a \oplus b = a^\perp \multimap b$

$!a = (?(a^\perp))^\perp$

$\forall$ $\exists$

$b \Rightarrow a = !b \multimap a$

$a^\perp$:



Very clear! Can we do a computer gam on Lattices? (Rocky's Boots!)

Theorem: $a^\perp$ is an E.S.

ab direct prod this too $f : b \to a$

$b \multimap a$ set of all maps this too

$?a$: { order ideals of a ordered by inclusion) this

COMBINING ONTOLOGICAL HIERARCHIES, AS ANTICIPATED BY PARKER-RHODES

Fritz Lehmann
4282 Sandburg, Irvine, CA 92715, U.S.A. Tel.: (714) 733-0566 Fax:(714) 733-0506
or 124 Parsons, Webster Groves, MO 63119, U.S.A. Tel.:(314) 968-8337
email: fritz@rodin.wustl.edu (or flehmann@orion.oac.uci.edu)

"... there may be several dozens of hierarchies." [Cercone et al., *The ECO Family*, 1992]
"Russell for instance imagines every fact as a spatial complex." [Wittgenstein, *Logical Notebooks*, 1913]

This article deals with a mathematical conceptual construction that is both philosophically significant and practically useful for performing very fast computer inference in Artificial Intelligence (AI). Since the Alternative Natural Philosophy Association was founded on some ideas of the late mushroom specialist Arthur Frederick Parker-Rhodes in the area of theoretical particle physics, I will tie this discussion to his other farsighted and highly independent work in the superficially unrelated area of formal mathematico-linguistic semantics.

Different conceptual hierarchies combine to form the main hierarchy of the world's concepts and relations. A description of something may include concepts from **part-whole hierarchies, space/time inclusion lattices, formalized measurement scales, shape classifications, color schemes, the lattice of quantifiers, goal/purpose hierarchies, ordered lexicons, case-relation systems, natural kind taxonomies, legal category systems, data models, class inheritance systems, granularity hierarchies, causal orders, plurality powerset systems, verisimilitude orderings, etc.** Any such different "ontological dimensions" can be combined mathematically to generate the full hierarchy of relational descriptions ordered by generality. In particular, a "determinable" (attribute with values) like *Color* or *Height* (as opposed to an object or a simple quality) is in reality a *poset factor* or separable factor in the hierarchy of all descriptions and therefore corresponds to a formal "ontological dimension". In physics it may be a fundamental variable or a "degree of freedom." In *Inferential Semantics*, A. F. Parker-Rhodes represented such ontological dimensions as lattice-structured **base-domains** of labels which label elements of the Rhema Graph (semantic network) of a description or proposition. [Parker-Rhodes 1978]

Concept-systems for particular subject areas, and ontologies for the world in general, are usually presented either as *hierarchic systems* (ordered sets) of concepts or as *collections of primitive concepts* from which more elaborate concepts may be built according to rules of composition. The latter approach is also inherently hierarchic since the supplied rules for building complex concepts from more primitive ones induce a virtual hierarchy of concepts, whether it is explicitly provided or not. Hierarchy here means not only a tree structure, but any ordered set or poset in which one member may be above another.

It is important in Artificial Intelligence, for example, that a complex concept system is factorable into smaller, contributing systems that each permit independent, fast(er) analysis. Every set of descriptions forms a partially ordered set (poset) of equivalence-classes ordered by inferability, which ordering is in part a function both of the **relational structures of the descriptions** and of various **orderings on their components.** In Order Theory an ordered structure (like a tree, lattice or general poset) is *factored* into smaller *quotient structures*. It is only because the world *is* ontologically structured that there can be a useful decomposition of the poset of descriptions into separate contributing factors. If everything were a disordered (but perfectly logical) flotsam & jetsam of predicates and relations then little would be gained. First order logic makes no contribution to this question; we are concerned here with particular higher order theories of the world's structure of concepts and relations.

I suggest this *large-scale structural level* analysis as an important aspect of *formal ontology*. The task is to mathematically factor the knowledge hierarchy (the partially ordered set of descriptions) into some kind of "product" of its contributing ontological subhierarchies. Divide and Conquer.

Reprinted from Proc. ANPA West 10 (1994)

In discussing combination of different ontological dimensions (subhierarchy factors) by way of two new "product"-like operators, I will point out a peculiar factor poset which is necessarily a "hidden participant" in the hierarchy of descriptions. Ideally, a complete and efficient encoding of all descriptions should be possible, based on a modest number of fundamental semantic primitives and "ontological dimensions". To accomplish this it is necessary to decide several philosophical questions, and build some mathematical structures based on those decisions. This leads from general ontology to graph and order theory.

A Simple Example: Faceted Classification

For classifying books, S. Ranganathan proposed "faceted classification" as an alternative to the tree of categories used in the Dewey Decimal system or the Library of Congress system. His COLON Classification [Ranganathan 1965] uses a generic subject-area and five *facets*: **Personality, Matter, Energy, Space and Time.** I drew the diagram below to suggest the large-scale factor structure of the COLON category hierarchy. The resulting hierarchy is neither a tree nor an unrestricted poset, but a *direct product* (Cartesian Product, cross product) *of trees.*

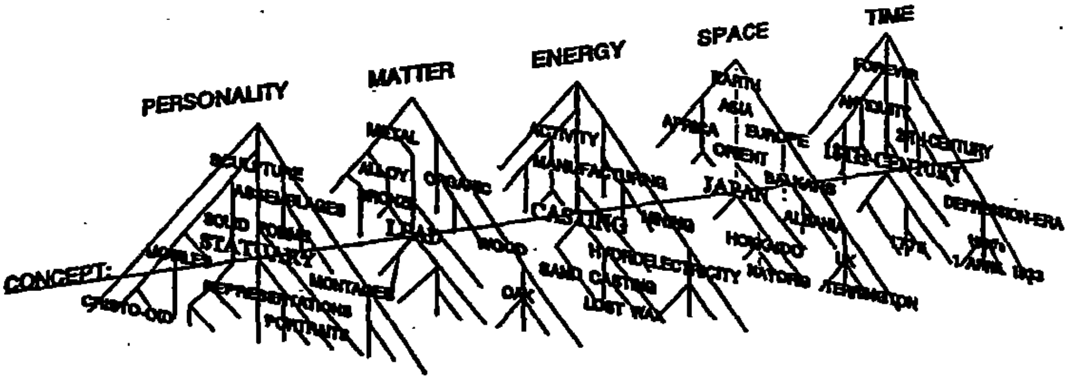


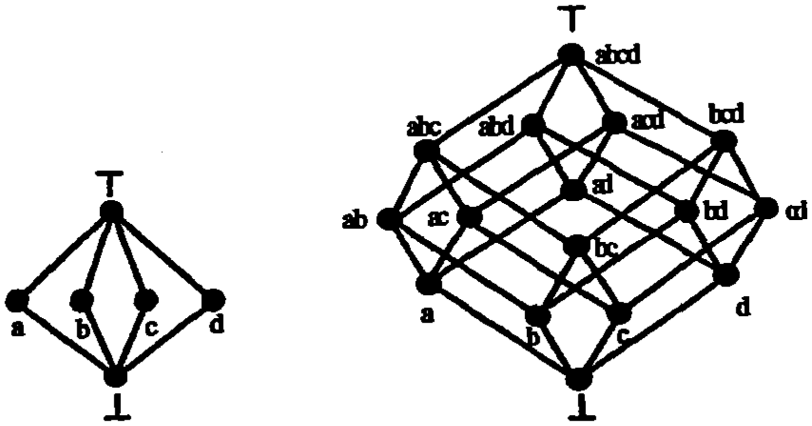
Figure 1: Separate tree structures for each facet impinge on a book about "18th Century Japanese cast lead statues"

One approach in AI achieves some restricted multiple inheritance of properties (from multiple superclasses) by treating an object as occurring in different independent (tree structured)*perspectives;* the resulting abstract poset of descriptions is again a direct-product of trees. [Mariño et al. 1990]. This is a step in the right direction, and has computational advantages, but two shortcomings are: A. a particular ontological dimension may have any (or no) order structure, not just a tree structure; and B. few descriptions are representable as simple direct-products or lists of features -- most have relational graph structures. For reasons of his own, Parker-Rhodes confined both his base-domains and his Rhema Graphs to being mathematical lattices. This too is too restrictive.

The combination of different **arbitrary posets** for each ontological aspect and interrelation of these aspects by a **relational graph** generates the full knowledge hierarchy. If you add arbitrarily nestable negations, this system has the descriptive power of typed (order-sorted) predicate calculus. But descriptive power of logical elements is not the subject here — the subject is the large-scale "architectural structure" of useful ontological theories made of those elements.

Chopping Away at the Boolean Lattice

One meaning of "algebraic" emphasizes the lattice-like structures induced by combinations of basic elements of a system. Let's start with the independent predicates of monadic predicate calculus. This is the Boolean world of Leibniz, a world without nontrivial structure. Without ontology. As we add structure to the concept system, the hierarchy of possible descriptions will be drastically reduced from an initial, enormous Boolean lattice on atomic primitive qualities. Diagrams will show the effects of various structural changes on the lattice. The goal is still Leibniz's dream of an *Ars Combinatoria* to generate everything from a small set of semantic primitives (and to do it in a way that enables great computational efficiency in automated reasoning).



Flat set of four independent qualities
{a,b,c,d}

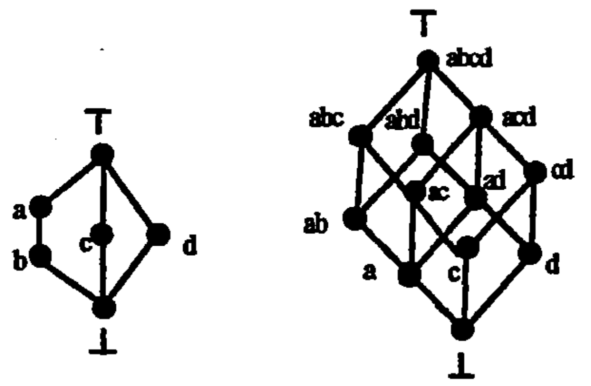
The Boolean lattice they generate
Powerset({a,b,c,d}) = 2^4

Figure 2

The approach here is to add more and more higher-order structure to the poset of descriptions, eliminating impossible, redundant and irrelevant models at each step, which mathematically amounts to chopping out large parts of the lattice at each step. The resulting hierarchy structure is induced by the dependencies in *a priori analytic* (e.g. terminological) knowledge which follows from logic, mathematics and formal definitions, as well as *a priori synthetic* knowledge, or hierarchic knowledge which follows from ontology or the conceptual systems, laws, and sets of long-persisting facts in our particular world.

Dependencies, Mutual Exclusion, and Partitions

A Boolean lattice is generated by its *atoms* (the elements just above or "covering" the bottom element) which are mutually *independent*, hence incomparable. These are the primitive predicates described above, and every point in the lattice represents some unrestricted combination of primitive predicates. In reality, most combinations cannot occur: there are no *wolves* that are *affidavits*, no *bachelors* who are *husbands*, and no *Nebraskans* who are not *Americans*. Dependency among the predicates is *a priori* second-order knowledge; you convert the initially unordered set of predicates into a partially ordered set of predicates, or taxonomy. Just as an *unordered* set of primitives generates a *Boolean* lattice of possible descriptions, any *partially ordered* set of primitives generates a unique *distributive* lattice according to Birkhoff's Representation Theorem for distributive lattices. [Birkhoff 1967, Davey & Priestley 1990] This is shown in Figure 3 (A Boolean lattice is a special case of a distributive lattice, and "atoms" are a special case of the distributive-lattice-generating *join-irreducible* elements. Join-irreducible elements in a lattice diagram are the dots with only one descending line.)

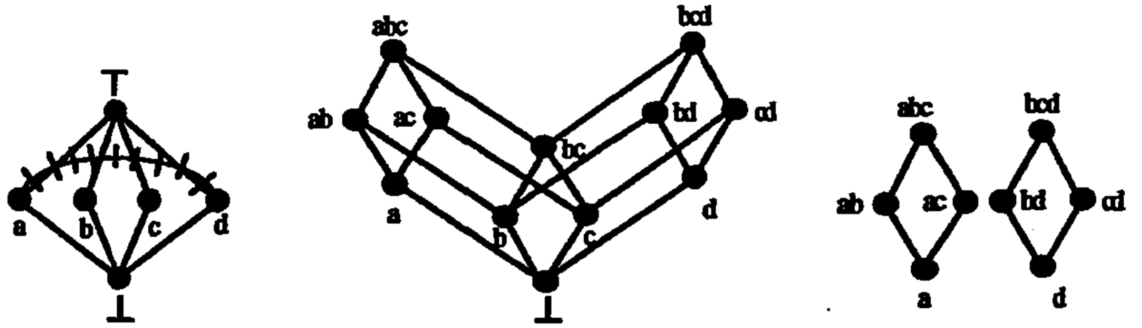


In this poset, a > b

The distributive lattice generated by the poset

Figure 3

Mutual exclusion is among the most common and useful restrictions in a concept hierarchy. If a figurine is 19 cm tall, it cannot also be 25 cm tall. Height, like many "scales", does not admit multiple values. In natural taxonomies, it is not possible for some living thing to be both an orchid and a baboon; these taxonomies are almost always presented as tree-structured hierarchies in which the child nodes of any parent node are mutually exclusive. For other hierarchies, like interval lattices, this principle does not apply. The effect of mutual exclusion on the lattice of descriptions is to rip off the top as shown in Figure 4:



Mutual exclusion: {a,b,c,d} but Not(a & d)

The resulting poset

Exhaustive too: a XOR d yields this poset

Figure 4

If a set of mutually exclusive predicates is also exhaustive, forming a mathematical *partition*, the reduction is even more drastic.

Determinables and Scales as Poset Factors or Ontological Dimensions

A "determinable" in philosophy is an attribute that can have certain values (called "determinates"). *Color* is a determinable for which *red* is one of its determinates. [Searle 1959, Hautamäki 1986, Way 1991] Carnap called it "a family of related properties." [Carnap 1950] (The only reason I use the over-technical "determinable" is that all the more familiar words for it like "attribute" are ambiguous.) The ubiquitous object-attribute-value triad used in databases and frame systems in AI has resisted philosophical consensus. What's an attribute? First Order Logic itself casts no light on this, since *Red*(*x*), *Color*(*x*,*red*), *Colored*(*x*), *Color*(*red*) are not recognized as related. A determinable is a certain *aspect* of something; its determinates (values) are usually interrelated by *dependence* (often mutual exclusion) whereas different determinables may be quite *independent* of one another. In addition, only certain values (determinates) are appropriate for a determinable, and only certain determinables are appropriate for a given type of object. (It doesn't make sense assign the value "pregnant" to a spin-angular-momentum variable, nor to specify any spin-angular-momentum at all for "sportsmanship".)

Determinables, attributes, "ontological dimensions", facets, perspectives, scales and *"BOXes" in hybrid knowledge representations* are all *factor posets* in the full hierarchy of descriptions; this is defined in the higher-order logic of structural relations among predicates (and among relations). The rather mathematical treatment of determinables as poset quotient factors should not distract you from the fact that these structures, like all assignments of objects to classes, are based on pragmatic interests in the real world. Size and color are separate ontological *dimensions* partly because repainting a large box will not help you get it through a small door.

Ranganathan's COLON Classification facets described above are determinables. So are the *categorical accidents* of Aristotle, like time, place, purpose, and quantity. Significantly for AI and linguistics, these correspond to the "case relations" like ACTOR, INSTRUMENT, PATIENT and LOCATION used in case grammars and as primitive relations in semantic networks like Wilks' Preference Semantics [Wilks & Fass 1992], Schank's Conceptual Dependency theory [Schank 1975, Lytinen 1992], Parker-Rhodes' Inferential Semantics [Parker-Rhodes 1978], or Sowa's Conceptual Graphs [Sowa 1984,1991,1992]. In fact, these case relations both *refer* to separate hierarchies and they, as relations, *occur* in a relational hierarchy like that of Parker-Rhodes shown above. The specialized reasoning components of Schubert's ECO group, like time, space, the "topic hierarchy", the color-solid scheme, etc., are all poset-structured determinables although they have not been formalized as such. [Cercone et al. 1992].

In Measurement Theory various *scale structures* have been found. [Roberts 1979] Ganter & Wille list fourteen different types of lattice-structured scale such as ordinal, biordinal, dichotomic, interordinal, *et al.* that have been developed for scientific measurement, each of which has its own kind of lattice structure. They use lattice-theoretic methods (completions of the direct-product of different scale-value lattices) to build a "formal concept lattice" [see Wille 1982, 1992] of possible descriptions using those scales as attributes. [Ganter & Wille 1989] A particular description has a value from each of a set of scale lattices; the lattice of all descriptions combines these scale lattices. Subsumption is automatic in the lattice. Similar work has been done independently in Russia in the field of "classification and meronomy" based on Yu. Shreider's work. [Rubashkin 1976, Raskina et al. 1976, Polyakov & Dunaev 1985] These theories have combined structures using some innovative "products" in addition to direct product of lattices. For every mathematical "product" that combines structures, there is a corresponding (but not always unique) factorization or decomposition of a composite hierarchy into its components.[1]

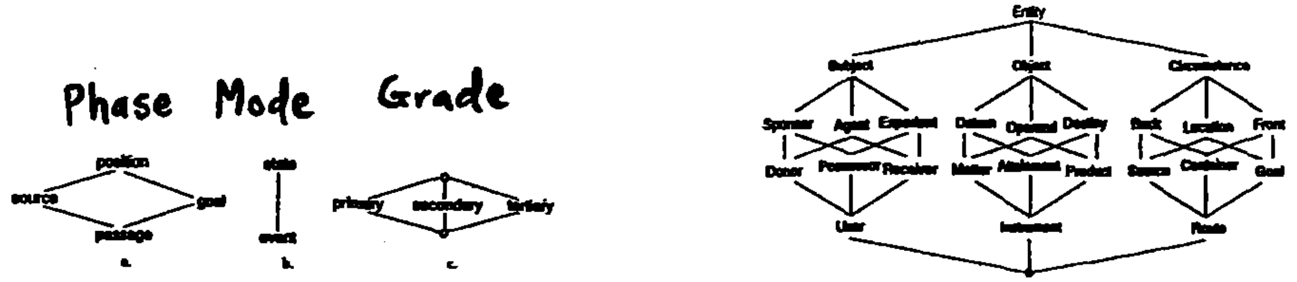


Figure 5. Three factors and their cross-product: Parker-Rhodes' Case Lattice

Parker-Rhodes' theory of inferential semantics [Parker-Rhodes 1978] builds the lattice of descriptions from constituent ontological sublattices. He includes a quantification lattice, a pronomial

---

[1] Wille's concept lattices developed so far do not deal with relational graph structure except in new work of Monika Zickwolff which makes this extension. [Zickwolff 1991] To accomplish the needed structured combination of ontological dimensions, her work (along with that described in this paper) should be combined with Ganter & Wille's structured scale/attribute dimensions.

identification lattice and a lattice of case relations; some of these are themselves direct-products of sublattice "factors". This comes closer to the structure I'm using. Figure 5 exhibits Parker-Rhodes' *Case Lattice* is as a direct product of three aspects of case relations, *Phase* (direction), *Mode* (passive/active) and *Grade* (causal status). A particular "case-relation" like INSTRUMENTAL has a value in all three. As second-order predicates, *Phase*, *Mode* and *Grade*, unlike INSTRUMENTAL, have no place in the resulting hierarchy of case-relations itself as members. Rather, they are its direct factor posets.

So the determinable-determinate relation is *not* that of supertype to subtype. It is wrong to say (as in [Way 1991]) that *Color* is to *Red* as *Red* is to *Crimson*. Although *Crimson* is a genuine subtype of *Red* in the hierarchy of colors ordered by generality, and both are first order predicates, *Color* is a second order predicate describing those first order predicates which are members of the ordered set of colors. Every determinable name is thus higher-order, determining a large-scale structural feature of the type hierarchy.

SEMANTIC NETWORKS

Since Bertrand Russell got the calculus of relations from Charles S. Peirce [Russell 1903, p.23n], symbolic logic has transcended the syllogism and dealt with n-adic relations along with monadic predicates. This alters the structure of the description hierarchy because John's being SON-OF-MARY and Mary's being MOTHER-OF-JOHN are no longer independent predications (Leibniz's relational predicates), but are identified as one instance of a *relation*. Any model of the world which has one such "predication" without the other is thereby excluded.

The next needed addition is that of relational cycles or circuits, which can only be specified using *identity of individuals* or *path equations*. This involves the most philosophically dangerous concept, that of "self". In the new, possibly cyclic, structure, we can describe a man scratching his nose. *His own* nose, not "some nose which is scratched by some man who scratches some nose which is scratched by some man who scratches ... etc." Only from *outside* the system you (your *"self"*) can see these two cases as distinct; the purely intensional description (without identity of individuals) applies also to an infinite chain of men each scratching the nose of the next man, or to a finite mutual nose-scratching cycle of any n men. Conveniently, an infinite number of models is eliminated by the identity, as is the need for infinitely long formal definitions of your-own-nose-scratching.[2]

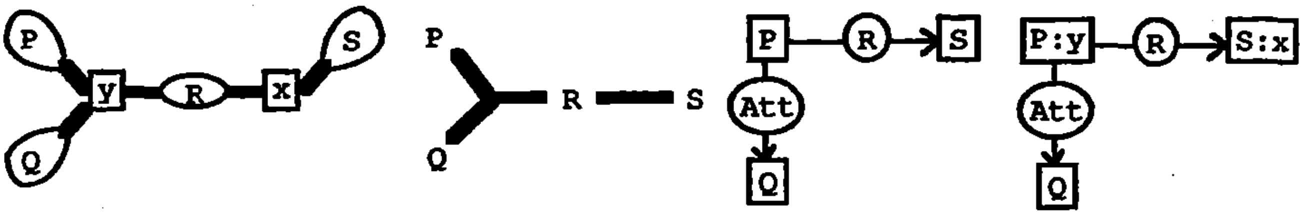
The graphic notation for type-labeled relational structures with type hierarchies, called semantic networks [Lehmann 1992, Sowa 1991], makes clear the graph-theoretic and order-theoretic nature of the development, and avoids the distracting notational artifacts (like bound variables, converses, arbitrary order embeddings, equations, inequations and quantifier precedence) of linear Predicate Calculus notation (though both are useful representations of the same underlying logical combinatorial structure).

In a semantic network, type-labeled nodes represent concepts and type-labeled directed edges represent relations between two concepts (or directed hyperedges for relations among more than two). The desired inference structure is a subsumption hierarchy of descriptive nets in which higher, more general descriptions subsume (and are implied by) lower, more specific descriptions. Each concept-node in a graph is labeled with the type of concept it represents; similarly each relation-edge in a graph is labeled with the type of relation it represents. The type label on each concept-node (and that on each relation-hyperedge) has its position in some external type subsumption hierarchy or taxonomy. Parker-Rhodes' Rhema Structures, Sowa's Conceptual Graphs, Order-Sorted logics, KL-ONE and Aït-Kaci' ε -types are

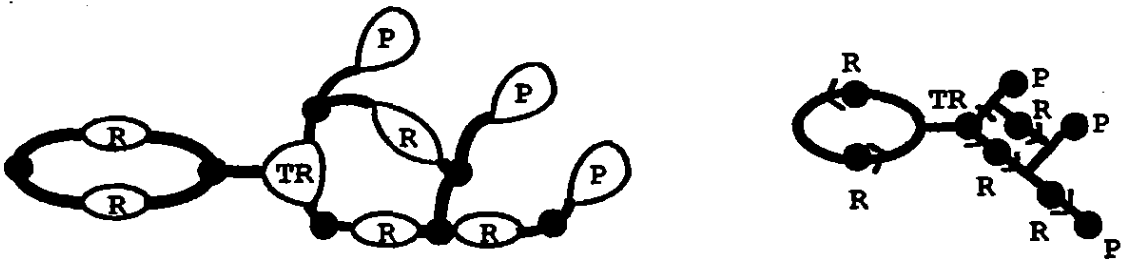
---

[2] That the notion of "self" is ultimately bogus within any system, I hope to prove someday. It is involved in incompleteness theorems, diagonal arguments, and the troublesome logical paradoxes. An apparent consequence is that one's own self cannot exist, hence cannot cease to exist (die).

all representable as this kind of network. The actual shapes, orientations and styles of the graphs are irrelevant to meaning. Only the topological (or rather the *combinatorial graph-theoretic*) structure matters.



$S(x) \wedge P(y) \wedge Q(y) \wedge R(x,y)$ Existential Graph Conceptual Graph Same, with referent field



Relational Structure has triadic relation and cycles It's a labelled directed hypergraph

Figure 6

On the top line above, the same assertion is shown in linear First Order Predicate Calculus notation, Charles Peirce's Existential Graphs, and John Sowa's Conceptual Graphs based on them (the CG uses circles in place of hyperedges and distinguishes P, a *concept type* like "cat", from Q, a *dependent attribute concept* like "red"). [Sowa 1984,1992] The bottom line shows a relational structure which would be tedious to express in linear First Order Predicate Calculus notation (P=predicate, R=dyadic relation and TR=triadic relation). The lower two nets are *duals* — on the left, dots are things and the lines (in this case labeled hyperedges) are relations, as in most semantic networks; on the right, however, the dots are relations and the lines are things, as in Existential Graphs' lines of individual identity.

The "Graph Dimension" in Ontology

Separate poset structures of *time, spatial inclusion, lexical categories, natural kind taxonomies, part-whole hierarchies,* etc. are the participating, contributing "ontologocal dimensions" or sources of the overall structure of the conceptual hierarchy, but in any system based on logical or linguistic descriptions **there is always an extra participant**, an "extra guest at the party" that is also a factor (formally) in determining the final structure. That is the "graph dimension," which I deem a formal-ontological hierarchy in its own right because it acts "algebraically" somewhat like the other hierarchy factors in determining the final hierarchy of all descriptions. It too is a kind of quotient poset factor.

Every linguistic description or logical "complex" has a semantic structure (its semantic net). For many categories of objects, ideas or situations, it is not adequate to describe them as ordered lists of values in different ontological dimensions (determinables) like height, weight and color. Most real-life category definitions or situation descriptions depend on some relational structure. For example, no simple combination of word-counts or similar "metrics" can distinguish a sound legal argument from a fallacious one. The same is true of computer programs, aircraft wiring schemes, or automatic plans in AI. These cannot be usefully categorized by the presence or absence of simple qualities, nor by tuples of values in a list of scales. This may be a fundamental drawback of common methods in physics and similar sciences, in which an object is characterized entirely by a vector of variable-values.

Given a set of concepts and relations, mapping set members to nodes of a graph structure converts the mere set of constituents into a "complex". Peter Simons, discussing the logical "complexes" of Wittgenstein's *Tractatus*, says "The structure of an atomic complex consists in the way its simples are concatenated together: this structure is not a further constituent of the complex," [Simons 1992, p 336] But in a sense at least, the pure graph structure of a complex *is* a further "constituent", though it is unlike the others. **The poset of connected graph inclusion** contributes as a factor to our derived hierarchy structure as much as any external subsumption-poset on ordinary constituents. The graph structure has an algebraic *effect* like the other constituents in that it systematically affects the large scale structure of the inferential hierarchy of descriptions by supplying a *quotient* structure.

A direct product of posets (like the one above in Figure 5) is easily factorable into its constituents by "direct factorization". Inconveniently, though, the poset of possible descriptions would be a simple direct product of factor dimensions *only* when the relational graphs of the descriptions have no symmetries, that is, when a particular value of one part of the graph cannot make it indistinguishable from another part. This is the case only when the relational graph is *rigid* meaning that its only automorphism (structure-preserving permutation as a labeled graph) is the identity.

Although I haven't quite understood Parker-Rhodes' *Theory of Indistinguishables* [Parker-Rhodes 1981], my impression is that, for him, two related individuals are "indistinguishables" at the ultimate, omniscient level of knowledge, if they occupy positions in the semantic network describing the entire system (universe) such that their positions in the network are symmetric in just the way described here. In such an *omniscient* semantic network there is no ability to label individual nodes with *additional* information or knowledge; nodes are distinguishable only by their explicit positions in, and relation to, the existing network. If the network is symmetric with respect to two nodes, they are distinct (two, not one) but totally indistinguishable in principle. The same is true in purest unlabeled graph theory, in which no two graphs can be isomorphic (else they would be one); if such a graph has any symmetry, there will be "indistinguishable" vertices, in Parker-Rhodes' sense. This of course contradicts Leibniz's notion of the identity of indiscernibles.

THE SUBSUMPTION ORDER

What is needed to generate the full hierarchy of relational descriptions is an alternative to direct-product: a "product" of the various supplied structured ontological dimensions that takes account of the relational graph structure of a description *and its possible symmetries*. Since the goal is a hierarchy of descriptions ordered by abstractness, also called generality or subsumption, it is important to have a criterion to determine when one description is subsumed by another. For people it is often obvious: "a fat California lawyer whose four rowdy daughters continually hit one another" is obviously subsumed by the more general "a fat American". [3]

The up and down terminology can be a bit confusing, especially since posets of descriptions have been drawn both ways. In Artificial Intelligence and traditional taxonomies the most general class (with

---

[3] Ideally there should be a compact code for each structure, and operations on the codes should preserve useful logical operations on the structures. Leibniz in his *Ars Combinatoria* attempted this by assigning a different prime number to each primitive concept, so that each composite concept is a multiple of its constituent primes. In the resulting *divisibility lattice* a concept-number is subsumed by all its divisors. Thus arithmetic division achieves subsumption testing. He could have simply used bits for primitives to form a Boolean lattice of sets — the divisor lattice unnecessarily yields a structured space of *multisets* since different multiples of the same prime factor affect the resulting number. [see Brink 1987] Still, this preserves *too little* structure, as we have seen above, since all structure among the primitives is lost. Gödel, in an elaboration of Leibniz, used his *Gödel-numbers* to encode logical propositions; each character place in a (fully ordered) string of symbols is assigned a consecutive prime number, and each symbol of the logical alphabet is assigned a number to which a prime may be raised (i.e. *character-position* is raised to the *symbolth* power). All these powers of primes are multiplied together to get the Gödel-number of the whole sentence, which is unique because raising to exponential powers is not commutative or associative (e.g. $3^5 \neq 5^3$ whereas Leibniz was stuck with $3 \times 5 = 5 \times 3$). This preserves *too much*, since every different syntactic variant (sentence, description) with the same meaning gets a different Gödel-number. The desired *semantic* encoding lies between these extremes. This is where the "graph dimension" comes in.

the least information and the most exemplars) is put at the top and more specific classes go below, whereas in domain theory Dana Scott decided to put a class *with more information* higher than one with less, so the most general class goes at the bottom. I follow AI and tradition, so TOP is general. Another source of confusion is the "sub"s — because here every descriptive graph is *sub*sumed *by* its *sub*graphs.

Ignoring negation and disjunction, there are three sources of structure for the partially ordered set of description-graphs, namely:
1. the poset of graph inclusion (there are variants discussed below),
2. the external (supplied) posets or ontologies of subsumption of single concepts, and
3. the external (supplied) posets or relational hierarchies of subsumption of single relations;
these are involved in three conditions that must be met in order for one relation-net to subsume another.

1. **The subsumed graph must include the subsuming graph.** In the simplest case there must a *subgraph isomorphism* projecting the upper graph, as a pure graph structure, into the lower graph. (The lower, more specific graph may be larger and more complicated.) For graphs $G_1=(V_1,E_1)$, $G_2=(V_2,E_2)$, there is a subgraph isomorphism (subgraph embedding) between G_1 and G_2 iff, for some subgraph $G_2'=(V_2',E_2')$ of G_2, G_1 is nonempty and there is an injection $\phi:V_1\rightarrow V_2'$ such that $\{u,v\}\in E_1 \Rightarrow \{\phi(u),\phi(v)\}\in E_2'$. Without the requirement that ϕ be injective (i.e. that $\phi(u)=\phi(v) \Rightarrow u=v$), there is a *projection* of G1 into G2 which might collapse together two separate nodes, so it may or may not be a subgraph embedding.[4]

2. **For every concept-node in the upper graph, its corresponding concept-node in the lower graph must be labeled as representing either the same concept or else a finer specification of it** such that an order-relation "$\leq_c$" in an external subsumption poset on concepts is satisfied. The lower node must be a more specific concept-node than the upper one, as determined by their places in the external concept-hierarchy poset, unless they are the same concept. There is a concept label function $\mathrm{LABEL}_c: V\rightarrow N_c$ where $P_c=(N_c,\leq_c)$ is a poset, and for all vertices $v\in V_1$, $\mathrm{LABEL}_c(\phi(v)) \leq_c \mathrm{LABEL}_c(v)$.[5]

3. **For every relation-edge in the upper graph,**
A: its corresponding relation-edge in the lower graph must be labeled as representing either the same relation or else a finer specification of it such that an order-relation "$\leq_r$" in an external subsumption poset on the *relations* is satisfied. The lower relation-edge must be a more specific *relation* than the upper one, as determined by their places in the external *relation-hierarchy* poset, unless they are the same relation; and
B. if the upper relation is asymmetric (not invariant under permutation of arguments), the projection of the upper relation into the lower relation **must not permute**

---

[4] As an alternative to subgraph embedding, you may choose to base structural subsumption on pure (intensional) "projection" as is used in Conceptual Graphs, also called "noninjective projection" [Chein & Mugnier 1992]; in essence it ignores individual identity, unlike the subgraph embedding, by allowing two nodes in a subsuming graph to "project" into the same nodes of the subsumed graph. Thus [PERSON] -> (AGT) -> [MARRY] -> (OBJ) -> [PERSON] would project into [MAN:x] -> (AGT) -> [MARRY] -> (OBJ) -> [MAN:x], that is, a man x who marries himself. Whether and when to allow this is a philosophically important issue, discussed above in connection with nose-scratching.

[4] This requirement is present in Conceptual Graphs, Order-Sorted logics, Parker-Rhodes' inferential semantics, Ait-Kaci's types, and in some KL-ONE variants. In inferential semantics the nodes may have to satisfy several different external partial orders, on multiple node-labels; recently Page & Frisch have moved in this direction in their generalizations of order-sorted logics [Page & Frisch 1991].

distinctly labeled or otherwise distinguishable concept nodes which the upper relation asymmetrically relates.[6]

There is a relation label function LABEL_T: $E \rightarrow N_T$ where $P_T=(N_T,\leq_T)$ is a poset, and for all edges $\{u,v\} \in E_1$, $\mathrm{LABEL}_T(\{\phi(v),\phi(v)\}) \leq_T \mathrm{LABEL}_T(\{u,v\})$. The same applies to hyperedges for triadic and higher-valence relations.

Requirement B is technical. The subgraph isomorphism of point 1 above applies to *undirected* graph edges, but relational structures are directed, or, rather, the arguments of a relation are *indexed*. The requirement here is not as strict as requiring an index-preserving directed graph morphism, since it allows for harmless permutations of arguments where either the relation is symmetric with respect to those arguments, or the concept nodes in the argument positions are entirely indistinguishable (in the same symmetry orbit of the whole labeled graph) in the lower graph.

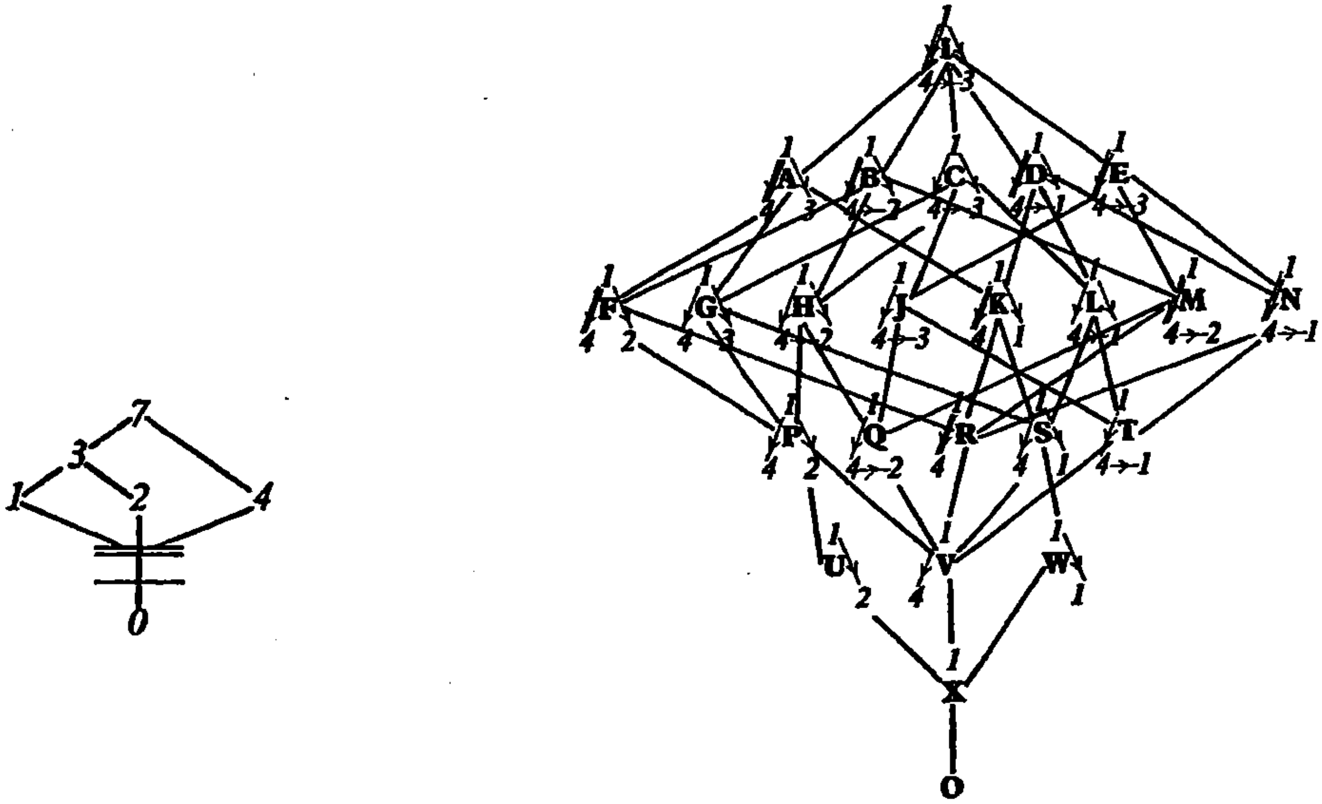
This is fairly standard for semantic networks and order-sorted logics [Sowa 1984, Aït-Kaci 1984, Parker-Rhodes 1978, Ellis 1991], but this graph-theoretic definition of subsumption of relational graphs avoids some notational artifacts of other systems. The explicit references to the poset of subgraph inclusion and the supplied poset of *relational* inclusion (from an external abstraction hierarchy of relations) are missing in most other systems or are left implicit.[7] Parker-Rhodes' notion of *multiple* labels on concepts representing multiple "aspects" or "facets" (hence multiple external subsumption lattices impinging on each node) is easily cured by adding those aspects as explicit predications in the relational graph.

It's clear from the following figure (in *Inferential Semantics*) that Parker-Rhodes was well aware of the essential construction of a type hierarchy induced by a graph labeled with labels from an ordered set. He chose to mix the hierarchy of relations with that of concepts, and as mentioned above he took pains to convert both the relational graph and the ontological dimensions ("base domains") of its labels into mathematical lattices. The latter was accomplished with a clever but nonetheless suspect "focus order" imposed on multiple relata of a relation. In doing this he avoided some difficulties of graph symmetries; I don't know if this was because of, or despite, his fundamental emphasis on symmetries in *The Theory of Indistinguishables* and in his theories of particle physics.

---

[6] A relation among three or more things may be symmetrical for some arguments and asymmetrical for others. The usual unordered meaning of Between(a,b,c), that a is between b and c, is invariant permuting b and c, but not permutations involving a.

[7] Using type hierarchies for relations, as opposed to concepts, is more rare.. NIKL and N-Ary KANDOR in the KL-ONE family include relational subsumption [Schmolze 1989] and it is discussed algebraically in [Brink & Schmidt 1992]. Ellis' and Levinson's work in Conceptual Graphs includes subgraph isomorphism and a relational hierarchy [Ellis 1991, Levinson 1992]. Most work on relational hierarchies has sprung from cognitive and lexical studies [Evens 1988] or has treated specific domains like spatiotemporal topology [Randell & Cohn 1992].



The structure of each rhema graph is indicated by *light* lines with arrows, and that of the *repertory* by heavy lines. 'I' represents the context 'rehearsal'.

Figure 7. Parker-Rhodes' induced lattice of lattice-labeled "rhema graphs". The large-scale structure (which he called the "context repertory") is determined by the the small "base domain lattice" at left from which the vertex and arc labels are taken. At left, the bottom three elements are arc labels "-","=", and "0", and the rest are vertex labels. His rhema graphs are rooted, directed, acyclic graphs (lattices in fact), a limitation which I do not obey.

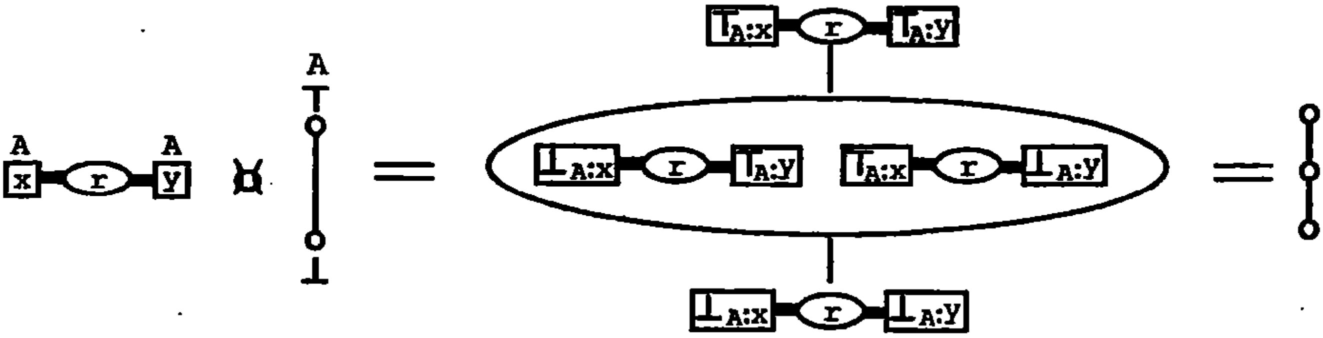
It's too bad more attention has not been given to*Inferential Semantics*, particularly Chapter V on "Labelled Graphs" and Chapters IX, X and XI on specific base domains (the Quantification Lattice, the pronomial Identification Lattice, and the Case Lattice displayed above in Figure 5). The obstacle to enthusiastic acceptance might have been the intimate mixture of very abstract mathematics (lattice and graph theory) with detailed attention to the vagaries of natural language, even intonation and prosody. This plus the author's convention-flouting independece of mind and penchant for justified but nonetheless obscuring neologisms.

In this article I'll specifically ignore negation and disjunction which were available in the Boolean lattice (as *complement* and *join* respectively) for predicates. Hence this system of relational structures is *vivid* in Levesque's sense and is not computationally vexed by the ability to nest negations arbitrarily. If we add negation and disjunction *of relations* we would get a relational algebra (technically a finitely generated cylindric set algebra, since we allow *n*-adic relations and not just the dyadic ones of a "relation algebra" in Tarski's sense); "relative products" (relational composition) and "diagonals" (equations between already related individuals) are already covered by the graph inclusion ordering. [Henkin et al. 1971-85, Burch 1992] .

In the graph dimension, symmetry and group theory enter the picture, and the new "products" and corresponding factorizations must take account of them

Two "product"-like operators: Skeleton Product and Fret Product

In order to establish the second-order structure of factors in the description hierarchy, a "product"-like operator "⋈" called **skeleton product** applies a controlling *poset-labeled graph* to one or more named hierarchies (subsumption posets) corresponding to the labels on the concept nodes and relation hyperedges of the controlling graph. This equals the direct product of hierarchies only in case the controlling relational graph is completely asymmetric. Next, a second "product"-like operator "X" called **fret product** handles subgraph isomorphism in comparison between two different controlling graphs (with its attendant group theoretic aspects), thus factoring the "graph dimension" poset out of every description. The poset of injective subgraph inclusion, by no means a lattice, is designated "$" and shown in Figure 12. Both products are formally defined below, but first a few simple examples.



$G = R(A{:}x, A{:}y)$, $A = (\{\top, \bot\}, \leq_A)$, $G \bowtie A = (\{R(\top, \top), \{R(\bot, \top), R^{-1}(\top, \bot)\}, R(\bot, \bot)\}, \leq)$

Figure 8

This is the simplest nontrivial skeleton product. The controlling relational graph at left, of symmetric R(x,y), has both related individuals labeled as members of the same type hierarchy, the "A" dimension. (To simplify the examples, R is untyped.) It is "multiplied" by the hierarchy "A" consisting only of types TOP and BOTTOM. The product is the three-element chain at right. Labeling the vertices with two disjoint source posets A and B makes the graph "rigid" and the result is the direct product A × B shown below in Figure 9.

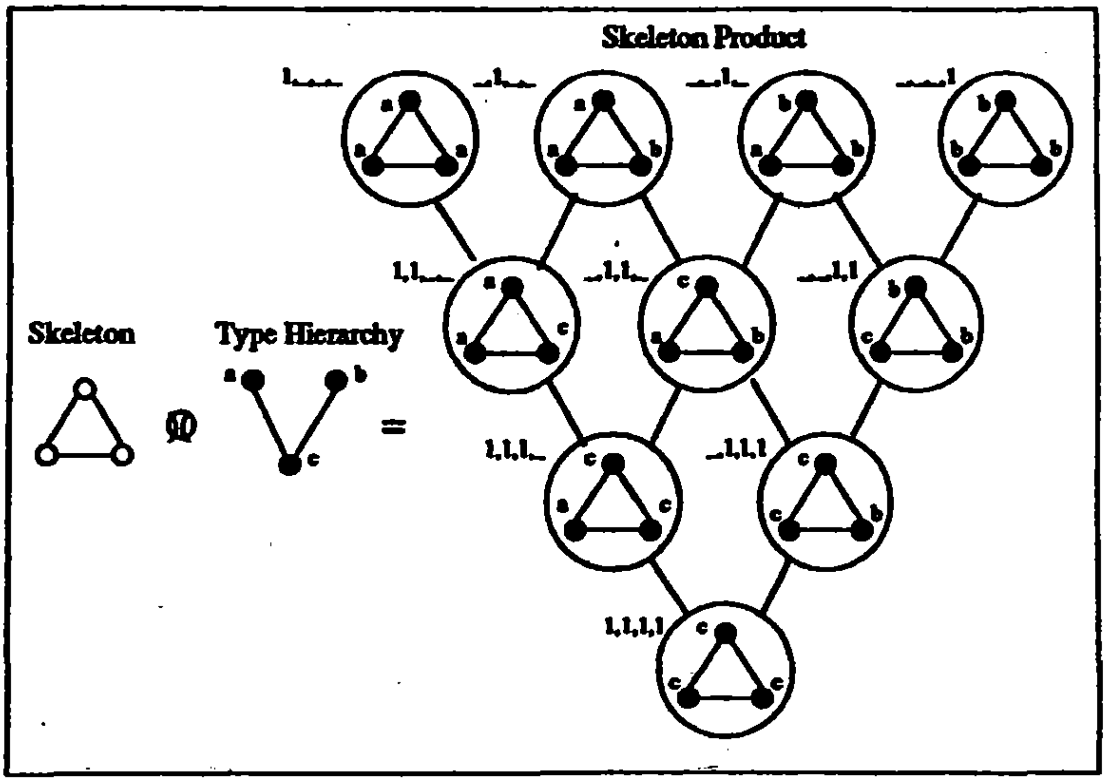


FIGURE 8A

The skeleton product X

of a triangle graph

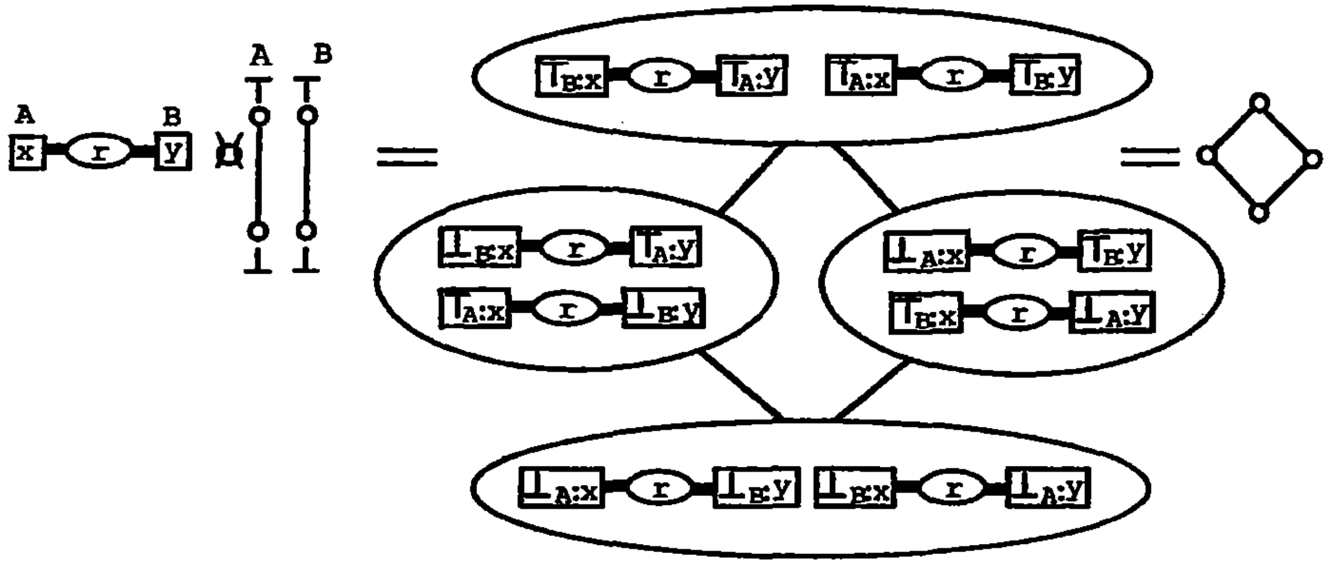


Figure 9

At the abstract level (without notational artifacts), there are no symmetric variants, and no "converses" of relations. To say EastOf(Missouri,Kansas) is to assert the *same relation* as in WestOf(Kansas,Missouri).[8] This is easily pictured by imagining any n-adic relation as a node (with n labeled attachment-spokes) floating in abstract space, as suggested by the Russell quote at the beginning of this article. Orientation of the node is irrelevant, and "converse" is a matter of linear notation, not fundamentals. The permutation of arguments is dealt with in the graph subsumption rule 3B above.

The skeleton product yields the poset of all descriptions in an equivalence class determined by common graph structure. Figure 10 is a skeleton product "multiplication table" for poset structures, in which the controlling graph is labeled from a single source poset with one to four nodes. Apparently the shape of the product structure depends on the automorphism groups of the controlling graph and of the constituent posets. The automorphism group of the three element chain Ch3 is S2 × S1 so the skeleton product Ch3 ⊗SP = [K2 ⊗SP] × [K2 ⊗SP], as is evident from the diagrams in the table (where "× " means direct product).

The number of nodes in the skeleton product of graph G and one poset SP is equal to the number of ways to label the graph with |SP| not-necessarily-distinct labels. [See Harary & Palmer 1973] The number does not depend on the structure of SP.[9]

Subsumption of differing graph structures depends on the *fret product* of the poset of graph inclusion (not just a single graph) and the source posets.

Symmetry and Connectedness

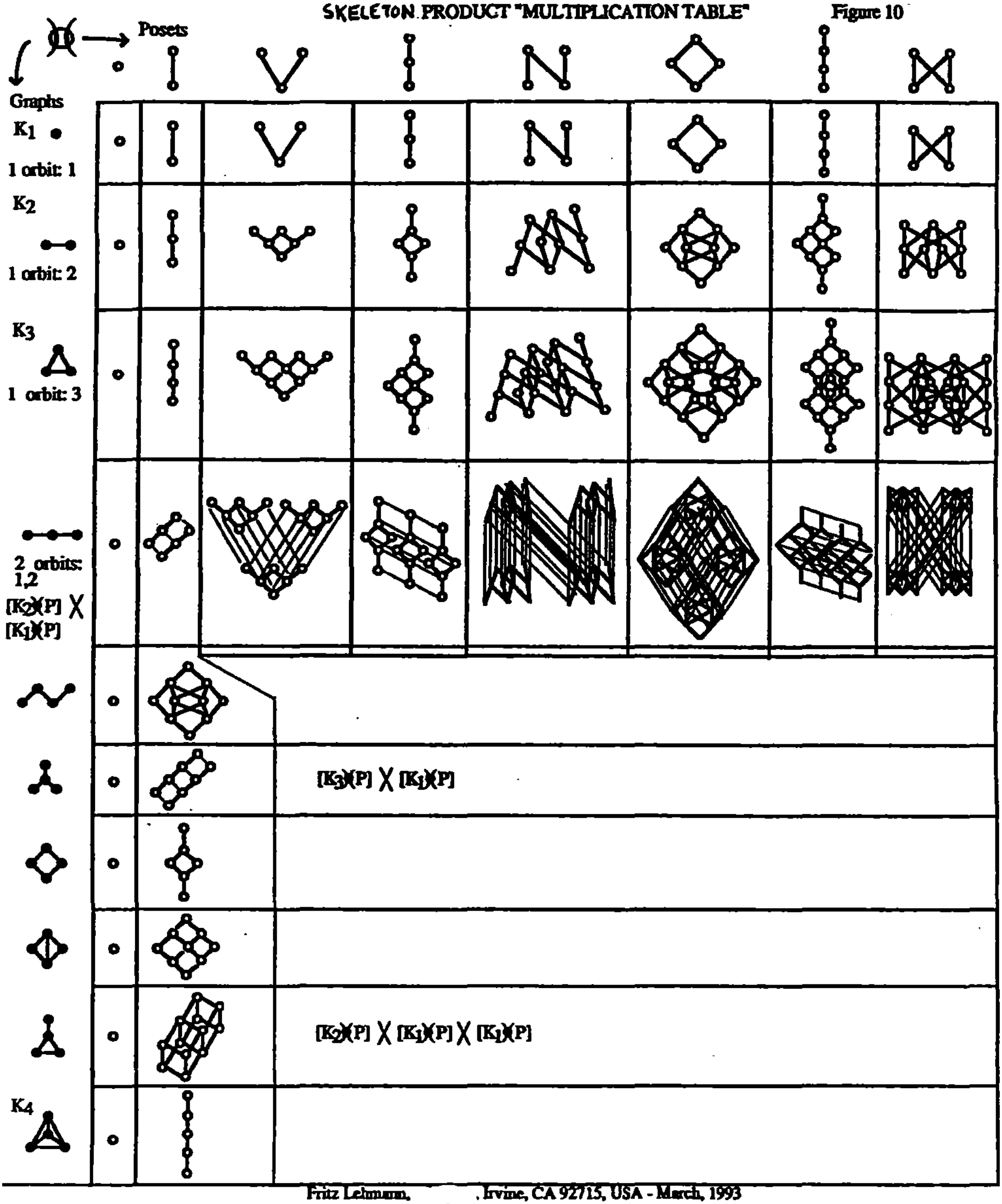
Although the inclusion poset is mathematically derived, its form is affected by *pragmatic ontological* decisions: First, it ignores symmetric variants, since we don't care about syntax. Second, it concerns only connected graphs, since no definition of an "x" is of the form: "An x is something which ____________ and there are many towns and cities in the world." All of a definition of an x must be

---

[8] This is a substantial simplification for higher valence relations; any triadic relation has five "converses" and a tetrad 23. There is generally one for each permutation of arguments other than the identity, as C.S.Peirce noted

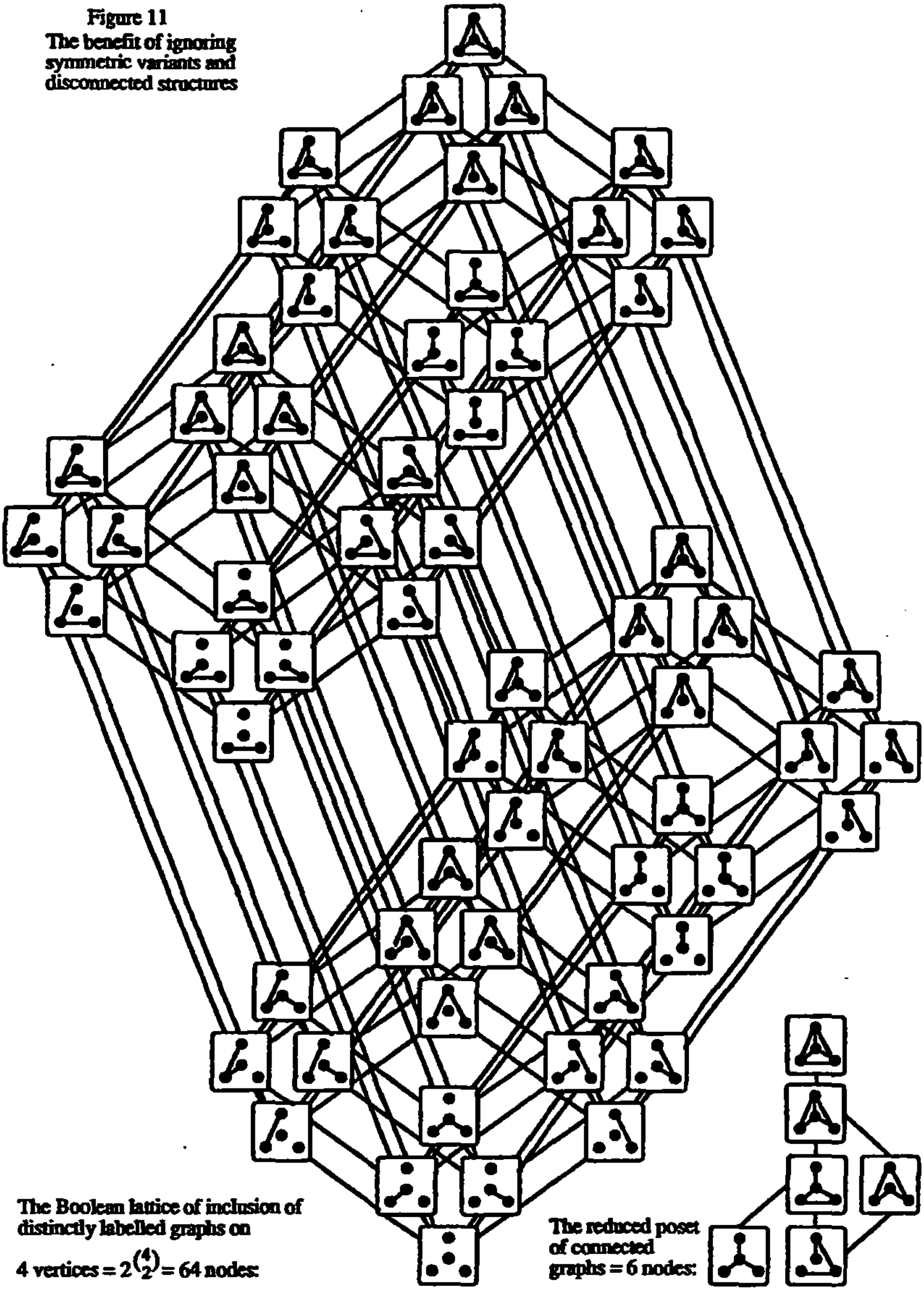
[9] Looking at Figure 10, my mother spotted the fact that that the posets in the third column of are the same as those in the second column, only turned sideways. I don't know whether this would always be true, or why.

SKELETON PRODUCT "MULTIPLICATION TABLE" Figure 10



Fritz Lehmann, 4282 Sandburg, Irvine, CA 92715, USA - March, 1993

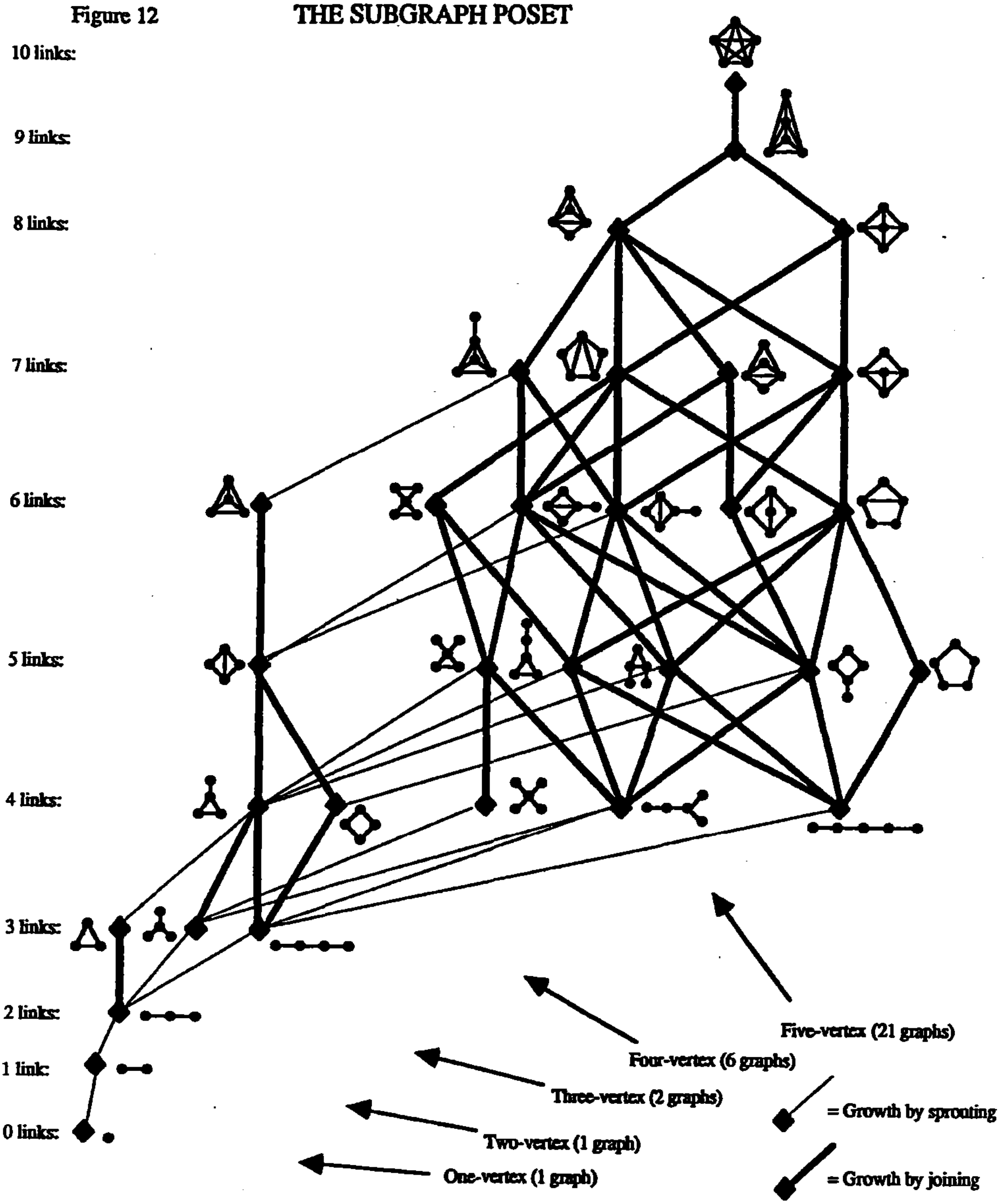
Figure 11
The benefit of ignoring symmetric variants and disconnected structures



The Boolean lattice of inclusion of distinctly labelled graphs on 4 vertices = $2^{\binom{4}{2}}$ = 64 nodes:

The reduced poset of connected graphs = 6 nodes:

Figure 12 THE SUBGRAPH POSET



16

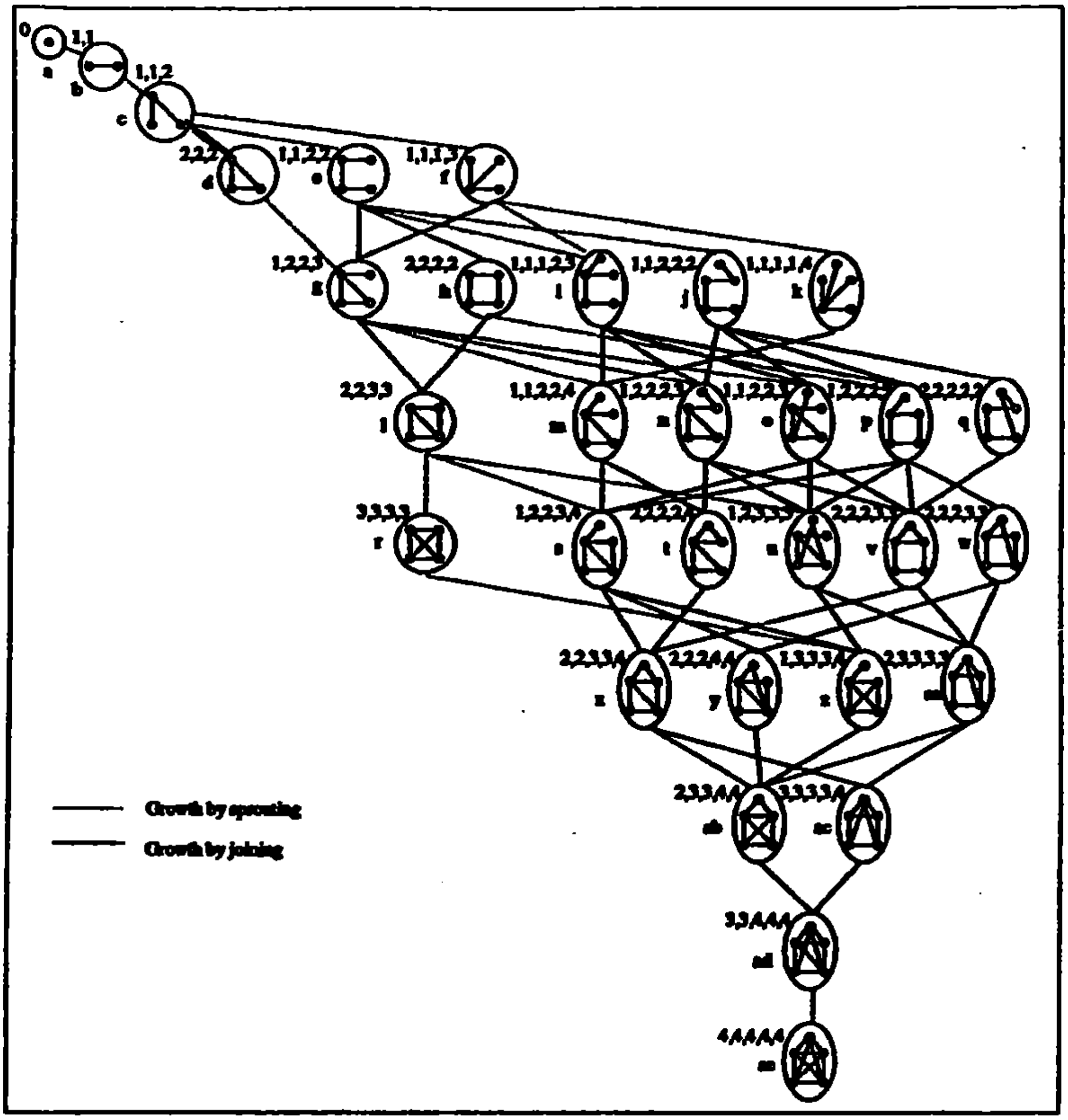


FIGURE 12A The names of the graphs in the poset of unlabeled connected graphs with 5 nodes or less, $\mathfrak{S}_5$ [Leh93] linked to x by some relational structure. These two pragmatic dictates vastly reduce the size of the graph inclusion poset. Without them, there would be 64 different graphs on 4 vertices, in the Boolean lattice on edges shown in Figure 11, but with these restrictions the number is reduced to six — the saving is even more dramatic as the number of vertices grows, although both grow exponentially. The number of *graphs on n distinctly labeled vertices* is $2^{\binom{n}{2}}$; the number of *connected graphs* is given by large equations [see Harary and Palmer 1973, Stein & Stein 1967]. For small values of n, the sequences are

n=	1	2	3	4	5	6	7
	2	4	8	64	1024	32768	2097152
	1	1	2	6	21	112	853

. The reduced poset of Figure 11 is in the center of $ (Figure 12).

Third, the resulting structure depends on a number of limitations on the semantic network graphs in question: whether the graphs include loops, whether they are restricted to, say, bipartite graphs (as would be useful for Conceptual Graphs), and whether non-injective graph projection is the basis of graph inclusion rather than injective subgraph embedding as a consequence of dropping the presumption of non-identity of distinctly named individuals.

Figure 12 shows the poset of injective graph inclusion for connected graphs, here denoted $. The poset for graphs with loops is a simple variant of $ and so is the poset of directed graphs. I believe this object $ will turn out to control the others. (Although the poset of non-injective directed hypergraph inclusion is different.)

The two kinds of link in the diagram of this poset, "twigs" and "joiners", correspond to the operations j1 and j2 (relative product of a relational structure or existential graph with itself or with a new predicate, respectively) described by Burch for Peircean Algebraic Logic (a form of Peirce's Existential Graphs), and independently defined by Chein and Mugnier for operations on Conceptual Graphs. [Burch 1992, Chein & Mugnier 1992] These correspond to diagonal elements in Tarski's Cylindric Algebras (or equations between variables) and composition of relations. [Henkin et al. 1971-85] They are also the two graph-grammar production rules needed to generate all graphs. This poset has various kinds of structures determining its particular form. The nodes are *ranked* by the number of edges *m* in a graph. The nodes then form families sharing a common number of vertices *n*. The number of graphs for each combination of *m* and *n* is given in [Harary & Palmer 1973]. While the twig links projecting up from every poset node count the *vertex orbits* in the graph, the joiner links projecting up count the *edge orbits* in the *complimentary* graph. The twig links projecting down count the univalent vertex orbits; the joiner links projecting down count the edge orbits, except for isthmus edges since disconnected graphs are excluded. Maybe $ has a further skeleton related to the subgroup-ordered poset of finite groups. The structure of each "family" of graphs on *n* vertices dangles like a marionette from its spanning trees, with the complete graph K_n hanging at the bottom.

The fret product: $ $\ddot{\times}$ SP itself is like $, but with G $\ddot{\otimes}$ SP substituted for each G occurring as a node of $, and with all links between nodes respecting $ as a *congruence* on the partial order. Figure 13 shows the simplest non-trivial fret product: $4 $\ddot{\times}$ 2, where $4 is the poset of graph inclusion for graph with up to 4 vertices and 2 is the simple 2-elment poset. it' a complex poset of 65 nodes arranged in 10 skeleton-product quotients. It is easy to see that this structure is a generalization of the "context rehearsals" of Parker-Rhodes a shown in Figure 7.

Formal Definitions of Skeleton Product and Fret Product

In the following, "vertices" refer to the vertices or points in a graph, whereas "nodes" refers to the members or points of a partially ordered set (poset). If a graph G = (V,E), where the V are its vertices and the E are its edges, then I will use V(G) for the set of vertices of G and E(G) for the set of edges of G.

The Simplest Fret Product: $\$_4 \divideontimes \mathbf{2}$

Subgraph inclusion (injective) of undirected graphs, all nodes labeled from the two-element poset, for $n <= 4$.

Poset $\mathbf{2} = \begin{smallmatrix}1\\ | \\ 0\end{smallmatrix}$

The universal structure fret factor:

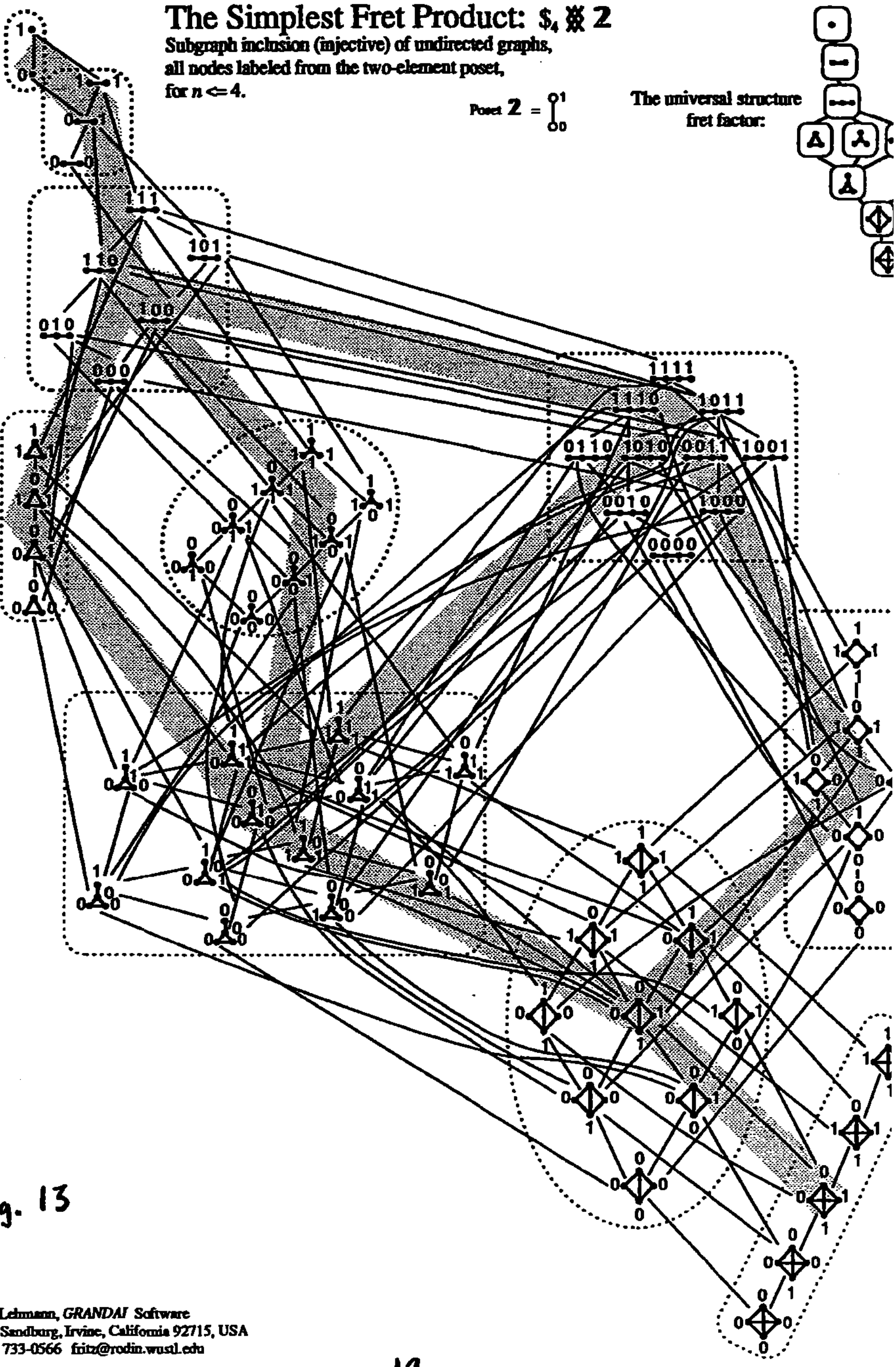


Fig. 13

Fritz Lehmann, *GRANDAI* Software
4282 Sandburg, Irvine, California 92715, USA
(714) 733-0566 fritz@rodin.wustl.edu

The *skeleton product* takes some particular graph and a set of posets (type hierarchies) and generates the subsumption poset of descriptions each of which has that graph as its relational structure. With one source poset it is the simply the induced order on a poset-valued graph. The *fret product* takes a set of posets (type hierarchies) and (using ta universal poset of graph-inclusion) generates the poset of all description-graphs (i.e. with all relational graph structures), ordered by subsumption. The vertex labels in the resulting description-graphs are types in the original type hierarchies. There is no provision here for negation or disjunction. Note that these are extended to handle directed *hypergraphs* (general relational structures) when triadic and higher relations are included.

1. **Skeleton Product.** The skeleton product is an operation on a set of named partially ordered sets (source posets) and a "controlling graph" with vertices each labeled with the name of a source poset. It yields a derived *poset of graphs* made of vertex-labeled graphs in which the resulting labels on the vertices of each graph are nodes from the source posets. The resulting order on the graphs is the subsumption order on a class of descriptions sharing a particular graph structure.

Let PLG, for "Poset Labeled Graph", be a graph and let SP, for "Source Posets", be a set of m disjoint posets $\{(P_1,\leq_1),(P_2,\leq_2),...(P_m,\leq_m)\}$. The vertices in PLG are each labeled with the name of a source poset in SP. The resulting skeleton product P is a poset of vertex-labeled copies of the graph of PLG; within every copy, each vertex label is a member (node) of the source poset for that vertex. The orderings of the source posets interact (by means of the controlling graph) to determine the resulting ordering on the product poset of graphs P.

DEFINITION 1: **The skeleton product operation** Φ is

Φ: *Poset-labeled graph* × *Set of posets* → *Poset of graphs*

in infix notation, PLG Φ SP = P, where

SP is a set of m disjoint posets $\{(P_1,\leq_1),(P_2,\leq_2),... (P_m,\leq_m)\}$, and

G is a graph (V,E), $|V|=n$, and PLG = G with poset-labels $P_1,P_2,...P_n \in$ SP, such that every vertex of PLG has one poset-label, i.e. there is a function PosetLabel:V(G) → SP, and

L is a set of vertex-labels and every vertex-label in L is also a member (node) of exactly one poset in SP, and

there are one or more graph-automorphisms Aut_k:G → G on G (irrespective of any labels), k=1...q where $|Aut(G)| = q$, including the identity, and

for graphs x and y and for each $Aut_k(x)$, if $y=Aut_k(x)$ then the function Im_k:V(x) → V(x) yields the image in y of each vertex in x under $Aut_k(x)$, and

the skeleton product $P = PLG\ \Phi\ \{(P_1,\leq_1),(P_2,\leq_2),... (P_m,\leq_m)\}$ is a poset $(P,\leq_P)$ such that, for $a,b \in P$, $a \leq_P b$ if and only if

a is the graph G with vertices $v_1,v_2...v_n$ labeled
with vertex labels $l_{a1}, l_{a2},...l_{an} \in L$, respectively,
b is the graph G with vertices $v_1,v_2...v_n$ labelled
with vertex labels $l_{b1}, l_{b2},...l_{bn} \in L$, respectively,
$VertexLabel_a$:V(G) → L and $VertexLabel_b$:V(G) → L are the resulting label functions on vertices of a and b, and
for some automorphism $Aut_k(G)$, k=1...q,

for every vertex v_i in a, i = 1...n,
  VertexLabel$_a$(v_i) is a node in the source poset PosetLabel(v_i),
    and VertexLabel$_b$(Im$_k$(v_i)) in b is a node
      in the source poset PosetLabel(Imk(vi)),
  PosetLabel(v_i) = PosetLabel(Imk(v_i)) = $P_j \in$ SP
  and VertexLabel$_a$(v_i) $\leq_j$ VertexLabel$_b$(Im$_k$(v_i)).

All the different source posets can be put in one big one by adding global top and bottom nodes. (The "separated sum" or the "fused sum" of the posets.) This adds extra, artificial nodes to the skeleton and fret products, though.

2. <u>**Fret Product.**</u> Whereas the skeleton product yields the subsumption hierarchy for descriptions sharing the *same* relational graph structure, the fret product yields the subsumption hierarchy for descriptions with *any* relational graph structures.

Any description graph is necessarily subsumed by all of its subgraphs. The fret product takes account of subgraph inclusion, by using the general poset of subgraph inclusion as the "controller" which relates the source posets. The top of the fret product is a single vertex labeled with the top node in the source poset. Since the poset of subgraph inclusion is infinite, accommodating infinitely large graphs, you may choose to use only the upper part that includes graphs with no more vertices than in the largest relational graph under consideration. I denote the poset of undirected graphs ordered by injective subgraph inclusion as "\$", or for graphs with n or fewer vertices "$\$_n$". As a technical matter, alternate graph-inclusion posets may be used for multigraphs, graphs with loops, noninjective projections, etc. Each has application to all kinds of descriptions.

Let SP be a set of source posets $\{(P_1,\leq_1),(P_2,\leq_2)...(P_m,\leq_m)\}$ as before. Let $\$ = (\$,\leq_\$)$ be the poset of connected graphs ordered by subgraph inclusion (also called graph homomorphism, subgraph isomorphism, injective projection, or embedding).

DEFINITION 2: The fret product operation ¥ is

¥: *Poset of graph inclusion* × *Set of posets* → *Poset of graphs*

in infix notation, \$¥SP = P, where
SP is a set of disjoint posets $\{(P_1,\leq_1),(P_2,\leq_2),...(P_m,\leq_m)\}$, and
\$ is the poset $(\$,\leq_\$)$ of unlabeled connected graphs ordered by
  subgraph inclusion (i.e. for connected graphs G and H, $G \leq_\$ H$ if
  and only if $V(G) \subseteq V(H)$ and there is a function $f:V(G) \rightarrow V(H)$
  such that for all $u,v \in V(G)$, if (u,v) is an edge in E(G) then
  (f(u),f(v)) is an edge in E(H)).
L is a set of vertex-labels, Lab a function Lab:$V(G) \rightarrow L$, and
  every vertex-label $Lab(v) \in L$ is also a member (node) of
  exactly one source poset in SP, and Src:$L \rightarrow SP$ is a
  function such that $Src(Lab(v)) \in SP$ is the name of that
  source poset, and
the fret product P = \$¥ $\{(P_1,\leq_1),(P_2,\leq_2),...(P_m,\leq_m)\}$ is a
poset $(P,\leq_P)$ such that, for $a,b \in P$, $a \leq_P b$ if and only if

a is a graph G_a labeled with vertex-labels in L,
b is a graph G_b labeled with vertex-labels in L,
G(a) and G(b) are the unlabeled graphs of a and b,
$G(a) \in \$$ and $G(b) \in \$$ and $G(a) \leq_{\$} G(b)$,
and for some injective graph homomorphism $f{:}V(G_a) \rightarrow V(G_b)$,
for all $v \in V(G_a)$, $Src(Lab(v)) = Src(Lab(f(v))$ and
$v \leq_{Src(Lab(v))} f(v)$.

If $G(a) = G(b)$, then $a \leq b$ in the fret product if and only if $a \leq b$ in the skeleton product $PLG \overset{\times}{\Phi} SP$ where $SP = \{Src(Lab(v)) | v \in V(G(a))\}$ and PLG is the poset-labeled graph in which every vertex-label in a (or in b) is replaced by Src(l).

Discussion and Further Work

The work here suggests that every description may be "fret-factorable" and that the poset \$ of graph inclusion (or that of directed hypergraph inclusion, injective or noninjective) is a universal element of every system of finite descriptions or statements in a language or logic. The two real-world facts that A: only connected semantic graphs matter, and B: symmetric variants don't matter, have a drastic and peculiar effect of the hierarchy. The resulting structure is ontological, rather than merely logical, and it is even partly epistemological in the sense that it is our limitation to finite relational conceptual structures (rather than "reality") which induces this particular kind of structure.

Ontological classification, like all classification, throws thing into equivalence classes based on some purpose for which the differences among objects in a class don't matter. A common ontology is useful to the extent that users share common purposes. Since purposes themselves occur in abstraction (and other) hierarchies, the hierarchy of purposes (tasks, missions, goals) presumably systematically alters the second-order structure of the full concept hierarchy. It would be good if formal changes in purpose could *automatically* alter the ontology in force.

For computers, doing subgraph isomorphism tests the hard way is believed to be intractable. For AI, my idea for the poset "\$" (or its variants) is to precompile the poset of its upper reaches into bit-strings embedded in a compressed Boolean lattice à la [Aït-Kaci et al. 1989], and tag all short to medium-length assertions with their graph bit-codes. Direct parallel logical bit-wise comparison of bit-strings would instantly answer the subgraph isomorphism test for small to medium graphs, and find (sets of) maximal lower bounds for unification. Used as a *filter* for isomorphism testing, it would eliminate isomorphism tests for almost all graphs of any size encountered in semantic networks, based on the *Pattern Associativity* work of [Levinson 1992]. Gerard Ellis has a similar notion of bit-string encoding of *incrementally compiled* posets of Conceptual Graphs, in the style of Levinson's Pattern Associativity [Ellis 1993]; he is now collaborating with me to pre-code the poset of *all* graphs up to a certain size [Ellis & Lehmann 1994]. For Boolean embeddings of posets see [Wille 1982, Aït-Kaci et al. 1989, Cohn 1992, Caseau 1993].

I calculated the skeleton and fret products by hand, so to save effort I confined this paper to undirected graphs ordered by injective subgraph isomorphism. Properly, it should be based on n-adic directed hypergraphs, ordered injectively or noninjectively depending on whether individuals with distinct names are presumed to be distinct. The system might then be extended to include nested negations of any part of a semantic net (hence covering all of predicate logic, as Peirce's Existential Graphs show [Roberts 1992]) by adding another "participant" sub-hierarchy dimension which would allow arbitrary negations; this seems to have been done already for Boolean operations on Aït-Kaci's epsilon types by Lucja Iwanska [Iwanska 1992].

A further goal is automatic factoring of an arbitrary system into its smallest skeleton or fret product of automatically generated "ontological dimensions" or "aspects".

REFERENCES

H. Aït-Kaci, R. Boyer, P. Lincoln & Roger Nasr, Efficient implementation of lattice operations, *ACM Transactions on Programming Languages and Systems*, 11(1)115-146, 1989

R. J. Brachman, On the Epistemological Status of Semantic Networks, in *Associative Networks*, N.V. Findler, Ed., Academic Press, Orlando, Fla., 1979

C. Brink, Some Background on Multisets, TR-ARP-2/87, Automated Reasoning Project, Australian National Univ., 1987

C. Brink & R. Schmidt, Subsumption Computed Algebraically, in [Lehmann 1992]

R. Carnap, *Logical Foundations of Probability*, Univ. of Chicago Press, Chicago, 1950

Y. Caseau, Efficient Handling of Multiple Inheritance Hierarchies, in OOPSLA-93, Washington, D.C., Sep. 1993

N. Cercone, R. Goebel, J. de Haan & S. Schaeffer, *The ECO Family*, in [Lehmann 1992]

M. Chein & M. L. Mugnier, Specialization, Where Do The Difficulties Occur?, *Proc. 7th Annual Workshop on Conceptual Graphs*, New Mexico State Univ. CRL, Las Cruces, N.M., 1992

A. G. Cohn, Completing Sort Hierarchies, in [Lehmann 1992]

B. A. Davey & H. A. Priestly, *Introduction to Lattices and Order*, Cambridge Univ. Press, Cambridge, 1990.

G. Ellis, Efficient Retrieval from Hierarchies of Objects Using Lattice Operations, *Conceptual Graphs for Knowledge Representation*, (Proc. International Conference on Conceptual Structures, Quebec City, Canada, August 4-7, 1993), G. W. Mineau, B. Moulin & J. Sowa, Ed.s, Lecture Notes in Artificial Intelligence 699, Springer, Berlin 1993.

M. W. Evens, Ed., *Relational Structures of the Lexicon*, Cambridge Univ. Press, Cambridge 1988

A. M. Frisch & A. G. Cohn, Thoughts and afterthoughts on the 1988 Workshop on the Principles of Hybrid Reasoning, *AI Magazine*, 11(5), 77-87, 1991

B. Ganter & R. Wille, Conceptual Scaling, in *Applications of Combinatorics and Graph Theory to the Biological and Social Sciences*, F. Roberts, Ed., Springer, New York, 1989

N. Guarino & L. Boldrin, Concepts & Relations, *International Workshop on Formal Ontology*, LADSEB/CNR, Padua, Italy 1993

F. Harary & E. M. Palmer, *Graphical Enumeration*, Academic Press, New York, 1973

A. Hantamäki, Points of View and their Logical Analysis, *Acta Philosophica Fennica*, 41, 1986

L. Henkin, J.D. Monk & A. Tarski, *Cylindric Algebras*, North-Holland, Amsterdam, part I 1972, part II 1985

L. Iwanska, A General Semantic Model of Negation in Natural Language,, in *Proc. KR-92, Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, Cal. 1992

F. Lehmann, Ed., *Semantic Networks in Artificial Intelligence*, Pergamon Press, Oxford, 1992; also in Special Issue on Semantic Networks in Artificial Intelligence, *Computers & Math. with Applications*, v.23(2-9), 1992

R. Levinson, Pattern Associativity and the Retrieval of Semantic networks, in [Lehmann 1992]

S. Lytinen, Conceptual Dependency and its Descendants, in [Lehmann 1992]

O. Mariño, F. Rechenmann & P. Uvietta, Multiple perspectives and classification mechanism in object-oriented representation, in *Proc. ECAI-90*, European Conference on Artificial Intelligence (Stockholm), Pitman, London, 1990.

R. N. McKenzie, G. McNulty & W. Taylor, *Algebras, Lattices, Varieties*, Wadsworth, Belmont, Cal., 1987

C. D. Page & A. M. Frisch, Generalizing Atoms in Constraint Logic, in *Proc. KR-91, Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, Cal. 1991

A. F. Parker-Rhodes, *Inferential Semantics*, Harvester Press/Humanities Press, Hassox, Sussex, Atlantic Highlands, N.J., 1978

A. F. Parker-Rhodes, *The Theory of Indistinguihables*, Reidel, Dordecht, 1981

O. M. Polyakov & V. V. Dunaev, Classification Schemes: Synthesis through relations, *Nauchno Technicheskaya Informatsiya*, *Seria 2*, 19(5), 15-21, (translated as *Automatic Documentation and Mathematical Linguistics* publ. by Allerton Press, New York), 1985

D. A. Randell & A. G. Cohn, Exploiting Lattices in a Theory of Space and Time, in [Lehmann 1992]

S. R. Ranganathan, *The Colon Classification*, 6th Ed., Rutgers State Univ., New Brunswick, N.J., 1965

A.A. Raskina, I. S. Sidorov and Yu. Shreider, Semantic Foundations of Object/Feature Languages, *Nauchno Technicheskaya Informatsiya, Seria* 2, 10(5), 18-25, (translated as *Automatic Documentation and Mathematical Linguistics* publ. by Allerton Press, New York), 1976

F. S. Roberts, *Measurement Theory with Applications to Decisionmaking, Utility and the Social Sciences*, Addison-Wesley, Reading, Mass., 1979

V. Sh. Rubashkin, Attribute and Value, *Nauchno Technicheskaya Informatsiya, Seria* 2, 3, 3-10, (translated as *Automatic Documentation and Mathematical Linguistics* publ. by Allerton Press, New York), 1976

B. Russell, *Principles of Mathematics*, 2nd Ed., London, 1903

R. Schank, *Conceptual Information Processing*, North-Holland, Amsterdam, 1975

M. Schmidt-Schauß, Subssumption in KL-ONE is undecidable, in *Proc. KR-89, Principles of Knowledge Representation and Reasoning*, R. J. Brachman, H. J. Levesque 7 R. Reiter, Ed.s, Morgan Kaufmann, San Mateo, Cal., 1989

J. Schmolze, Terminological Reasoning Systems Supporting N-ary Terms, in *Proc. KR-89, Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, Cal. 1989

J. Searle, Determinables and the notion of resemblance, *Proc. of the Aristotelian Society*, Supp. Vol. 33, 141-158, 1959

P. Simons, *Logic and Philosophy in Central Europe from Bolzano to Tarski*, Kluwer, Dordecht, 1992

J. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Reading, Mass., 1984

J. Sowa, Semantic Networks, in *Encyclopedia of Artificial Intelligence*, S. Shapiro, Ed., John Wiley, New York, 1987

J. Sowa, *Principles of Semantic Networks*, Morgan Kaufmann, San Mateo, Cal., 1991

J. Sowa, Conceptual Graphs as a Universal Knowledge Representation, in [Lehmann 1992]

M. L. Stein & P. R. Stein, *Enumeration of Linear Graphs and Connected Linear Graphs up to p = 18 Points*, Report LA-3775. Los Alamos Scientific Laboratory, Los Alamos, New Mexico, Oct. 1967

E. Way, *Knowledge Representation and Metaphor*, Kluwer, Dordecht, Holland, 1991

Y. Wilks & D. Fass, The Preference Semantics Family, in [Lehmann 1992]

R. Wille, Restructuring Lattice Theory, in *Ordered Sets*, I. Rival, Ed., NATO ASI No. 83, Reidel, Dordecht, Holland, 1982

R. Wille, Concept Lattices and Conceptual Knowledge Systems, in [Lehmann 1992]

L. Wittgenstein, *Logical Notebooks* 1914-1916, 2nd Ed., Blackwell, Oxford 1979

W. Woods & J. Schmolze, *The KL-ONE Family*, in [Lehmann 1992]

M. Zickwolff, *Rule Exploration: First Order Logic in Formal Concept Analysis*, Dissertation, Dept. of Math., Technische Hochschule Darmstadt, Darmstadt, Germany, 1991

24

To appear in proceedings of
TAPSOFT'95
Theory and Practice of
Software Development
Aarhus, Denmark
May 22-26, 1995

Rational Mechanics and Natural Mathematics

Vaughan Pratt*
Dept. of Computer Science, Stanford University, Stanford, CA 94305-2140
pratt@cs.stanford.edu

February 17, 1995

Abstract

Chu spaces have found applications in computer science, mathematics, and physics. They enjoy a useful categorical duality analogous to that of lattice theory and projective geometry. As natural mathematics Chu spaces borrow ideas from the natural sciences, particularly physics, while as rational mechanics they cast Hamiltonian mechanics in terms of the interaction of body and mind.

This paper addresses the chief stumbling block for Descartes' 17th-century philosophy of mind-body dualism, how can the fundamentally dissimilar mental and physical planes causally interact with each other? We apply Cartesian logic to reject not only divine intervention, preordained synchronization, and the eventual mass retreat to monism, but also an assumption Descartes himself somehow neglected to reject, that causal interaction within these planes is an easier problem than between. We use Chu spaces and residuation to derive all causal interaction, both between and within the two planes, from a uniform and algebraically rich theory of between-plane interaction alone. Lifting the two-valued Boolean logic of binary relations to the complex-valued fuzzy logic of quantum mechanics transforms residuation into a natural generalization of the inner product operation of a Hilbert space and demonstrates that this account of causal interaction is of essentially the same form as the Heisenberg-Schrödinger quantum-mechanical solution to analogous problems of causal interaction in physics.

1 Cartesian Dualism

The Chu construction [Bar79] strikes us as extraordinarily useful, more so with every passing month. Elsewhere we have described the application of Chu spaces to process algebra [GP93], metamathematics [Pra93, Pra94a], and physics [Pra94b]. Here we make a first attempt at applying them to philosophy.

It might seem that traditional philosophical questions would be beyond the scope of TAPSOFT. Bear in mind however that Boolean logic as the basis for computer circuits was born of philosophy (and a little statistics). Only slightly more recently, program verification has drawn heavily on more sophisticated logics such as first order, modal, and higher order. Computers being thinking machines, computer science should not neglect the philosophical literature on thinking. It is easy to dismiss "all that stuff" as obsoleted by technology. However good truths, like good wine, must be served at the proper time. We would like to think of our application of Chu spaces to Descartes' inspiring yet short-lived theory of mind-body dualism as a convincing example.

---

*This work was supported by ONR under grant number N00014-92-J-1974

Reprinted from PROC. ANPA WEST 11 (1995)

Cartesianism is a "philosophy of everything" founded by René Descartes in the 1630's. Its point of departure was to reject all authority and question everything including the questioner's existence. Descartes resourcefully bootstrapped himself back into existence with an instance of the liar paradox, the absurdity of questioning his own questioning, constructivized as *Cogito, ergo sum*. Emboldened by this success, Descartes posed many more questions whose imaginative answers formed the basis of Cartesianism. This rationalist philosophy flourished for half a century until the march of science contradicted too many of its answers for it to remain a viable grand unified theory of anything. Some of the questions however remain philosophically challenging even today.

A central tenet of Cartesianism is mind-body dualism, the principle that mind and body are the two basic substances of which reality is constituted. Each can exist separately, body as realized in inanimate objects and lower forms of life, mind as realized in abstract concepts and mathematical certainties. According to Descartes the two come together only in humans, where they undergo *causal interaction*, the mind reflecting on sensory perceptions while orchestrating the physical motions of the limbs and other organs of the body.

The crucial problem for the causal interaction theory of mind and body was its mechanism: how did it work?

Descartes hypothesized the pineal gland, near the center of the brain, as the seat of causal interaction. The objection was raised that the mental and physical planes were of such a fundamentally dissimilar character as to preclude any ordinary notion of causal interaction. But the part about a separate yet joint reality of mind and body seemed less objectionable, and various commentators offered their own explanations for the undeniably strong correlations of mental and physical phenomena.

Malebranche insisted that these were only correlations and not true interactions, whose appearance of interaction was arranged in every detail by God by divine intervention on every occasion of correlation, a theory that naturally enough came to be called occasionalism. Spinoza freed God from this demanding schedule by organizing the parallel behavior of mind and matter as a preordained apartheid emanating from God as the source of everything. Leibniz postulated *monads*, cosmic chronometers miraculously keeping perfect time with each other yet not interacting.

These patently untestable answers only served to give dualism a bad name, and it gave way in due course to one or another form of monism: either mind *or* matter but not both as distinct real substances. Berkeley opined that matter did not exist and that the universe consisted solely of ideas. Hobbes ventured the opposite: mind did not exist except as an artifact of matter. Russell [Rus27] embraced *neutral* monism, which reconciled Berkeley's and Hobbes' viewpoints as compatible dual accounts of a common neutral Leibnizian monad.

This much of the history of mind-body dualism will suffice as a convenient point of reference for the sequel. R. Watson's Britannica article [Wat86] is a conveniently accessible starting point for further reading.

The thesis of this paper is that mind-body dualism can be made to work via a theory that we greatly prefer to its monist competitors. Reflecting an era of reduced expectations for the superiority of humans, we have implemented causal interaction not with the pineal gland but with machinery freely available to all classical entities, whether newt, pet rock, electron, or theorem (but not quantum mechanical wavefunction, which is sibling to if not an actual instance of our machinery).

2 Dualism via Chu Spaces

We propose to reduce complex mind-body interaction to the elementary interactions of their constituents. Events of the body interact with states of the mind. This interaction has two dual forms. A physical event a in the body A *impresses* its occurrence on a mental state x of the mind X, written $a \Dashv x$. Dually, in state x the mind *infers* the prior occurrence of event a, written $x \models a$. States may be understood as corresponding more or less to the possible worlds of a Kripke structure, and events to propositions that may or may not hold in different worlds of that structure.

With regard to orientation, impression is *causal* and its direction is that of time. Inference is *logical*, and logic swims upstream against time. Prolog's backward-chaining strategy dualizes this by viewing logic as primary and time as swimming upstream against logic, but this amounts to the same thing. The basic idea is that time and logic flow in opposite directions.

Can a body meet a body? Only indirectly. All direct interaction in our account of Cartesian dualism is between mind and body. Any hypothesized interaction of two events is an inference from respective interactions between each of those events and all possible states of the mind. Dually, any claimed interaction of two states is inferred from their respective interactions with all possible events of the body.

The general nature of these inferences depends on the set K of values that events can impress on states. The simplest nontrivial case is $K = 2 = \{0,1\}$, permitting the simple recording of respectively nonoccurrence or occurrence of a given event in a given state. In this case body-body and mind-mind interactions are computed via a process called residuation. Specifically, event a *necessarily precedes* event b when every state x witnessing the occurrence of b also witnesses a. This inferred relationship is calculated formally by left residuation, which we describe in detail later. The dual calculation, right residuation, *permits* a transition from state x to state y when every event a impressing itself on x does so also on y. That is, any transition is permitted just so long as it forgets no event. These simple-minded criteria are the appropriate ones for the small set $K = 2$.

For $K = 3$ more complex rules for inferring necessary precedence and possible transition obtain, including the possibility of forgetting (to be written up). At $K = 8$ we have groups and semigroups, the latter embedding all *abstract* category theory [PT80]. For K the set (not field) of complex numbers, right and left residuation are naturally taken to be the respective products $\langle\varphi|\psi\rangle$ or $\varphi^*\psi$ and $|\psi\rangle\langle\varphi|$ or $\psi\varphi^*$, corresponding to respectively inner product and its dual outer product in a Hilbert space.

This conveys the flavor of our proposal. We now equip these general ideas with enough algebraic structure and properties to make the proposal interesting, useful, and we hope convincing.

The following analogy serves to fix ideas. The numbers ±1 are connected in two ways, algebraic and geometric. The algebraic connection is via the operation of negation, an involution ($- - x = x$) that connects them logically by interchanging them. The geometric connection is via the interval $[-1,1]$ of reals lying between these numbers, a closed convex space connecting them topologically. We refer to these connections as respectively the *duality* and *interaction* of -1 and 1. The connections themselves might respectively be understood as mental and physical, but this takes us beyond our present story.

We regard each point of the interval as a weighted sum of the endpoints, assuming nonnegative weights p, q normalized via $p + q = 1$, making each point the quantity $p - q$. An important property of interaction is that it includes the endpoints, namely as the special cases where one of p or q is zero. An important property of duality is that it extends to interaction, namely via the calculation $q - p = -(p - q)$.

We shall arrange for Cartesian dualism to enjoy the same two basic connections and the two associated properties, with mind and body in place of -1 and 1 respectively. Ideally the duality would be a negation-like involution that interchanges their roles; no information is lost in this transformation, and the original mind or body is recovered when the transformation is repeated. And ideally the interaction would turn out to be the long-sought solution to dualism's main conceptual hurdle. Chu spaces achieve both of these in a very satisfactory way.

The counterparts to ± 1 in our Chu space formulation of Cartesian dualism are the respective categories **Set** and **Set**^{op}. That is, at 1 we place the class of all sets, each understood as a pure body. At -1 we place what would appear at first sight to be the same sets, which we propose to construe as pure minds.

Our first distinction between body and mind will be the trivial one of using different variables to range over these sets: A, B over bodies, X, Y over minds. The second distinction will be in how the two kinds of sets transform into each other. Later we make a third distinction within the objects themselves by realizing the two kinds as Chu spaces with dual form factors: sets tall and thin, antisets short and wide.

Bodies transform with functions. We turn the class of bodies into **Set** by first superimposing on it the graph whose edges comprise all functions, with each function $f : A \to B$ connecting the set A to the set B. We then promote this graph to a category by equipping it with the standard composition rule for functions, as an instance of composition of binary relations, along with an identity function $1_A : A \to A$ at every set A.

Minds transform with *antifunctions*. An antifunction $g^{\perp} : X \to Y$ is a binary relation from X to Y whose converse is a function $g : Y \to X$. Adopting the composition rule for binary relations as with **Set** then yields a category dual to **Set**, one that is equivalent, in fact isomorphic, to **Set**^{op} (the result of merely reversing all the edges of **Set**), which we simply identify with **Set**^{op}.

These graphs are not isomorphic, even without their respective compositions. A quick way to tell them apart is to look for a vertex whose only edge to it is a self-loop. This vertex occurs only in **Set**, namely as the empty set. Or look for a vertex whose only edge *from* it is a self-loop; this too is the empty set, but in **Set**^{op}. The reader will think of other tests.[1]

We now argue that sets are physical and antisets mental. Since the only difference is in how they transform, any distinction between mental and physical must be either dynamic in the sense of being transformational, or algebraic in the sense that structure regulates transformation. We present both types of argument (which themselves can be understood as respectively operational hence mental and denotational hence physical).

Functions identify and adjoin. The function $F : A \to B$ identifies just when it fails to be injective: $f(a) = f(b)$ means that f identifies a and b. It adjoins just when it fails to be surjective: $f : A \to B$ first transforms A *onto* $f(A)$, then adjoins to it $B - f(A)$ to become *into*.

Antifunctions copy and delete. The antifunction $g^{\perp} : X \to Y$ makes multiple copies just when its converse $g : Y \to X$ fails to be injective: $g(y) = g(y')$ means that $g^{\perp}$ sends copies of $g(y)$ to both y and y', *inter alia*. It deletes just when g fails to be surjective: $g^{\perp} : Y \to X$ deletes exactly $Y - g(X)$.

Identifying and adjoining are canonically denotational tasks that mathematicians are accustomed to performing on their spaces, groups, and other algebraic objects. This is the realm of the physical.

Copying and deleting are canonically operational tasks that logicians and computer scientists are accustomed to performing on their proofs, spreadsheets, and other symbolic objects. This is

---

[1] Example: look for any vertex having exactly one edge to it from each vertex, and infinitely many edges out. There are lots of these in **Set**, namely the many singletons, all isomorphic, but none in **Set**^{op}.

the realm of the mental.

In additional to these transformational arguments we can contrast the discrete or dust-like physical structure of sets with the rigidly intermeshed mental structure of Boolean algebras.

A set is an algebra with no language at all, and no equational theory beyond the equational tautologies $x = x$. There is therefore no mental plane to speak of in sets, making them the most physical of all the objects of traditional concrete (set-based) mathematics, if not of all category theory (and perhaps even there, cf. [RW94]).

$\mathbf{Set}^{\mathrm{op}}$ is equivalent to the category of complete atomic Boolean algebras (CABA's). But the free CABA generated by the set X is the power set 2^{2^X}. Hence the Boolean operations of each arity X, X empty, finite, or infinite, consist of *all* functions from 2^X to 2. This is the maximum possible language compatible with CABA homomorphisms; not only is every arity represented but every operation of that arity.[2] Furthermore the equational theory of CABA's is maximally consistent in the sense that no new equation can be added without collapsing the entire algebra to a singleton. A CABA as the ultimate know-it-all is as mental as any object of traditional concrete mathematics can be.

We have thus established that the two isolated points **Set** and $\mathbf{Set}^{\mathrm{op}}$ represent respectively the physical and the mental. We now proceed with the promised construction. At this point the situation is as for ± 1 on their own: we have two isolated graphs, and we seek a duality and an interaction.

The duality analogous to negation is simply the converse operation for binary relations, which evidently interchanges **Set** and $\mathbf{Set}^{\mathrm{op}}$.

The interaction analogous to the interval $[-1, 1]$, which includes the points it connects as part of the interval, consists of all Chu spaces and a graph superimposed on them, which includes as subgraphs **Set** and $\mathbf{Set}^{\mathrm{op}}$. That is, the interaction consists of adding further vertices and edges, in addition to those already present, to populate an interval from $\mathbf{Set}^{\mathrm{op}}$ to **Set**.

A *Chu space* $\mathcal{A} = (A, X, \models)$ *over* a set K consists of a set A of *points*, an antiset X of *states*, and an $X \times A$ matrix $\models$ with entries drawn from K.[3] These provide the vertices of the interval.

This ontogeny of the Chu space recapitulates the phylogeny we are working towards. A and X are respectively the body or object and mind or menu of the space, $\models$ is their interaction, and matrix transposition is the duality interchanging mind and body to yield the dual Chu space $\mathcal{A}^{\perp} = (X, A, \models^{\smile})$.

Points have *necessary* existence, all being present simultaneously in the physical object A. States are *possible*, making a Chu space a kind of a Kripke structure [Gup93]: only one state at a time may be chosen from the menu X of alternatives.

Lafont and Streicher [LS91] were the first to single out Chu spaces as a case of the more general Chu construction $\mathbf{Chu}(V, k)$ [Bar79, Bar91], namely $V = \mathbf{Set}$, worthy of separate attention as a natural model of linear logic [Gir87] embedding topological spaces, vector spaces, and coherent spaces. They referred to these objects as games, understanding $\models$ as the payoff matrix of a von-Neumann-Morgenstern two-person game.

There is a chicken-and-egg question here as to whether Chu spaces are more naturally understood as a game or a player of a game. As players, the spaces $\mathcal{A}$ and $\mathcal{B}$ play the *interaction game* $\mathcal{A} \otimes \mathcal{B}$, their tensor product. This interaction has featured prominently in our own research as an operation we called *orthocurrence* [Pra85, Pra86]. We originally identified orthocurrence as ordinary

---

[2] One can add further operations, for example modal logic adds $\Diamond$. However CABA homomorphisms respect none of these additional operations whatsoever.

[3] Contrast this with a vector space over a field k, which requires k to be equipped with the four rationals; here K is simply a set with no additional structure.

product in a cartesian closed category of partially ordered multisets (pomsets), but subsequently generalized it to the tensor product of any closed category [CCMP91, Pra93, GP93, Pra94a]. In all cases we took as our basic example the interaction of trains and stations described on the train station wall by the daily schedule. Whereas ordinary product must be capable of being projected *consistently* onto either component, tensor product requires only that each row or column of the resulting rectangular body of the space (how stations appear to conductors, and trains to stationmasters) meet all the constraints imposed on each of the two constituents of the product, the concept of bilinearity. The tensor product constitutes a larger Chu space, which can in turn be a player in a yet larger game.

The representation $\mathcal{A}\otimes\mathcal{B}$ takes the physical viewpoint. The logic of the game may be understood in terms of its dual $(\mathcal{A}\otimes\mathcal{B})^{\perp}$, which is equivalent to either of $\mathcal{A}\multimap\mathcal{B}^{\perp}$ or $\mathcal{B}\multimap\mathcal{A}^{\perp}$. In the former, we take Alice's point of view as our premises and view Bob as the goal. This view dualizes Bob to make his body, which Bob proudly thinks of as his strong points, appear to Alice as Bob's possible Achilles' heels (wrists, etc.). At the same time Bob's mind, which Bob thinks of as his possible options, are seen by Alice as Bob's tricks, all of which she must be simultaneously on her guard against.

A *Chu transform* $(f,g) : (A,X,\models) \rightarrow (A',X',\models')$ consists of a function $f : A \rightarrow A'$ and an antifunction $g^{\perp} : X \rightarrow X'$, namely the converse of a function $g : X' \rightarrow X$, satisfying the *continuity condition* $g(x') \models a \;=\; x' \models' f(a)$ for all $a \in A$ and $x' \in X'$. These provide the edges of the graph on the interval of all Chu spaces running from $\mathbf{Set}^{\mathrm{op}}$ to $\mathbf{Set}$. They compose via $(f',g')(f,g) = (f'f, gg')$ to make the graph a category, denoted $\mathbf{Chu}_K$.

The function f transforms the body of the space denotationally, identifying some points and adjoining others, but neither deleting nor duplicating any. At the same time the antifunction g transforms the mind of the space operationally, i.e. as a symbolic object such as a program or a proof, deleting some states to further constrain the degrees of freedom of the space and copying some as needed so as not to infringe on the degrees of freedom of the newly adjoined points (transformations need only preserve the structure of what they transform and cannot be held responsible for what goes on in the adjoined points). However g never identifies states, which would be logically inconsistent for states having distinct rows, and never adjoins states having new rows, which would be logically unsound (the image could enter a state not permitted its source).

To understand better this last point, let $\mathrm{row} : X \rightarrow (A \rightarrow K)$ and dually $\mathrm{col} : A \rightarrow (X \rightarrow K)$ denote the functions satisfying $\mathrm{row}(x)(a) \;=\; x{\models}a \;=\; \mathrm{col}(a)(x)$. Continuity may then be rephrased in terms of rows: $\mathrm{row}(g(x')) \;=\; \mathrm{row}'(x')\circ f$, verified via $\mathrm{row}(g(x'))(a) \;=\; g(x'){\models}a \;=\; x'{\models'}f(a) \;=\; (\mathrm{row}'(x')\circ f)(a)$. That is, every row of B when composed with f must be some row of A, with g a function selecting a suitable row index. When $K = 2$ this is equivalent to requiring that g behave as f^{-1} on rows viewed as characteristic functions of subsets of A'. But then the requirement that every row of $\mathcal{A}'$ be mapped by f^{-1} to some row of $\mathcal{A}$ is recognizable as the condition for a function between topological spaces to be continuous, where rows are understood as open sets.

For technical reasons Chu transforms are usually associated with a fixed K, calling for a distinct category $\mathbf{Chu}_K$ of Chu spaces for each set K. A set theorist should have no difficulty with Chu spaces over different K's transforming into each other, but the resulting category would to begin with lack a tensor unit, an annoying omission when one begins to press the rich algebraic structure of $\mathbf{Chu}_K$ into service.

The structure of $\mathbf{Chu}_K$ is that of linear logic [Gir87], which can be understood as the logic of four key structural properties of $\mathbf{Chu}_K$: it is concrete, complete, closed, and self-dual (which therefore makes it also cocomplete and coconcrete). The associated linear logic connectives are respectively $!A$, $A \oplus B$ (and unit 0), $A\multimap B$ (and left unit 1), and $A^{\perp}$, which form a complete basis for linear logic. $\mathbf{Chu}_K$ is complete but perhaps for syntactic simplicity linear logic weakens completeness to finite products. Furthermore it is not yet agreed whether induction is a necessary element of concreteness.

Just as $\{-1,1\} \subseteq [-1,1]$, so are sets and antisets made part of the category of Chu spaces, as follows. The set A is identified with the Chu space $\mathcal{A} = (A, K^A, \gamma)$ where for each $x : A \to K$, $\gamma(x,a)$ denotes the application $x(a)$. The function $f : A \to A'$ is identified with the pair $(f, f^\perp) : (A, K^A, \gamma) \to (A', K^{A'}, \gamma)$ where $f^\perp : K^{A'} \to K^A$ is defined by $f^\perp(g)(a) = g(f(a))$. When $K = 2$, $f^\perp$ can be seen to be the usual inverse-image function f^{-1}, making this topology's continuity condition as remarked earlier. We call the Chu space $\mathcal{A}$ a *realization*[4] of the set A in $\mathbf{Chu}_2$.

Dually the antiset X is identified with $(K^X, X, \gamma^\smallsmile)$ where $\gamma^\smallsmile$ is converse application, satisfying $\gamma^\smallsmile(x,a) = a(x)$, and the antifunction $g^\perp : X \to X'$ (i.e. the function $g : X' \to X$) is identified with the pair (f,g) where $f : K^X \to K^{X'}$ is defined at each $h : X \to K$ by $f(h)(x') = h(g(x'))$. This constitutes a realization of $\mathbf{Set}^{\mathbf{op}}$ in $\mathbf{Chu}_2$.

Just as the duality of ±1 extended to $[-1,1]$, so does the mind-body duality of $\mathbf{Set}$ and $\mathbf{Set}^{\mathbf{op}}$ extend to $\mathbf{Chu}_K$. The dual of $\mathcal{A} = (A, X, \models)$ is $\mathcal{A}^\perp = (X, A, =\!\mid)$, while the dual of the Chu transform (f,g) is (g,f). Moreover the duality of sets and antisets achieved via converse of their transforming binary relations is also achieved via Chu duality for their realizations in $\mathbf{Chu}_K$.

To each finite Chu space $\mathcal{A}$ we associate integers P and Q measuring respectively the *discipline* and *versatility* of $\mathcal{A}$, in terms of the amount by which the space fails to be a set or an antiset respectively. Write $\|A\|$ for the number of *distinct* columns of the matrix, and likewise $\|X\|$ for the number of distinct rows. Let $P = K^{\|A\|} - \|X\|$ and $Q = K^{\|X\|} - \|A\|$, both nonnegative. For $K \geq 2$ these cannot vanish simultaneously or we would have an integer solution to $K^{K^A} = A$. Hence we can safely define nonnegative reals $p = P/(P+Q)$, $q = Q/(P+Q)$ satisfying $p + q = 1$. We take $p - q$ as the location of $\mathcal{A}$ in the interval $[-1,1]$ itself, giving a sense in which $\mathbf{Chu}_K$ lies between $\mathbf{Set}^{\mathbf{op}}$ and $\mathbf{Set}$. Notice that this procedure assigns sets and antisets to 1 and -1 respectively, while exactly square Chu spaces are sent to 0.

Although the position of a Chu space in $[-1,1]$ gives some indication of its form factor, these positions turn out not to populate $[-1,1]$ densely. For example at $K = 2$ the intervals $[\frac{1}{3}, \frac{1}{2})$ and $(\frac{1}{2}, \frac{2}{3}]$ contain no Chu spaces, since Chu spaces that are only one away from being square are below $\frac{1}{3}$ or above $\frac{2}{3}$, and indeed the interval is riddled with such holes. One imagines being able to distribute Chu spaces more uniformly along $[-1,1]$ with the help of say $\|A\|/\|X\|$, but in choosing such a formula it would help to have some reason for wanting a dense distribution.

This viewpoint is a compromise between those of set theory and category theory. Set theory monistically constructs everything from the single category of pure sets. Category theory pluralistically constructs a plethora of categories. Chu spaces are like sets in that there is only one category $\mathbf{Chu}_K$ of them (modulo the parameter K). $\mathbf{Chu}_K$ is dualistic in that it postulates the two categories $\mathbf{Set}$ and $\mathbf{Set}^{\mathbf{op}}$, neither of which is singled out as having priority over the other, and connects them via interaction to form the single much larger category $\mathbf{Chu}_K$. Some impression of its size may be had from the theorem [Pra93, p.153-4] that $\mathbf{Chu}_{2^k}$ realizes the category of *all* $k-ary$ relational structures and their homomorphisms standardly defined. For example Chu_8 realizes the category of ternary relational structures, which in turn realizes the category of groups and group homomorphisms (since its multiplication is the ternary relation $xy = z$), and realization is transitive.

---

[4] A *representation* is a full embedding of one category in another, i.e. a full and faithful functor $F : C \to D$. A *realization* is a concrete representation; that is, C and D are concrete categories, meaning they have underlying set functors $U_C : C \to \mathbf{Set}$ and $U_D : D \to \mathbf{Set}$, with which F commutes, $U_D F = U_C$, i.e. the realizing object has the same underlying set as the object it realizes [PT80, p.49].

3 The Meaning of Interaction

Thus far we have constructed interaction as no more than a formal notion. We now relate it to our intuitions about causal interaction.

It is ironic that Cartesian philosophy, whose guiding dictum was to question everything, should question causal interaction *between* the mental and physical planes before that *within* the planes. The latter problems must have posed an insufficient challenge to the Cartesians. We argue that the converse is the case: between is actually easier than within!

We interpret interaction as causality. Causality is directional, but the direction depends on whether we have in mind physical or mental causality. We interpret $x \models a$ ambiguously as the time elapsed between the occurrence of the physical a and its impression on the mental state x, and as the truth value of a as a proposition.[5] The former is physical causality or *impression*, flowing forward in time from events to states. The latter is mental causality or *inference*, flowing backwards in time from the thought of a to the inference of a's occurrence. In this way time flows forward (from the usual point of view) while logic flows backward. This is *primary* interaction, and it occurs only *between* the mental and physical plane.

We thus see that the seat of causal interaction in Cartesian duality is not the pineal gland but the identification of impression and inference. We write $x \models a$ as expressing equally the impression of event a on subsequent state x and the deduction by state x of the prior occurrence of event a. The Cartesian dictum *cogito, ergo sum* is the case of this where x is the thinker's state and a the event of his or her existence.

As a proponent of more dynamic logics than traditionally contemplated in logic [Pra76, Pra90a] we point out the atemporal quality of this dictum, a hallmark of classical logic. Examined closely, our analysis shows that Descartes' dictum properly tensed becomes *cogito, ergo eram* (I was), an epitaph both of whose tenses the liar paradox renders true in perpetuity. Our thoughts follow from our events but not conversely and hence may survive them without logical contradiction. A particularly good one may far outlive its source.

We pass now to interaction *within* each plane, whether physical or mental, which we derive as *secondary* interaction from the primary form with the aid of *residuation*, a pair of operations on binary relations that constitutes dynamic implications forwards and backwards in time;. For $K = 2$, $=\!\mid$ as a matrix of 0's and 1's is an ordinary binary relation: the event a either is or is not related to state x. This relation is understood ambiguously as a two-valued distance in either time space ($a =\!\mid x$, physical) or information space ($x \models a$, mental).

Given any two contrary binary relations $R \subseteq U \times V$, $T \subseteq U \times W$, their *right residual* $R\backslash T$ [WD39, Jón82, Pra90b] can be defined equivalently as follows.

(i) As the operation satisfying $R;S \subseteq T$ iff $S \subseteq R\backslash T$. (Think of this as defining division on the left by R, with inequalities where one would expect an equality. The case $R = 0$, all entries 0, requires no special attention.)

(ii) As the largest relation $S \subseteq V \times W$ such that $R;S \subseteq T$.

(iii) As the set of all pairs (v, w) in $V \times W$ such that $uRv \to uTw$ for all $u \in U$.

(iv) As that operation monotone in its right hand argument that satisfies modus ponens, $R;(R\backslash T) \vdash T$, and also $T \vdash R\backslash(R;T)$, where $\vdash$ is read as $\subseteq$. This makes $R;-$ and $R\backslash-$ pseudoinverse operations which when composed either decrease or increase their argument depending on

---

[5] The reader may be understandably concerned at this identification of physical events and ostensibly mental propositions. However a Boolean proposition about events in A is of type 2^{2^A} and *each exponentiation* dualizes, whence two of them return us to the physical plane. The truly mental propositions are the constituent descriptive clauses of a physical DNF formula, each describing a possible world.

the order of composition.

(v) As the relation $(R^{\smile};T^{-})^{-}$ where $R^{\smile}$ is converse (transpose) and T^{-} is complement (change all 0's to 1's in the matrix and vice versa). This can be written more neatly as $(T\dagger;R)\dagger$ where $T\dagger$ denotes $T^{-\smile}$. If we think of residuation $R\backslash T$ as a form of implication $R \rightarrow T$, and composition as a form of conjunction, and allow for the noncommutativity of relational composition (relative product), then this corresponds to the classical principle $A \rightarrow B \;\equiv\; \neg(A \wedge \neg B)$, as well as to linear logic's $A \multimap B \;\equiv\; (A \otimes B^{\perp})^{\perp}$.

It is a straightforward exercise to show the equivalence of these definitions; see [Pra90a] for further discussion.

Definition (v) reveals the contravariance of the operation in R, and its covariance in T, composition being monotone in each argument, a form of bilinearity. We therefore call residuation *sesquilinear*, in anticipation of the next section.

Now consider $\Dashv\backslash\Dashv$ in the light of condition (iii). This instance of residuation is a binary relation on X. For all x, y in X, $x(\Dashv\backslash\Dashv)y$ holds just when row x implies (is a subset of) row y for every event, i.e. when $x \rightarrow y$ is valid. Now $x \rightarrow y$ says that in order to be able to get from x to y, every event a whose occurrence is recorded in x must still be recorded in y. Thus $\Dashv\backslash\Dashv$ consists of those pairs (x, y) which as transitions do not entail taking back the claim that an event has already happened.

This makes $\Dashv\backslash\Dashv$ the natural transition relation on X. This is a partially ordered *automaton*. Elsewhere we have used higher dimensional automata to argue that automata could be reliably paired up as the dual of schedules [Pra92]. We find Chu spaces a very appealing extension of this duality.

The left residual T/S, where $T \subseteq U \times W$, $S \subseteq V \times W$, is the dual of the right. We settle for defining T/S as the set of all pairs (u, v) in $U \times V$ such that $vSw \rightarrow uTw$ for all $w \in W$ (cf. (iii)), and ask the reader to infer the other four equivalent formulations corresponding to (i)-(v) above.

The left residual $\Dashv/\Dashv$ is, by dual reasoning to $\Dashv\backslash\Dashv$, that binary relation on A containing (a, b) just when for all $x \in X$, $b \Dashv x$ implies $a \Dashv x$. This makes it the natural temporal precedence relation on events, namely a *schedule* of events, an alternative to automata theory and Kripke structures that has attracted our attention as a reliable model of true concurrency since 1982 [Pra82].

When we unravel the primitive causal links contributing to secondary causal interaction we find that two events, or two states, communicate with each other by interrogating *all* entities of the opposite type. Thus event a deduces that it precedes event b not by broaching the matter with b directly, but instead by consulting the record of every state to see if there is any state volunteering a counterexample. When none is found, the precedence is established. Conversely when a Chu space is in state x and desires to pass to state y, it inquires as to whether this would undo any event that has already occurred. If not then the transition is allowed.

If one truly believed that the universe proceeded via state transitions, this might seem a roundabout and inefficient way of implementing those transitions. However it seems to us, particularly in view of the considerations of the following section, that the more likely possibility is that the universe only *seems* to proceed via state transitions, due perhaps to our ancestors having ill-advisedly chosen monism as the natural world view, perhaps millennia before the rise of Cartesianism, perhaps only some years after its decline. What we conjecture actually happens is that events signal states forward in time, or equivalently that states infer events backwards in time, and the world we imagine we live in is simply what that process looks like to its inhabitants when interpreted monistically.

Why this theory as opposed to any other? Well, certainly no other theory has satisfactorily explained the causal interaction of real mental and physical planes as conceived by Descartes.

Whether monism is an equally satisfactory alternative for Descartes' problem is a good question. But for the other applications of Chu spaces considered here, namely concurrency, metamathematics, quantum mechanics, and logic (see below), it seems to us that monism simply cannot compete with dualism.

4 Quantum Mechanics

When time and truth are complex-valued as in quantum mechanics, right residuation is replaced by the sesquilinear operation of inner product $\langle\varphi|\psi\rangle$. This is a complex-valued *correlation* between wavefunctions $\langle\varphi|$ and $|\psi\rangle$, which are given as points of a *Hilbert space*, a metrically complete vector space which is made an inner product space with this operation.

The correspondence with Chu spaces is as follows. Any given choice of basis of Hilbert space defines a set of propositions, one per basis vector. Each coordinate of a given state vector relative to that basis is interpreted as the complex truth value of the corresponding proposition in that state. Relative to that basis, a state vector then corresponds to a row of $\models$, or a column of $=\!\!|$. Right residuation is defined even for one-state spaces, and is in form the logical counterpart to inner product. The right residual of a one-state space with itself is simply the identity relation on that state, this being the only partial order possible. The inner product of a wavefunction with itself is a scalar, namely its length squared, but quantum mechanics is a projective system where lengths are only physically meaningful in proportion: the length of a single state is no more informative in QM than is the identity partial order on a singleton.

A *mixed state* is a set of pure states and a distribution giving their relative probabilities. Such a distribution can be understood as a quantitative form of disjunction, making a mixed state the quantum mechanical counterpart of a Chu space. Here $\langle\varphi|\psi\rangle$ for mixed states corresponds to the right residual of two Chu spaces. The inner product of a mixed state with itself yields a square matrix of transition probabilities between its constituent pure states. The right residual of a Chu space with itself yields a square matrix of transition *possibilities* when $K = 2$, and a suitably richer relation for larger K, where the possibilities begin to depend on choice of quantale for K, taking us beyond the scope of this paper.

The outer product $|\psi\rangle\langle\varphi|$ produces an operator which transforms Hilbert space. Viewed as a transformation of basis vectors of Hilbert space, such an operator establishes correlations between attributes. The corresponding operation on Chu spaces is left residuation, which likewise produces a (two-valued) correlation between events, which we may identify with attributes.

This perspective leads to the following reconstruction of the emergence of modern quantum mechanics in 1925-26. Classical physics, and the old quantum mechanics, took between-state correlations as basic. Newton's laws, or their expression in terms of Lagrange's equations and the energy-difference Langrangian, were couched in terms of space and time, with velocity v being the derivative of position with respect to time, and momentum being mv. Hamilton made the bold move of taking momentum to be an independent quantity in its own right, observing that two equations per dimension based on a total-energy Hamiltonian yielded an elegantly symmetric reformulation of Langrange's one equation per dimension. From the perspective of classical physics this was no more than an ingeniously symmetric but otherwise unimproved variant of the basic laws of motion.

The new quantum mechanics made Hamilton's "causal interaction" of momentum and position primitive, and derived the classical laws as secondary. Furthermore they used the same logic, only as a complex-valued fuzzy logic rather than a two-valued logic, to achieve this end. This made momentum-space interaction a simple interaction, and the derived momentum-momentum and space-space interactions more complex. These can be understood as having to go both backwards and forwards in time for their complete effect, the basis for Cramer's transactional account of quantum mechanics [Cra86], which Leslie Lamport drew to my attention in 1987.

5 Conclusion

We have advanced a mechanism for the causal interaction of mind and body, and argued that separate additional mechanisms for body-body and mind-mind interaction can be dispensed with; mind-body interaction is all that is needed. This is a very different outcome from that contemplated by 17th century Cartesianists, who took body-body and mind-mind interaction as given and who could find no satisfactory passage from these to mind-body interaction. Even had they found a technically plausible solution to their puzzle, mind-body interaction would presumably still have been regarded as secondary to body-body interaction. We have reversed that priority.

One might not expect mind-body duality as a mere philosophical problem to address any urgent need outside of philosophy. Nevertheless we have offered solutions to the following practical problems that could be construed as particular applications of our general solution to Descartes' mind-body problem, broadly construed to allow scarecrows and everything else to have minds.

What is the conceptual basis of concurrent computation? What is the essence of quantum mechanics? On what foundation should mathematics be based? What is the right logic to reason with?

Concepts for concurrent computation. Our research focus since 1980 has been concurrent computation. Our conclusion is that programmers should be able to move as freely as possible between declarative and imperative modes of thought about the same program. We are now convinced that the duality of schedules and automata, as the realization of the duality of body and mind respectively in the world of programming, provides a better conceptual foundation for concurrent programming than any other model.

Essence of quantum mechanics. We claim that quantum mechanics has not previously been reduced to lay terms by physicists, who have been content to leave the subject as a mysterious jumble of properties of Hilbert space that the working physicist can become acclimatized to and even confident with after sufficient exposure. Mind-body duality and interaction explains respectively complementarity and the inner product in relatively elementary terms making a clear connection with other structures such as the above model of computation and the following foundation for mathematics. The central role of the mental plane in this account of quantum mechanics makes it a *rational* mechanics.

Foundations of mathematics. We implicitly settle for relational structures as the objects of mathematics when we so restrict the models of first-order logic. But this has the unfortunate side effect of excluding some popular mathematical structures, most notably topology, which would appear to require a second order theory. Chu spaces over 2^k realize all k-ary relational structures [Pra93, p.154-3] as well as topological spaces when $K = 2$ [LS91], all as objects of the one category, yielding a novel degree of morphism-sensitive typelessness for foundations. The above connection with quantum mechanics suggests that mathematics based on Chu spaces be thought of as *natural* mathematics, sharing with nature the essential principles of duality and interaction.

Choice of logic. We envision two logics, elementary and transformational. Elementary logic has its usual meaning as the logic of individual objects such as sets, groups, and Boolean algebras. It serves to reason about relationships between elements of such objects. These objects are traditionally understood as relational structures but they can also more generally be understood as Chu spaces as per the preceding paragraph.

Transformational logic bears superficial resemblances to elementary logic but serves to reason about interactions between objects rather than relationships within objects. The structural basis for object interaction is the homomorphism or structure-preserving morphism, from which flows all other interaction structure such as duality, limits, tensor products, homsets, and size (cardinality or concreteness).

The most promising transformational logic seems to us to be Girard's linear logic [Gir87]. Chu_K is a constructive model of linear logic in the sense that it interprets the sequents of linear logic as sets of proofs rather than as Boolean or intuitionistic truth values. Nonconstructive models of linear logic such as phase spaces seem to us at best a curiosity. As to alternative constructive models, for want of any convincing counterexamples we conjecture mildly that these can all be satisfactorily subsumed by Chu spaces, the case $V =$ Set of the general Chu construction $\mathrm{Chu}(V, k)$. We have yet to be shown a V that improves on Set for any significant application of the Chu construction.

References

[Bar79] M. Barr. **-Autonomous categories, LNM 752.* Springer-Verlag, 1979.

[Bar91] M. Barr. *-Autonomous categories and linear logic. *Math Structures in Comp. Sci.*, 1(2), 1991.

[CCMP91] R.T Casley, R.F. Crew, J. Meseguer, and V.R. Pratt. Temporal structures. *Math. Structures in Comp. Sci.*, 1(2):179–213, July 1991.

[Cra86] J.G. Cramer. The transactional interpretation of quantum mechanics. *Reviews of Modern Physics*, pages 647–687, 1986.

[Gir87] J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50:1–102, 1987.

[GP93] V. Gupta and V.R. Pratt. Gates accept concurrent behavior. In *Proc. 34th Ann. IEEE Symp. on Foundations of Comp. Sci.*, pages 62–71, November 1993.

[Gup93] V. Gupta. Concurrent Kripke structures. In *Proceedings of the North American Process Algebra Workshop, Cornell CS-TR-93-1369*, August 1993.

[Jón82] B. Jónsson. Varieties of relation algebras. *Algebra Universalis*, 15:273–298, 1982.

[LS91] Y. Lafont and T. Streicher. Games semantics for linear logic. In *Proc. 6th Annual IEEE Symp. on Logic in Computer Science*, pages 43–49, Amsterdam, July 1991.

[Pra76] V.R. Pratt. Semantical considerations on Floyd-Hoare logic. In *Proc. 17th Ann. IEEE Symp. on Foundations of Comp. Sci.*, pages 109–121, October 1976.

[Pra82] V.R. Pratt. On the composition of processes. In *Proceedings of the Ninth Annual ACM Symposium on Principles of Programming Languages*, January 1982.

[Pra85] V.R. Pratt. Some constructions for order-theoretic models of concurrency. In *Proc. Conf. on Logics of Programs, LNCS 193*, pages 269–283, Brooklyn, 1985. Springer-Verlag.

[Pra86] V.R. Pratt. Modeling concurrency with partial orders. *Int. J. of Parallel Programming*, 15(1):33–71, February 1986.

[Pra90a] V.R. Pratt. Action logic and pure induction. In J. van Eijck, editor, *Logics in AI: European Workshop JELIA '90, LNCS 478*, pages 97–120, Amsterdam, NL, September 1990. Springer-Verlag.

[Pra90b] V.R. Pratt. Dynamic algebras as a well-behaved fragment of relation algebras. In *Algebraic Logic and Universal Algebra in Computer Science, LNCS 425*, pages 77–110, Ames, Iowa, June 1988, 1990. Springer-Verlag.

[Pra92] V.R. Pratt. Event spaces and their linear logic. In *AMAST'91: Algebraic Methodology and Software Technology*, Workshops in Computing, pages 1–23, Iowa City, 1992. Springer-Verlag.

[Pra93] V.R. Pratt. The second calculus of binary relations. In *Proceedings of MFCS'93*, pages 142–155, Gdańsk, Poland, 1993. Springer-Verlag.

[Pra94a] V. Pratt. Chu spaces: complementarity and uncertainty in rational mechanics. Technical report, TEMPUS Summer School, Budapest, July 1994. Manuscript available as pub/bud.tex.Z by anonymous FTP from Boole.Stanford.EDU.

[Pra94b] V.R. Pratt. Chu spaces: Automata with quantum aspects. In *Proc. Workshop on Physics and Computation (PhysComp'94)*, Dallas, 1994. IEEE.

[PT80] A. Pultr and V. Trnková. *Combinatorial, Algebraic and Topological Representations of Groups, Semigroups, and Categories*. North-Holland, 1980.

[Rus27] B. Russell. *The analysis of matter*. Paul, Trench, Trubner (also Dover 1955), 1927.

[RW94] R. Rosebrugh and R.J. Wood. An adjoint characterization of the category of sets. *Proc. Amer. Math. Soc.*, 122(2):409–413, 1994.

[Wat86] R.A. Watson. Descartes and Cartesianism. In *The New Encylopaedia Britannica*. Encyclopaedia Britannica, Inc., 1986.

[WD39] M. Ward and R.P. Dilworth. Residuated lattices. *Trans. AMS*, 45:335–354, 1939.

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.

Executive Council: Dr. John Amson, Dr. Ted Bastin, Mr. Anthony M. Deakin, Dr. Tom Etter, Ms. Arleta Griffor, Prof. Louis Kauffman, Prof. Clive W. Kilmister, Dr. Michael Manthey, Prof. H. Pierre Noyes, Dr. David Roscoe, Dr. Fredric S. Young, Prof. Rainer Zimmerman.

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: David Roscoe, Department of Applied Mathematics, Sheffield University, Sheffield S3 7RH. [Tel: 01142 235 2785, Email: d.roscoe@sheffield.ac.uk]

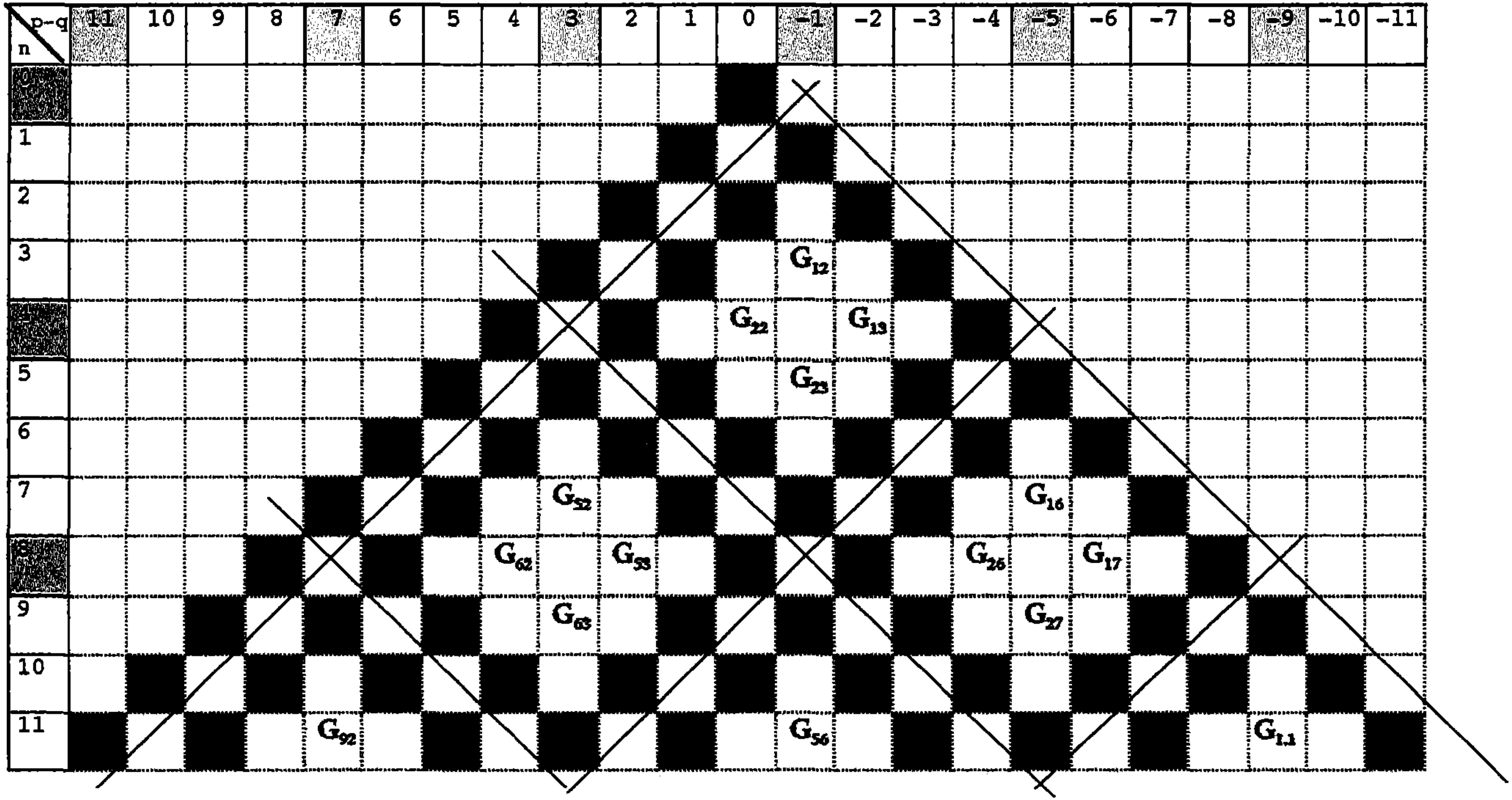
Proceedings Editors:

Philosophical Aspects: Arleta Griffor, 1 Venetia Rd, London N4 1EJ. [Tel: 0208 340 7985, Email: a.griffor@physics.bbk.ac.uk]

Scientific Aspects: Keith Bowden.

Advisory Board: M. Horner (Chairman), Profs. G.F. Chew (Berkeley), C. Isham (Imperial College), M. Redhead (Cambridge), N. Cartwright (LSE), C. W. Kilmister (London, retired), H. Pierre Noyes (Stanford, retired).

CLASSIFICATION OF GROUPS $G_{p,q}$



n \ p-q	11	10	9	8	7	6	5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
0												■											
1											■		■										
2										■		■		■									
3									■		■		G_{12}		■								
								■		■		G_{22}		G_{13}		■							
5							■		■		■		G_{23}		■		■						
6						■		■		■		■		■		■		■					
7					■		■		G_{52}		■		■		■		G_{16}		■				
8				■		■		G_{62}		G_{53}		■		■		G_{26}		G_{17}		■			
9			■		■		■		G_{63}		■		■		■		G_{27}		■		■		
10		■		■		■		■		■		■		■		■		■		■		■	
11	■		■		G_{92}		■		■		■		G_{56}		■		■		■		$G_{1,1}$		■

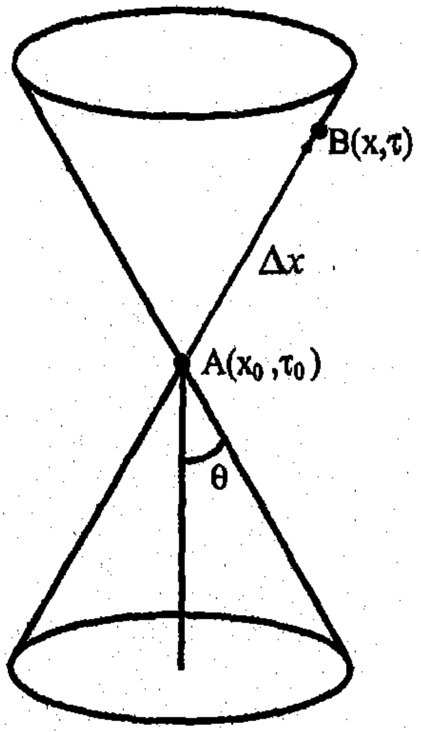


Figure 1: The Hypercone

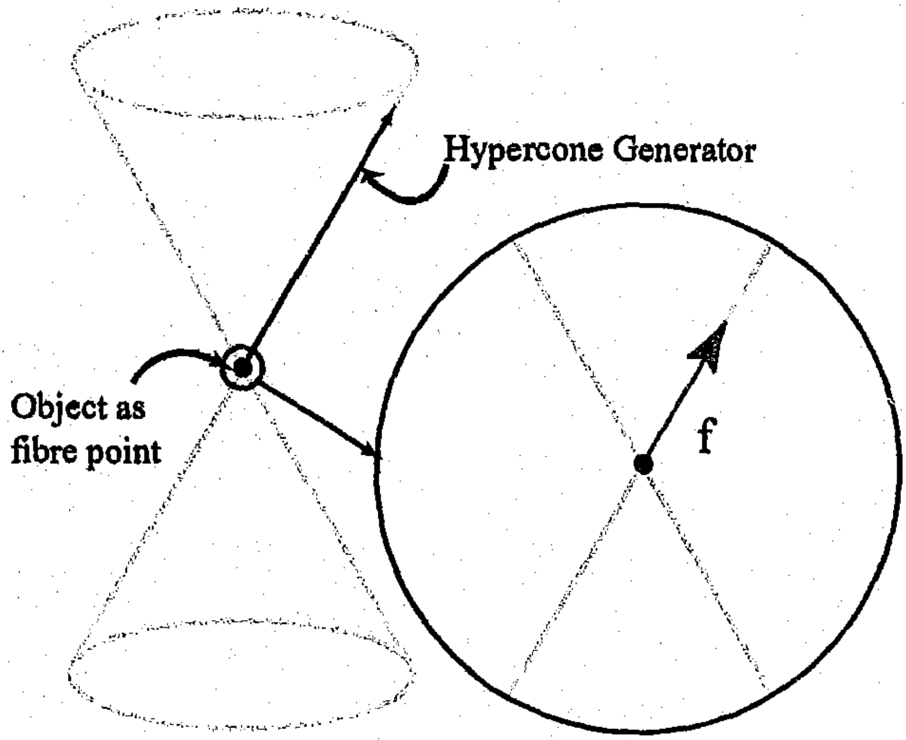


Figure 2: The f-generator

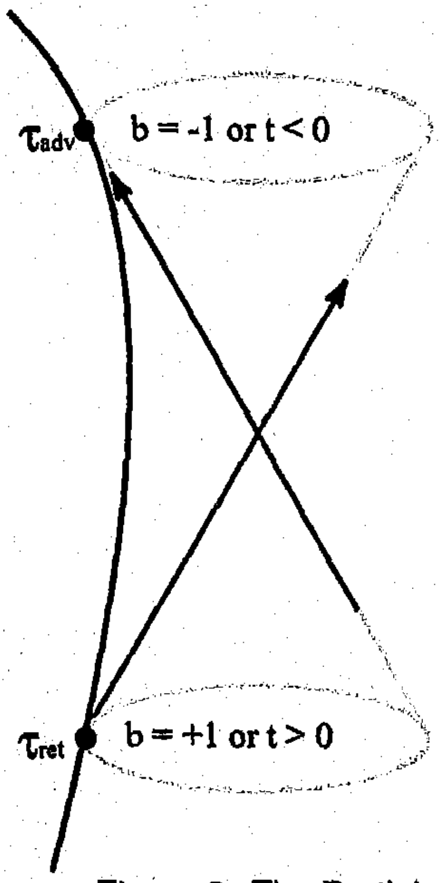


Figure 3: The Particle Solutions

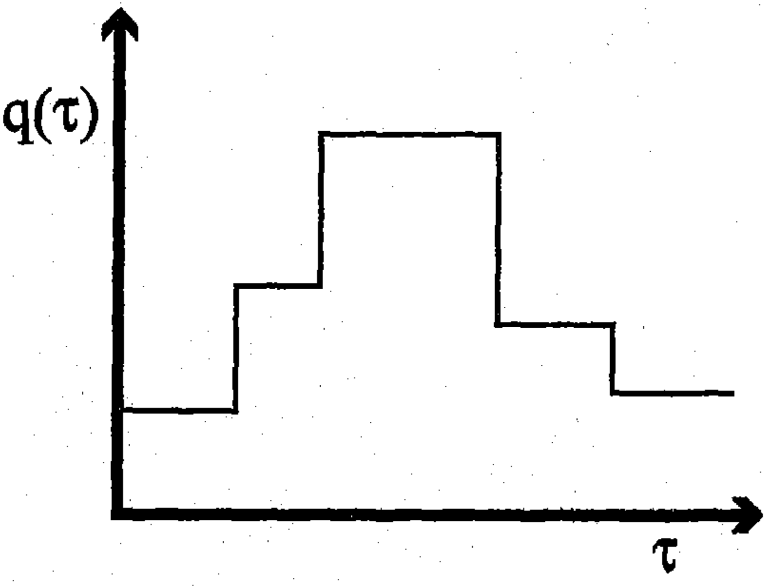


Figure 4: The Discrete Charge Function