SLAC-PUB-4526
June 1989
(T/Noyes)

FOUNDATIONS OF A DISCRETE PHYSICS*

DAVID MCGOVERAN† AND PIERRE NOYES

Stanford Linear Accelerator Center,
Stanford University, Stanford, California 94309

ABSTRACT

Starting from the principles of finiteness, discreteness, finite computability and absolute nonuniqueness, we develop the ordering operator calculus, a strictly constructive mathematical system having the empirical properties required by quantum mechanical and special relativistic phenomena. We show how to construct discrete distance functions, and both rectangular and spherical coordinate systems (with a discrete version of "π"). The richest discrete space constructible without a preferred axis and preserving translational and rotational invariance is shown to be a discrete 3-space with the usual symmetries. We introduce a local ordering parameter with local (proper) time-like properties and universal ordering parameters with global (cosmological) time-like properties. Constructed "attribute velocities" connect ensembles with attributes that are invariant as the appropriate time-like parameter increases. For each such attribute, we show how to construct attribute velocities which must satisfy the "relativistic Doppler shift" and the "relativistic velocity composition law," as well as the Lorentz transformations. By construction, these velocities have finite maximum and minimum values. In the space of all attributes, the minimum of these maximum velocities will predominate in all multiple attribute computations, and hence can be identified as a fundamental limiting velocity. General commutation relations are constructed which under the physical interpretation are shown to reduce to the usual quantum mechanical commutation relations.

1. INTRODUCTION

The purpose of this paper is to present a self-contained mathematical foundation for the modeling of diverse phenomena—in particular, physical phenomena—and to demonstrate its utility.

Contributed to the 9th Annual International Meeting of the Alternative Natural Philosophy Association, Cambridge, England, September 23–28, 1987. To appear as Chapter 2 in DISCRETE AND COMBINATORIAL PHYSICS: Proc. of ANPA 9, H. P. Noyes, ed.; published by ANPA WEST, 25 Buena Vista Way, Mill Valley, CA 94941

*Work supported by the Department of Energy, contract DE-AC03-76SF00515.

† Permanent address: Alternative Technologies, 150 Felker Street, Suite E, Santa Cruz, CA 95060, USA.

Twentieth century foundational mathematics is caught on the horns of several dilemmas. Perhaps the most difficult of these dilemmas is also the most ancient: the separation of description and process or, as more usually encountered, the separation of mind and body. This dilemma manifests itself in the split-mind with which the practitioner of mathematics must operate. On the one hand, we perform finite computations by prescribed methods; on the other, we must keep forever in mind that these are artificial limitations of space, time, energy and symbolism—as is evident in the ever present use of ellipses and the infinity symbol. The description ignores the process of describing.

Somehow the student of mathematics must simply accept the fact that we never quite complete (and in principle cannot complete) many tasks of either description or describing, but must extrapolate. Such acts of faith are deeply embedded in the foundations. Of course, one should not be too concerned that counterfactual paradoxes arise as a result of following the faith with fervor or that one can prove that a mathematical system, if moderately powerful, cannot be both consistent and complete[1]. One must simply accept. At once, the student must pretend that the system is faithful (generates trustworthy results) and unfaithful (is either inconsistent or incomplete).

Twentieth century foundational physics is caught on the horns of a similar dilemma. The practitioner of laboratory physics appeals to the theoretician to completely describe his practice in an objective manner. Again, on the one hand, we perform finite measurements and computations by prescribed methods, while on the other hand we are asked to accept the fact that these are artificial limitations of space, time, energy and symbolism. Again the description ignores the process of describing. Dirac[2], seems to have been acutely aware of this separation of practice and formalism in dealing with the physical interpretation of discrete eigenvalues versus a range of eigenvalues:

"An eigenstate of ξ belonging to an eigenvalue ξ' lying in a range is a state which cannot strictly be realized in practice, since it would need an infinite amount of precision to get ξ to equal exactly ξ' .... Thus an eigenstate belonging to an eigenvalue in a range is a mathematical idealization of what can be attained in practice. All the same such eigenstates play a useful role in the theory and one could not very well do without them. Science contains many examples of theoretical concepts which are limits of things met with in practice and are useful for the precise formulation of laws of nature, although they are not realizable experimentally, and this is just one more of them. It may be that the infinite length of the ket vectors corresponding to these eigenstates is connected with their unrealizability, and that all realizable states correspond to ket vectors that can be normalized and that form a Hilbert space."

Neither the general nor the special theories of relativity readily admit of quantization. These theories are formulated within the space-time continuum using differential geometry. In conflict with this, quantum events are unique, discrete, irreversible, nonlocal, and yet indivisible. Conventional quantum theory tries to embed them in a space-time continuum, which is the source of many conceptual difficulties such as the "collapse of the wave function," the EPR "paradox" and the infinities of second quantized field theory. The properties of quantum events are more fundamental mathematically and conceptually than the properties of an abstract continuum.

One cannot construct a basis which is adequate for this thinking and for the description of phenomena with a language which is dependent on an embedding of discrete structures in a continuous one. We will develop terminology afresh, without the taint of a continuum (and infinities). Our point of view is more process-oriented than just descriptive: it must be possible to generate the structures and the properties which we explore.

L. E. J. Brouwer and others have attempted to constructivize mathematics since 1907 based upon severe and successful criticisms of classical mathematics. As noted by Bishop[3], "Many mathematicians familiar with Brouwer's objections to classical mathematics concede their validity but remain unconvinced that there is any satisfactory alternative." These are valid criticisms, but so are similar criticisms of various constructive attempts, which fail to recognize any of what we feel are some of the more essential aspects of the practice of mathematics. In particular, mathematics which is not process-oriented, context-sensitive, finite, discrete and constructive (primarily in Bishop's sense)[4] is of little use in practice, since the Universe in and about which mathematics is to be used is all of these things. The Universe is only knowable as a complete, consistent system: there exist no black holes arising from undecidability, halting problems, incompleteness or magic of any kind. It is not knowable or understandable in terms of its parts alone. We are strict, constructive, *systems* mechanists.

While we contend that the mathematical foundation presented here will indeed prove useful outside of physics (and we have reason to believe it will), the focus of this paper is restricted to demonstrating the utility of the mathematics for physics.

In order to construct a discrete basis for physics, we limit ourselves from the start to a finite number of symbols (e.g., 0, 1) and to an order parameter defined in terms of primitive recursion. In ordinary language, this allows us to count up to (or down from) some finite integer N which we specify in advance. No construction will be allowed to exceed this integer without additional articulation of the extant theory. This additional articulation will be consistent with and guided by our approach. These restrictions allow us to d–map our construction onto any "operational" description of physics in a sense even more strict than Bridgman's "pointer readings" and the finite specification of what operations are needed to make "pointer readings" are allowed only if we can reduce the operations to "counting." That this apparently impoverished starting point leads to interesting physics will be demonstrated in what follows. In particular, we achieve a fresh understanding of a number of the best established physical facts.

The context-sensitive process of ordering is fundamental in this: simple but subtle notions of ordering, carefully formalized, result in a rich mathematical structure. If one insists on finiteness, discreteness and a strong constructive approach, the power of the system is surprisingly undiminished from that of continuum mathematics[5]. For example, where others have claimed that a finite, discrete topology was indefinable, we assert that the proper notion of open set defined within the formalism is in fact more constructive than the usual definition from point set topology or Intuitionistic Zermelo–Frankel (IZF) set theory, and clearly avoids the paradoxes generated by the usual continuum-oriented definition of open neighborhood or open set.

In this paper, five principles will be introduced which should not strain the reader's credulity: finiteness, discreteness, finite computability, absolute nonuniqueness and strict constructionism. Then, after presenting eight key concepts (indistinguishables, d–sorts, ordering operators, d–sets, open d–sets, d–subsets, parameterization, dimension or basis and attributes) within the context of a larger development, the following consequences will be constructed: the 3+1 dimensional structure of space-time, a combinatoric construction of π, identification of the speed of light constant, the Lorentz Transformations, the relativistic Doppler shift, the relativistic composition law for d-velocities, the uncertainty principle, superluminal correlations without supraluminal communication, a combinatoric construction of the exponentiation operator, the commutation relations for linear and angular momentum, the de Broglie relations, the relativistic mass change, identification of Planck's constant and momentum conserving events.

1.1 Principles

We will develop a theory which, both in terms of the constructs and operations defined on those constructs, possesses the properties expressed in the following five principles.

Principle I: The theory possesses the property of strict finiteness.

By finiteness, we mean that no infinities or infinitesimals are allowed in the theory. By infinities, we mean an x such that x is larger than any finite y in the system. By an infinitesimal, we mean an x such that x is smaller than any finite y in the system and is not identical to 0. In particular, no x in the system can be arbitrarily large or small. Furthermore, and in keeping with strict finiteness, we require finite definability of any derived (constructed) system, subsystem or attribute of a system.

Principle II: The theory possesses the property of discreteness.

By discreteness, we mean that the depth of partitioning by recursive descent (as by Dedekind cuts) or construction by recursive ascent (as in the construction of the transfinites) is bounded in advance from outside the theory. This absolute bound on the practice is a pragmatic constraint. Over the course of any effort, a particular bound will evolve by refinement.*

* As an example, consider any practice which is realizable on a physical computer. The bound is fixed in advance by the amount of accessible storage. It is our point of view that altering this bound constitutes an alteration of the system (computer plus algorithm) which cannot be understood or modeled within the system. Thus, a system which allows for changes to this bound is ill-defined. If the behavior of a program written to run on a computer having a certain amount of memory is in any way dependent on the amount of memory available, then it is clear that changing the amount of memory available requires the programmer to reevaluate the program for unplanned behavior. If the programmer is wise, this is taken into account by coding "system parameters" into the program such that the system alterations will be "automatically" handled.

Principle III: The theory possesses the property of finite computability.

By finite computability, we mean that the theory is constructive in the following strict sense. It must always be possible to specify any procedure or argument used in the theory as an algorithm having a finite number of finitely definable steps and consuming a finite amount of memory. Such a theory is Turing computable, but only theories which are both Turing computable and which use a finite tape are finite computable. Those which use countably infinite tapes or countably infinite algorithms are excluded by this principle.

Principle IV: The theory possesses the property of absolute nonuniqueness.

Simply put, we assume indistinguishability and uniformity unless we have constructively stated otherwise. By absolute nonuniqueness, we mean that no property which serves to single out or distinguish a construct within the theory from any other construct within the theory may be used in the construction in the absence of an explicitly stated computational mechanism. That is, we will invoke a finite algorithm within the theory whenever a property is to be used in a construction and will otherwise be required to deal with the absence of that property (i.e., by probabilistic means). Any two differently labeled, but otherwise indistinguishable, constructs must be treated as interchangeable in the absence of such an algorithm. Thus, the only *a priori* property that is acceptable is recognition of a lack of information as evidenced by indistinguishability.†

Principle V: The formalism used in the theory is strictly constructive.

Following Bishop[6], and in addition to the preceding Principles I–IV, we will argue by constructive means. As such, proof by contradiction will be considered to be justified, since we are restricted by Principle I to finite situations. The only way in which we may show that an object exists is to give a finite means for constructing it. Bishop would say "finding it," but we do not accept the idea of *a priori* existence of nonfundamental objects. Complex (derived) objects are constructed, not found.

A property P is called definable in the system, if, for every object x constructively shown to exist, x has a property P or it does not. This is different from saying that it has the property P or else it has the property "NOT P." If this cannot be said, then the property P is not constructively defined or even definable within the finite system. Within these constraints on the allowed subject matter, we will deny arguments by the principle of omniscience and of limited omniscience, except (again contrary to the position of Bishop) where the latter may be supported by a finite search. Because our theory is finitary, we embrace the Law of the Excluded Middle (as would Bishop).

We call this position **strict constructionist** because we understand it to be more restrictive than the constructive positions of both Bishop[7] and Beeson[8], which are among the more restrictive statements of the position, and clearly more so than Brouwer.

† As we will see, this very general principle is at the heart of most invariance principles, including the assumption of equal *a priori* probabilities, isotropy, homogeneity and relativity.

2. MATHEMATICAL FOUNDATIONS I

In this chapter and the next two, we develop a strict constructive mathematical system which we refer to as the **ordering operator calculus**. This system will be shown to have sufficient power to be a foundation for, or simply replace, significant aspects of conventional mathematics including set theory, lattice theory, differential topology, real and complex analysis and differential geometry.

2.1 Preliminary Concepts

Several concepts will be taken as fundamental in the development of our theory. These concepts are well-known to computer scientists and are rigorously defined by them. Nonetheless, we will provide definitions which limit the scope and applicability of the terms, since our usage will in general be more restrictive. It is especially important for the reader to keep in mind that we do *not* import the additional theoretical framework which is normally accepted within computer science and discrete mathematics.

Recursively Definable

By **recursively definable**, we shall mean simply that an abstract term is definable with a finite number of steps from simpler terms and values.

Computable

By **computable**, we shall mean that an effective procedure has been given by which an abstract construct can be constructed in a finite number of steps and with finite resources. We shall use the term *recursive* in a manner similar to that used in recursive function theory, in that it includes both recursive and iterative algorithms and is not restricted to mean a "recursive procedure call" in the computer programming sense.

Computational Complexity

By the **computational cost** $C(O)$ of an abstract, finite, discrete algorithm O, we shall mean a measure of the time cost and the space cost of the algorithm. Each of these is usually expressed as a procedure, which shows how to compute from the cardinality and/or ordinality of the domain upon which the algorithm operates (usually called the *size* of the problem), and yields a measure of the **computational time cost** $C_t(O)$ or the **computational space cost** $C_s(O)$ of the algorithm in time-like units (e.g., CPU cycles or algorithmic steps) or space-like units (e.g., bits), respectively. Note that, for us, these costs include the cost of running and storing the algorithm itself.

It is considered normal to express the **computational complexity** measure in terms of the dominant term of the appropriate polynomial, logarithmic, exponential or combinatorial expression; we will consider this to be shorthand for the exact

expression. An algorithmic procedure $g(n)$ will be said to be of computational complexity

$$O[f(n)]$$

read "of order $f(n)$," if there exists a rational constant c such that

$$g(n) \leq cf(n)$$

for all n.* By the **total computational cost** of an abstract, finite, discrete algorithm O, we shall mean the result of a procedure which computes for each pair of inputs $C_t(O)$ and $C_s(O)$ a finite number $C(O)$ in a finite number of steps. Such a procedure (which in classical mathematics is representable by a polynomial expression) is said to represent a **computational metric.**

Representational Resources

By **representational resources** of an abstract, finite, discrete system, we shall mean the maximum of the spatial complexities of those algorithms which may be expressed within the system, without appeal to either spatial or time resources outside the system.

2.2 The Concept of Order Revisited

Ensembles

Consider a collection of mathematical (in the sense that physical properties are neither implied, nor are they denied) objects about which we have no knowledge, other than their quantity (cardinality), together with a collection of (mathematical) operators for selecting some of those objects. We call this collection of objects an **ensemble**, because it differs from the usual set-theoretic notion of a collection in ways which we now explain.

Ordering Operators

The notions of distinguishability and indistinguishability of such objects are relative. Without a stated computational mechanism, we are required to assume indistinguishability in keeping with Principle IV.† When asked if two objects are distinguishable, one must respond with a question, "distinguishable with what algorithm?". If presented with such an algorithm, we can think of that algorithm as inducing a property on the objects on which the algorithm operates; then the question of distinguishability becomes, "distinguishable with regard to such and such a property." Indeed, whether the objects are "truly" indistinguishables or not in the sense of

* It is usually permissible that a finite number of values of n violate the inequality. We do not allow this.

† This is not an ontological statement.

Parker-Rhodes[9] is irrelevant: our inability to directly access the objects makes the properties of the computational mechanism used on the objects the essential knowledge in building our theory.

We choose a single means of establishing structure in our formalism, namely, the generalized concept of ordering relation called an **ordering operator.** These computational mechanisms are defined as having the following properties:

1) they are only defined on a finite ensemble (a domain);
2) the ensemble must have fixed cardinality N;
3) they take as single input a label;
4) each label carries an embedded unique inaccessible sequence number;
5) they operate on the ensemble or some portion of it;
6) they generate as output one or more labels;
6) the labels successively generated are not necessarily unique;
7) the labels so generated constitute a finite ensemble;
8) the mechanism has a stop rule,
9) the details of the mechanism, including the stop rule, are not inferable.

Note that without either the identification of the ensemble and the input or recognition of the output, there is no knowledge that the operator has been used. By recursively applying this mechanism, we generate an ordered sequence of labels. Clearly, the ordering operator counts as a generating function in the sense used by Kilmister[10], although it does not require the same mathematical foundation and has additional computational power. Since we lack knowledge about the nature of the indistinguishables, we need to specify a few more characteristics of the mechanism of ordering operators. Having done this, ordering operators then also serve an essential function in our axiomatic system as general rules of inference, since they determine precisely what can be constructively exhibited or evaluated.

By **indistinguishables**, we mean that, given the ordering operator mechanism, the objects in an ensemble come in two forms which we now define. By **identicals** we mean that there exists no algorithm constructed within the formalism which serves to distinguish two objects. Thus, identicals is‡ what one gets when an ordering operator operates twice on the "same" object. By **twins**, we mean that the algorithm used to manipulate the objects does not distinguish them, but that there exists some algorithm constructed within the formalism which does distinguish them. Thus, twins are what one obtains when the ordering operator operates on two objects, but does not distinguish between them in its output; that is, the objects seem to us to be the same within the context of the specific ordering operator. Thus, two objects are indistinguishables only for a specific alogrithm.

‡ The grammatical "agreement" as used here is intentional.

The output which results from using the ordering operators in either of these first two cases is two indistinguishable but sequence ordered object descriptions which we will call **labels** for short.§ Thus, the ensemble of objects has no inherent ordinality as far as we are able to know.

Above, we said that the ordering operator operates on the ensemble. Specifically, we mean that the ordering operator picks a finite number of indistinguishables, given a label as input. If the operator picks more than one object, successive recursions of the operator via input of a label generate one sequence numbered label per object in the subcollection, until the subcollection has been exhausted. The sequence numbers "stick onto" the objects, and their significance can only be recognized by the ordering operator that generates them (it maintains the equivalent of a symbol table which allows it to look up the sequence number(s) associated with a label and vice versa); thus, other ordering operators simply ignore the sequence numbers if operating on the same ensemble. The subcollection is then returned to the ensemble. Further input of a label returns one to the initial situation.

Note that this mechanism allows the operator to generate both total and partial orderings of the labels. The ordering operator algorithm has a stop rule (it halts in a well-defined manner) and will not allow, without repetition, recursive generation of more than a fixed and finite number of labels. The process is defined with (a) the operator and (b) a unique starting label. For some label input, the number of labels output by recursively feeding the output label into the input (i.e., **recursive generation**) is a maximum. The maximal label output of the operator and the ordering operator, itself, are **mutually defining**. Thus, given a finite ordering of labels on a given ensemble, we define an operator, and vice versa. Finally, the complexity of the operator mechanism (i.e., the algorithm) is too great to be represented by the labels alone. We would also have to know the intrinsic nature of the ensemble, but this can only be investigated with (other) ordering operators.

Suppose that a particular ordering operator O on a specific ensemble C (its "domain" in this instance) is given as input a specific label L_0. Let the resulting output of O be the label L_1. On input of L_1, O generates L_2, etc., up to some finite number of labels N. These labels need not be unique; however, each corresponds to some object in the ensemble C. Suppose that this correspondence is such that at least one label has been generated for each object in the ensemble C. If we keep a record of the labels so generated, we can be certain that rerunning the generation will produce an identical record, given the same ensemble C and the same initial label L_0. Indeed, if we begin with L_1, O produces an identical record with exception of the missing entry for L_0. If we begin with L_2, O produces an identical record with missing entries for L_0 and L_1.

If several successive entries in the record are equivalent except for the sequence in which they were generated, we cannot know whether the objects in the underlying ensemble to which these labels correspond are twins (indistinguishable but distinct) or

§ We suggest the use of **tags** where the term labels would be otherwise confusing as, for example, in Noyes[11] where label refers to a particular kind of label in our sense of the term.

identicals (indistinct). This can only be known by detailed knowledge of the objects in the ensemble and the algorithm by which the operator works. One might argue that in reproducing a record starting from equivalent records, that the recursion would not terminate since the labels are equivalent. However, note that conveying both the label and its record entry sequence is required for entry if the record is to be reproduced. The use of the notation L_n is not accidental. Two pieces of information are conveyed; the label and a number representing the sequence in which it was produced.

Define an operator $O^{\#}$ associated with O that behaves as follows: if the sequence number is left out, then $O^{\#}$ selects a default sequence number for the particular label consistent with the possible sequence numbers with which it might be produced. Thus, the operator may generate labels in two modes: with the sequence number or without it. When the sequence number is excluded, the recursive use of the ordering operator is similar to a sampling algorithm, subject to the constraint of an ordering relation. This ordering relation is not, in general, transitive. In this case, it is possible for $O^{\#}$ to generate L_M multiple times, given only label L for input. The output of $O^{\#}$ must then be ordered on the output sequence numbers to recover the ordering relation which $O^{\#}$ mutually defines. Unless we refer to sampling with repetition allowed, we will mean that $O^{\#}$ has as input and output both the sequence number and the label. However, we will ordinarily refer only to the input and output labels, the sequence being assumed and, since we then do not distinguish between O and $O^{\#}$, we will simply use O notationally.

Note that without the underlying objects, the algorithm for the ordering operator cannot in principle be defined, since the nature of the algorithm will depend upon the nature of the ensemble. Furthermore, depending upon which ensemble an ordering operator operates on, the statistical distribution of labels so generated (with or without repetition) is determined by the intrinsic character of the objects in the ensemble; namely, the cardinality of the ensemble and whether there exist indistinguishables or not, and how many. At best, having run through the operator once, we may use the output "sequence"* as a "look-up" table for additional runs, but only in the sense of checking off what has been generated so far and, in the case of a partial ordering, what is left.

2.3 Classifications of Fundamental Objects and Operators

The elementary unstructured object of our mathematics is taken to be a d–sort. We define a **d–sort** as any ensemble of n indistinguishable objects having cardinality n, of which it is NOT asserted that every pair of members is either identical or distinct[12]. A **perfect d–sort** is a d–sort for which every pair of members is either identical or twins (indistinguishable). We allow the members of a d–sort to be labeled by an

* We do not mean to imply that the output is "sequential" or totally ordered by the use of this term.

ordering operator, noting however that such internal labeling is inaccessible except via the ordering operator which performs the labeling.†

In order to formally define the relationship between ordering operators and the objects of our formalism, we have need of some additional definitions.

An **ordering relation** $\leq$ is a binary relation (a relation on two arguments). From time to time, we will take the liberty of writing $y \geq x$ in place of $x \leq y$. We shall mean by the symbol for **equivalence** $=$, a binary relation such that the two arguments are either identicals or twins; that is, they are the members of a perfect d–sort. A **successor function** $'$ is an ordering relation such that P1 and P2 are satisfied:

P1: Given arbitrary objects a, b, b' and x in a d–sort S such that if $a \leq x \leq b$, $a \leq b$, $a \leq b'$, and either $a = x$ or $x = b$, then $b = b'$ (uniqueness).

P2: There exists an e in d–sort S such that $e \leq x$ for all x in S (infimum).

A **recursive enumeration** E is an ordering operator which provides or recovers a label for each member of a d–sort. It is, therefore, an effective procedure for listing the members of a d–sort, with repetition allowed. More formally, a recursive enumeration is a rule with successor function ($'$) such that, given a label for object x in a d–sort S as input, the recursive enumeration generates a label for object x' in S, not necessarily distinct from x.

Theorem 1: Neither the enumeration nor the successor function on a given d–sort are unique.

A **partial ordering relation** $\leq$ is a binary relation between two members of a d–sort S which, for all arguments x, y, or z in S, satisfies the following conditions:

P3: For all x, $x \leq x$ (reflexive).

P4: If $x \leq y$ and $y \leq x$, then $x = y$ (antisymmetric).

P5: If $x \leq y$ and $y \leq z$, then $x \leq z$ (transitive).

A **parameterization** is a partial ordering relation induced on a d–sort by an ordering operator O such that, given a label x for a member x of S as argument, the parameterization generates a label x' for the successor of x, called x'. That is, the partial ordering relation satisfies P1 and P2, as well as P3, P4 and P5. We may refer to a ordering operator O as a parameterization if O is used to induce a parameterization.

A **total ordering relation** is a partial ordering relation which satisfies P6.

P6: Given arbitrary x and y in d–sort S, either $x \leq y$ or $y \leq x$.

† Parker–Rhodes was insistent that indistinguishables could not be labeled at all. In this respect, we suspect that Parker–Rhodes would have likened d–sorts to a kind of multiset[13]. With a suitable mapping we could identify the notion of an ordering operator with Parker–Rhodes' functor in which case the ordering operator would produce ordinal sorts. However, this would relegate our theory to the domain of sort theory, an extra degree of ontological freedom and notational complexity which we cannot afford.

We shall deem it convenient, at times, to speak of a particular type of recursive enumeration. In particular, we will want a recursive enumeration without repetition, and in which the binary relation is a partial ordering relation.

A **rule of correspondence** is a total ordering relation induced by an ordering operator O on a d–sort of cardinality 2.

A member x of a d–sort S of cardinality 2, with partial ordering, is called a **supremum** or **sup** if, for member x and arbitrary y in S, $y \leq x$ and it is not the case that $x = y$. Similarly, a member z of S is called an **infimum** or **inf** if, for members y and z, $z \leq y$ and it is not the case that $y = z$.

Theorem 2: Every total ordering relation induced on a finite d–sort defines a supremum and an infimum.

Theorem 3: A rule of correspondence defines a supremum and an infimum.

A member a of a d–sort S with ordering operator O inducing an ordering relation $\leq$ on S is an **upper bound** if there exists a member b such that for the d–sort S' consisting of a and b, with ordering operator O' inducing the ordering relation $\leq$, a is the sup of S'. Similarly, a member b of a d–sort S with ordering operator O inducing an ordering relation $\leq$ is a **lower bound** if there exists a member a such that for the d–sort S' consisting of a and b, with ordering operator O' inducing an ordering relation $\leq$, a is the inf of S'.

Note that we have used partial ordering in the definitions of sup and inf so that we may create **d-chains** of d–sorts with the same partial ordering. Defining the concepts of upper and lower bound in this way insures transitivity across d–sorts; then these concepts take on the usual lattice theoretic definitions and the uniqueness of the sup and inf (in a given d–sort of cardinality n with specified partial ordering) is assured[14].

2.4 Constructed Objects: From d–Sorts to Coordinates

A **d–set** is a d–sort with ordinality m imposed by one or more recursive enumerations. The ordinality m of the d–set is just the cardinality of the d–sort of labels given by the recursive enumeration. Note that there may be more than one such recursive enumeration associated with the d–sort.

We may now classify the generations of an ordering operator in terms of the cardinality and the ordinality of the labels it generates. When these are the same, the output is a d–set and the O is said to be a **total ordering operator**. When these are not the same, the output is a d–sort and O is then said to be a **partial ordering operator**. Like sets in set theory, d–sets have no members that cannot be counted uniquely, while d–sorts have members that can not be counted uniquely. Unlike sets, a d–set is only defined with respect to one or more *specific* ordering operators. It, like all other objects in our system, is *constructed*.

We say an enumeration E is **monotonic increasing** if it gives labels to elements in the order of the recursive enumeration of the d–set on which it is defined. Similarly, we say E is **monotonic decreasing** if it gives labels to elements in reverse of the order of the recursive enumeration of the d–set. An enumeration E will be said to be **nonmonotonic** if it cannot be shown to be either monotonic increasing or monotonic decreasing constructively.

We are now in a position to define an important concept of topology; the notion corresponding to an open set. Note that our definition in no way appeals to the notions of continuity (in the usual sense of the word) or infinitesimals.*

The **boundary** of a d–set S defined with ordering operators O_i generating ordering relations R_i, consists of those elements of the underlying d–sort which, when operated on by any of the O_i, generates a label which is either a sup or inf of the R_i.

A d–set S' is said to be an **open d–set** with respect to a d–set S defined with ordering operators O_i, if S' is just S without the elements which would generate the boundary of S. Thus, for the defining ordering operators of the d–set S, none of the defining ordering operators of the d–set S' on an element of the underlying d–sort of S' generates a label which is either a supremum or an infimum of the defining ordering operators for S. It will generally be the case that d–sets are formed from multiple ordering operators. The extension of the definition to more complex d–sets having multiple ordering operators is straightforward.

Clearly, from the definition of ordering operator (i.e. an ordering operator and its productions are mutually defining), S' is, itself, a closed d–set, but *for a different ensemble of ordering operators.* This makes clear the importance of the notion that a d–set is a d–set by virtue of the defining ordering operators.

Note also that this eliminates the possibility of d–sets with deleted points. The transitivity of an ordering relation is, itself, defined constructively. Thus, the transitivity of any ordering relation is broken if a point is "deleted" and new ordering relations are induced, resulting in a new d–set. By defining open d–set as above, we have insured that there is no means of specifying a *classical boundary* for the *a* set independent of the construction of the set, as is done with a classical (infinite or continuum) set. Thus, every enumeration of the elements of a d–set would have to be an infinite enumeration, either by allowing (infinite and therefore not constructively definable) repetition (in the case of d–sorts) or without repetition (i.e., only if the d–sort is itself infinite and, in which case the d–set cannot be constructed as the ordering operator, is finitely definable only when all the members of the d–sort can be finitely specified). Thus, for d–sorts of sufficiently large cardinality, open d–sets are not distinguishable from the open sets of the classical definition.†

The fundamental concept of local topology is now within reach; i.e., an open neighborhood. A **d-subsort** (or **d–subset**) S' of a given d–sort (or d–set) S is itself a d–sort (d–set) which is defined with the same ordering operators, and for which x is an element of S' if and only if x is an element of S. We say that the S contains

* In addition, the definition is purely constructive and recovers the "classical" definition for sorts of sufficiently large cardinality n. We shall define what we mean by "sufficiently large" in a future paper in which we will discuss measurement—both abstract and physical—using the terminology developed here. Briefly, one measures the number of partitions n of a model by a d–map G to a d–sort of cardinality n'. Then, if $n > n'$ for all n' chosen, n is "sufficiently large," as no finiteness will be detectable independent of the measurement.

† This will be more obvious after we provide a constructive definition of a smooth recursive enumeration, below.

S'. Note that this precludes the possibility of "supersets" of d–sets being equivocated with d–subsets contained in the d–sets and, thus, the "set of all sets" of Russell's Paradox. For arbitrary elements of a d–sort S, an **open neighborhood of x** is any open d–sort containing an open d–sort containing x.

A **closed d–sort (d–set)** is a d–sort with defining ordering operators O_i, such that at least one element of the d–sort is either a supremum or an infimum for at least one of the O_i.

A **d–map** G on a d–sort S is an ensemble of rules of correspondence defined on S (i.e., inducing a d–sort of cardinality 2), such that there exists a d-subsort of S with members x_i all of which are the infimums of the rules of correspondence, and there exists a d-subsort of S with members all of which are the supremums of the rules of correspondence. We refer to one of these d-subsorts as the **domain** of the d–map G and to the other as the **range** of G. The range is said to be a d-subsort of some d–sort called the **image** of G. The domain and range have ordinality 1.

The **union** of two d–sets S and S' is a d–set S'', whose members consist of the members of either or both of S and S'. Similarly, the **intersection** of two d–sets S and S' is a d–set S'', whose members consist of the members of S which are also members of S'. The **symmetric d–set difference** of two d–sets S and S' is a d–set S'', whose members are either members of S but not of S', or members of S' but not of S.

A d–map M on a d–sort S is said to be **one-to-many**, if and only if the cardinality of the domain is less than the cardinality of the range. Such a d–map is called an **operation**. A d–map M on a d–sort S is said to be **many-to-one**, if and only if the cardinality of the domain is greater than the cardinality of the range and **one-to-one**, if and only if the cardinality of the domain is equal to the cardinality of the range. Such a d–map is called a **function f()**. A d–map M on a d–sort S is said to be **onto**, if and only if every element of S is either a supremum or an infimum of the rules of correspondence. If a d–map M on S is both many-to-one and onto, it is called an **isomorphism**. M is said to be order-preserving or **isotone** if, given an ordering operator on two elements x_1 and x_2 in the domain of M, there exist corresponding elements y_1 and y_2 in the range of M, which are also valid arguments of the induced ordering relation in the following sense: given that x_1 corresponds to y_1 and x_2 corresponds to y_2, it follows that if $x_1 \geq x_2$ then $y_1 \geq y_2$ and similarly, if $y_1 \geq y_2$ then $x_1 \geq x_2$.

Theorem 4: An isomorphism is isotone.

Argument:

If we could not make this choice, there would exist some order on the d-subsorts, and the d–map would not exist. Note that such an ordering relation on the range or domain provide structure and thus increase the ordinality of the d-subsort.

QED

A **bisection** of a d–sort S is any one-to-one d–map defined on S and having both domain and range in S, in that a one-to-one d–map divides the d–sort into a range and a domain.

A **partitioning** is an ensemble of d–maps on a d–sort S such that no element is in the domain or range of any other d–map. Thus, a partitioning of the d–set S is a selection of disjoint d–subsets from the ensemble of d–subsets of S, which are disjoint union S (their union is equivalent to S). Note that a partitioning provides a natural means of "dividing" a d–set into parts, each d–sort distinguishable from the other.

We designate by $\{\}$ a d–set S and by $\{|\}$ the bisection of S. For convenience, we label the d–sorts thus created by a bisection of S, L and R. We call the label for $\{\}$ the **identity label**. The bisection $\{L|R\}$ is the bisection of S into d–sorts L and R. We have defined bisection in such a way that it is invertible. Thus, we may speak of the inverse process from time to time and call this **adjoining**. Note that for any d–set of ordinality 2, bisection yields 2 d–sorts of ordinality 1, namely L and R. Thus, adjoining L to R yields a d–set of ordinality 2. Similarly, adjoining L to S or S to R yields a d–set of ordinality 3. Note that adjoining L to itself (or R to itself) yields a d–sort of cardinality 1. We call this recursive adjoining, beginning with a single d–sort, the **von Neumann recursion**.

We leave the details of this recursion to the reader, but note that it differs from the original recursive process defined by von Neumann and the more recent explication by Conway[15], only in that we do not define a cardinal 0, nor do we require the (∞) real number line, since we are merely generating labels. Note that neither 0 nor ∞ are defined for us, since we cannot show how to construct (or even find in Bishop's terminology) either of them constructively and finitely. For us, the infimum is the successor of 0 and ∞ is the successor of the supremum of a recursive enumeration. Clearly, these are not constructable within the context of the specific recursive enumeration nor are they unique for the collection of recursive enumerations in the system, even though they are ordered with respect to its properly generated labels.

By a **number**, we mean a label given to an element x of a d–set via a monotonic recursive enumeration. We define a **primary enumeration** as follows: first establish a unique label to represent identicals (i.e., the identity element); then, a recursive enumeration which labels the initial element with that of the identity element, and which, on recursion, generates labels for which there exists an isomorphism with the recursively generated elements of the von Neumann recursion is called a primary enumeration. A primary enumeration generates the integers up to the cardinality of the d–sort.

We call a ternary relation $+$ **addition** if, operating on a d–set S of arithmetic elements (i.e., *numbers*), for arbitrarily chosen elements x, y, z, there exist elements of S x' and constant element e_1, such that the following relations hold:

$$x + (y + z) = (x + y) + z$$

$$x + e_1 = x$$

$$x + x' = e_1$$

$$x + y = y + x \quad .$$

We call a ternary relation $\times$ **multiplication** on a d–set S with addition if, for arbitrarily chosen elements of S, w, x, y and z, there exists elements of S w' and e_2, such that the following relations hold:

$$w \times (x + y) = (w \times x) + (w \times y)$$

$$(w + z) \times x = (w \times x) + (z \times x)$$

$$(w \times z) \times x = w \times (z \times x)$$

$$w \times x = x \times w$$

$$w \times e_2 = w$$

$$w \times w' = e_2 \quad .$$

Note that addition and multiplication are both intended to be relations manifested by the ordering operators and could be defined much like the reverse-polish notation calculator which has a single (arithmetic label) display and accepts one input (label) at a time. For operators, closure is not defined explicitly. The existence of a unique starting label and a stop rule guarantees that, for some input labels, the operator will simply stop and perhaps generate a special label. This is quite similar to an "overflow" or "underflow" condition in a physical calculator. In practice, arithmetic closure simply guarantees that a calculator is well-behaved and does not suddenly generate a symbol which is not a number. Our operators have this deterministic element built in, and so there is no need of a closure property.

A **reparameterization** is an ordering relation induced on a d–set by an ordering operator, such that, given a label for the element x of the d–set as an argument, the reparameterization gives a label for the successor of x; namely, x'. A reparameterization is not a primary enumeration.

A **segment [x,y]** for x and y in d–set P, is the d–set of all elements z which satisfy $x \leq z \leq y$. A partially ordered d–set is said to be **locally finite** if every segment is finite. Clearly, all partially ordered d–sets within the ordering operator calculus are locally finite.

Enumerations on d–sets may be divided into two classes. **Normal enumerations** are those for which, though not necessarily monotonic, there exists an isomorphism to the von Neumann recursion and which begin with the identity label. **Subclass enumerations** are those for which there exists an isomorphism to the von Neumann recursion, and which (up to redundancy) establish the identity label as the final label of the d–set.

Theorem 5: A subclass enumeration is a reparameterization.

A **fractional enumeration** is a subclass enumeration in which the labeling of each element is given in comparison to the final label of the d–set. The labeling thus generated follows a recursion relation induced by the inverse algorithm for multiplication. A monotonic subclass enumeration on a d–set of nonfinite cardinality would define the **real number line segment on [0,1]**. Since this is strictly not defineable within the formalism, we define the **discrete real number line segment [0,1]** to be a monotonic subclass enumeration on a d–set S of finite cardinality. We call the cardinality of S the **precision** of the segment. A reparameterization of the discrete real number line segment [0,1], with final label n and initial label m, defines the **discrete real number line segment** $[m,n]$.*

Random versus Arbitrary

As noted in the introduction, where information regarding the construction of a property is not available, we shall be required to deal with the property by probabilistic means. In order to do this, we must introduce a concept of randomness which is constructive and finite. We are now in a position to do so.

Kolmogorov[17] and Chaiten[18] have defined the measure of randomness of a string in terms of the length of its shortest description, an inherent property of individual strings. Namely, if the space complexity of the algorithm is greater than the length of the string it produces, then the string is random. Unfortunately, this definition is not acceptable for three reasons: (a) it allows for infinitely long strings and infinitely complex algorithms, (b) it is nonconstructive (i.e., it does not tell how to construct a random string) and (c) the set of Kolmogorov-random strings is nonrecursive. A number of extensions have been considered, but none give an effective procedure for writing pseudorandom generators.

Suppose that the algorithm for an particular computation O is not known. Select a computational metric. Let the computational cost $C(O)$ of representing the algorithm for O be greater than the **representational resources** n within the finite discrete system S to be constructed. Under certain conditions which we now determine, the algorithm may not be discovered or even constructed within S.

Theorem 6:
An algorithm O with computational cost $C(O)$ is indistinguishable from a "true" random number generator within a discrete, finite system S with representational resources $R(S)$ whenever

$$R(S) < C(O) + log_2 C(O) \qquad (1)$$

(where the operations $<$, $+$ and log_2 have their usual meanings). Call O an **arbitrary binary number generator.**

* Note that if the reparameterization is a one-to-many d–map with range of cardinality greater than the cardinality of the system N, and if monotonic decreasing, we obtain $(m,n]$, and if monotonic increasing, $[m,n)$, and the adjoin of these segments is (m,n). Furthermore, this provides a formal definition of the hierarchical nature of the real numbers. As pointed out elsewhere[16] the equivocation of this hierarchy of classes is the source of a number of apparent paradoxes.

Argument:

Consider (1) a system composed of a Universal Turing machine with a finite memory, and (2) a binary number generator G. Such a system is incapable of deciding whether or not the number generator produces repeating binary strings of length n whenever the memory is smaller than an amount m equal to $n + \log_2 n$.

Suppose that the Turing machine takes as input a particular substring of length n output by G, and we wish it to determine whether or not the number generator G is producing this substring repeatedly. Select as a computational metric the computational space cost C_s, without regard to the computational time cost C_t. The Turing machine must consume an amount of memory equal to n in order to store the string; then, the computational space cost C_s for any computation on the substring, including direct comparison with a second input substring, is at least as great as C_s, for a count of the number of symbols n in the substring ($log_2 n$). Thus, $n + log_2 n$ sets a lower bound on the computational space cost $C_s(O)$ for any algorithm which may be selected to make the decision.

It follows that the system cannot decide whether or not the target string has been produced if it has memory less than $n + log_2 n$. But this means that the system cannot distinguish between number generators which produce repeating strings and random numbers. Clearly, the symbols in the repeating strings will occur with equal probability, as required for a random distribution. However, since the system cannot detect that a given string is repeating, it cannot detect that some string of cyclicity n is repeating. Thus, for systems with less than $n + log_2 n$ memory, a generator producing repeating strings of minimal cyclicity n is indistinguishable from a generator producing random numbers.

QED

This theorem means that we may actually construct ordering operators which are "perfect" pseudorandom generators, in our terms more properly called "arbitrariness generators." Thus, ordering operators can be constructed from existing ordering operators, and not all ordering operators need have *a priori* existence. According to the Theorem, such a situation will give rise to a nondeterminism born of computational complexity and representational impoverishment:* we cannot predict the output of the ordering operator, because we could not even express the complete algorithm if it were "known."

Given this situation, it is possible to understand the ordering operator foundations as arising from a complex, though finite system of space complexity $n + log_2 n$ greater than the space complexity n of the finite system in which we are working. The complexity of the *a priori* ordering operators is greater than the space complexity n of the known system. Note that this does not introduce an infinite regress, since we need postulate an extension of the known system only once to account for conditions of "randomness" and indeterminacy. The notion of *truly random* can have no meaning within the theory.

* The variables involved in an ordering operator's algorithm may rightly be called von Neumann hidden variables. This does not mean hidden variables in the usual "quantum mechanical" sense, since we do not have a Hilbert space. It is interesting to note that von Neumann had similar ideas when he referred to systems with "partial knowledge."

d–Spaces

A d–set S' is said to be a **permutation** of the d–set S, if the only difference between them is the partial ordering relation. Consider a d–set S partitioned into n mutually disjoint d–subsets.† These d–subsets need not be formed by equipartitioning of S, although this is what will usually be meant. In general, however, we will denote cardinality of each of the n partitions by $m_1, m_2, m_3, \ldots m_r, \ldots m_n$, respectively. For each d–subset r, by definition, there exists an ordering operator which generates m_r distinct labels. Call an ordered d–set of the labels, one from each of the n partitions, an **n–tuple** or **d–point**. Form a d–set R from all the possible ordered n–tuples of labels. We call such a d–set R a **d–space**.

A d–space S on which addition holds for the d–points of S and for which there exist elements (defined via a primary enumeration) between which multiplication holds, is called a **vector d–space**. The d–points of a vector d–space are called **d–vectors**. By a **d–basis**, we mean an ensemble of n totally ordered d–sets.

A **d–curve** on a d–space is a d–set of d–points for which there exists at least one basis d–set which can be mapped 1–1 onto the d–curve, and for which there exists at least one total ordering on the d–basis, which is preserved by the d–map. A **smooth d–map** is defined here to mean that there exists a partial ordering over the domain and a partial ordering over the range of the d–map, such that the n^{th} enumeration in the partially-ordered domain maps to the n^{th} enumeration in the partially-ordered range (i.e., the d–map is isotone but not necessarily 1–1), and for every reparameterization of the domain there exists a reparameterization of the range which is isotone.

The **derivate** of a recursive enumeration $f(n)$ is the number x, where

$$P7\text{: } x = \frac{f(n) - f(n+h)}{h} .$$

Thus, x is just the divided difference $[n, n+h]$ of $f()$. Because the primary enumeration generates the integers, h is just 1 if $f(n)$ is a primary enumeration. Then x is also the **forward difference**. Indeed, most of the results of the calculus of finite differences may now be taken over intact[19].‡

A recursive enumeration $f(n)$ (or rule of correspondence or a d–map), is **locally differentiable** over some d–sort S if the corresponding primary enumeration exists; then there exists a number x such that $f(n) + x = f(n+1)$. Although there may exist a recursive enumeration on a d–sort, the primary enumeration need not exist if, for example, the d–sort is only partially ordered. A smooth d–map, which is locally differentiable for all the labels in the domain of the d–map, corresponds to the classical notion of a continuous function.

† We could just as well begin with n mutually disjoint d–sets S_n, and form a new d–set S which is the union of the disjoint d–sets; however, this would require care in specifying the ordering operator on which S is defined.

‡ Note that for d–set with sufficiently large cardinality m, one may reparameterize P7 to read $P7'$: $f(n) + x\frac{1}{m} = f(n + \frac{1}{m})$, which reduces to the classical definition (L'Hopital's Rule) in the limit of large m, though without appeal to infinitesimals.

The series formed by summing a recursive enumeration $f(n)$ for successive values of n is said to **converge** if the recursive enumeration is monotonic decreasing, and **diverges** if the recursive enumeration is monotonic increasing.

If the domain of a function, with range defined on the discrete real number segment [m,n], depends on the parameterization chosen for the range, then we call it a **real-valued function**; otherwise we call it a **scalar function**.

For recursive functions with multiple arguments, we define the **partial derivate** $f_i'(x)$ as

$$f_i'(x_1, x_2, \ldots x_n) = f(x_1, \ldots, x_i + 1, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n) \quad .$$

With this definition, we recover the inverse function theorem for inverse suitably defined over d–sets. Further, the determinant (computed in the usual manner) will go to the infimum of the d–set, if the x_i are dependent. If the x_i are independent, then we are guaranteed that there exist linearly independent recursive functions

$$f_i(x_i)$$

such that some linear combination of the f_i yield f.

A **chart** is an 2–tuple consisting of a neighborhood N and a d–map from N to some d–space R^N, whose N disjoint d–subsets are each defined on a discrete real number line segment. If it is possible to construct a system of charts in such a way that each d–point of a d–space M is in at least one neighborhood, we call this system an **atlas**. A d–space M with an atlas A is called a **d–manifold**. Each d–map on a manifold M associates with each element of M an n–tuple of the space called the **coordinate** of the element under this d–map. A manifold can therefore also be understood as a d–set of d–points (N–tuples) where, for each d–point of the d–set there exists an open neighborhood which has a smooth one-to-one d–map onto an open d–set of R^N for some N.

A **coordinate d–basis** x^i parameterized on the generations t of an ordering operator O^i of a d–space S, is a basis such that the d–sets of the basis have no element (indeed, no d–point) in common (they are mutually disjoint), other than a uniquely and arbitrarily identified d–point called the **origin**.

Each d–space is characterized by a unique number n, which is the maximum number of disjoint d–subsets of S of equal cardinality, such that the union of the d–points formed from these disjoint d–subsets is indistinguishable from the d–space S. This number is called the **d–dimension** of S. For a coordinate d–basis of d–dimension n, we may refer to one of the disjoint d–subsets as a ***d–coordinate*** of S.

Having defined a vector d–space, we can assume the usual definition for linear combination, linear independence, maximal linear independent set (d–set), basis, dimension, components, metric functions, inner product, etc. We may also define the usual continuum notions, as long as we adhere strictly to the conditions for d–sets. We can now include eigenvalues and eigenvectors as usual, except that a range of eigenvectors is not allowed. Keep in mind that each recursive function may have multiple arguments. Thus, the d–vector at d–point P of M is not just a real number, except in the case when the number of arguments is one, and even then it may have a sign. It is truly a discrete vector with n components.

For any coordinate system x^i in an open neighborhood of a d–point on a vector d–space, the coordinates define a coordinate d–basis x^i (since there are n linearly independent d–vectors in the tangent d–space, these being the vectors formed from first derivates at the d–point of the underlying d–space vectors). Good coordinates are those where the x^i are linearly independent—this is just the condition on them to provide a 1–1 d–map to some neighborhood of the d–point in M onto a region in R^N.

Notice that we have nowhere restricted the definitions of the elements of a given d–sort or d–set: the structure is always extrinsic, thus supplying a local topology. This also means that we can use our constructs as fundamental elements in the definitions we have just given, thus generating a further layer of recursion. In this way, we will be able to define hierarchical structures.

An **inner product n()** is a recursive function on a vector d–space V, which satisfies the following if x is a d–vector of V and a is any (discrete) real number, and $|a|$ is the value independent of sign:

P8: $n(ax) = |a| \times n(x)$.

An inner product n is a **distance function** or *norm* in a vector d–space V, if it satisfies the following, where x and y are vectors of V and 0 is the inf of the d–set of all such vectors:

P9: $n(x) \geq 0$ and $n(x) = 0$ iff $x = 0$.

P10: $n(x + y) \leq n(x) + n(y)$.

Note that the relations of $+$ and $\times$ need not be the usual addition and multiplication. For example, $\times$ can be multiplication modulo 2.

By treating d–sets in R^N, we are in no difficulty, as long as we remember that defining the members of the image d–set must be recursive. Our recursion serves to maintain the class of elements in the d–set (insuring the existence of what is usually called, and which we will continue to refer to as, an **equivalence class**). In practice, one may use the standard notation and properties of continuum mathematics as a kind of shorthand (effectively making a Dedekind cut to obtain the real number line segment). If this is done, we must remind ourselves that in so doing, we have changed class membership (e.g., 1/2 is not in the same class as 0 and 1): we have effected a reparameterization. Since our definition of d–set is dependent upon the ordering operator which generates it, this means we must reconstruct any relationships between the d–set and any other d–sets.

In a sense, the d–set of d–points S in R^N is the union f of the image d–sets for all classes in the domain. Thus, we may define a distance function on S. We may also define a distance function on a d–set (a non–Hausdorff space), but it will be "multivalued" in the sense that the ordering between elements need not produce a single chain; thus, there may be more than one "path" between elements. If we take the distance function such that the number of elements traversed is minimal, then at best we must assume that elements in the string defining the distance (i.e.,

the minimal simple chain between any two elements) may in fact be twins under the equivalence class defined by the distance function.*

A **bilinear and symmetric** inner product n satisfies P11:

P11: $n(x + y)^2 + n(x - y)^2 = 2[n(x)]^2 + 2[n(y)]^2$.

If we say that a curve passes through a d–point P of a manifold M, it follows from the definitions that M is a recursively enumerable d–set and the curve is then a monotonic recursive function f on M. Thus, we say that the d–points of M (or objects of M) form an equivalence class ordered under f.

A derivate of f at P is then the **motion** along f at P (how fast is the ordering parameter increasing and in what direction + or −). Furthermore, for monotonic f, there are $n!$ distinct orderings of M without redundancy, if M has n objects and a sufficiently large number of reparameterizations for a given f (i.e., adding $1, 2, \ldots n$ to f).

If a is an element of a d–set A and b is an element of d–set B, then a and b are **comparable** if and only if, for some ordering operator O, $a \leq b$ or $b \leq a$.

We say that a **covers** b when the segment $[a, b]$ has two elements.

Theorem 7: If M consists of objects which are not in the same equivalence class, then we order them such that there exists an f for each class in M; then we may cover M by a suitable choice of

$$x^i(P_i)$$

with $f(x_1, \ldots x_i, \ldots x_m)$ for m classes in M. Thus, we can establish a basis for M.

There are, of course, many such bases. A coordinate basis is, then, one for which M is covered and the f_i do not order the same P_i (i.e., they are maximal and (linearly) independent).

A **d–fiber** consists of the d–set of all the derivates for all the possible parameterizations at a d–point of the base manifold M. A **projection d–map** assigns each d–fiber to a d–point of M.

A **product space** $M \bigotimes N$ consists of all ordered pairs (a, b), with a in M and b in N.

A **vector field** is a rule which chooses precisely one tangent d–vector from the tangent d–space at each d–point and assigns this to the point. For every vector field, there exists a curve, just as every curve has a tangent vector at each point. A d–set of curves which cover a manifold is called a **congruence**.

It will be useful to step outside the theory from time to time in order to understand the relationship between the ordering operator calculus and other mathematical endeavors. Certain computational phenomena in the practice of standard mathematics as applied to laboratory physics may be accounted for in this way. For this purpose, we introduce two special terms. If a recursive function were to be defined on

* Clearly, we do not care if the d–maps are strictly recursive; they may be analytic in the usual sense, as long as we keep in mind the constraints on the space.

an infinite set, then it would be said to be **analytic**. The **analytic interpolation** of a recursive function defined on a d–set over the segment $[m, n]$, is just an analytic recursive function for which there exists a d–map between the reparameterization of some discrete real number line segment $[m, n]$ of the recursive function and some monotonic sequence belonging to the infinite set generated by the analytic recursive function.

Theorem 8: A monotonic recursive function on a finite d–set of cardinality n has at most $n - 1$ derivates.

From time-to-time, we may say that some aspect of our construction is **global**, by which we mean that it is characteristic of or applicable to the entire d–space. Similarly, we may say that some aspect of our construction is **local**, if it is characteristic of or applicable only to some proper d–subspace.

d-Vector Functions

A **one-form** is a recursive function which generates a (discrete) real number for each d–vector on M and follows the usual linearity. The formation of this number is called the **contraction** of the one-form on the d–vector. A **metric** is a linear, symmetric function of two d–vectors (the "dot" product).

The recursive enumeration for a general d–set provides a parameterization for recursive functions defined on the d–set. Clearly, the parameter takes values from 1 to n over a d–subset of cardinality n. Note that the function deals, in general, not with cardinality but with ordinality, and this is arbitrary under the permutation group. The input and output are only symbols. Consider finite d–sets only. Interpretation as having cardinality n induces (via the function) an ordering on the d–set; thus, some structure is supplied. The function is intrinsically a mechanics of typography—how we can combine and use symbols is a recursive function.

In general, the notion of a recursive enumeration of a d–subset of the recursive d–set goes over to a parameterization under a d–map: $J \rightarrow R^1$. The parameterizations must cover the d–set. If they do so independently without repetition, then we have a **coordinate parameterization.**

Theorem 9: Exterior differentiation (defined as usual) commutes with any differentiable mapping of the manifold.

It is interesting to note that the cohomology groups depend only on the topological structure of M, and *not* on its differentiability. That is to say, cohomology theory passes over from standard differential geometry to the present theory almost intact, as there is no dependence on the definition of differentiability. This is particularly important for applications in physics, where Gauss' and Stokes' theorems are of such great use.

A Note on Computing Numeric Roots

It will often be the case that d-functions are needed, which make use of various roots, such as the square root. For us, not all numbers have a "square" root, meaning

two equivalent roots, and similar comments hold regarding higher-ordered roots, such as the "cube" root. Where such references are made in the remainder of this paper, we refer to the so-called "symmetric" root. Symmetric roots are defined as being a rational root of the number plus or minus some other rational number. Thus, in general, any square may be expressed as

$$(a - \epsilon) \times (a + \epsilon) = a^2 - \epsilon^2$$

such that ϵ is a rational fraction up to the precision of the computation. Note that this definition literally inverts the Pythagorean construction of the irrationals, but in a manner which requires no irrationals. This is, of course, just the operational definition which is taken in finite computation such as that using Newton's Method. One performs a recursive computation until the error (our ϵ) is sufficiently small. The nonequivalence of the two *root*'s in the ordering operator calculus holds a special significance: it suggests that the d–space is intrinsically noncommutative and that a commutative d–space is meaningful only if constructed of perfect squares, perfect cubes, etc.

2.5 Constructed Objects: Partition Lattices

In this section, we provide the concepts necessary to make the appropriate connections to generating functions and the finite operator calculus, as well as incidence algebras and von Neumann's theory of games. These concepts will prove useful when we begin the process of interpretation of physical phenomena.

An **order ideal** in a partially ordered d–set P is a d–subset Z of P, which has the property that if x is an element of Z and $y \leq x$, then y is an element of Z.

The **product** $P \bigotimes Q$ of two ordered d–sets is (p, q), where p is an element of P and q is an element of Q endowed with order $(p, q) \geq (r, s)$ whenever $p \geq r$ and $q \geq s$. The **direct sum** or ***disjoint union*** $P \bigoplus Q$ consists of elements x and y with order $x \leq y$ if and only if

(i) x, y are elements of P and $x \leq y$ in P

or

(ii) x, y are elements of Q and $x \leq y$ in Q.

The **blocks** of a partition of a d–set S are the d–subsets of S making up the partition. A partition $\prod$ is a **refinement** of a partition $\sum$ if every block of $\prod$ is contained in a block of $\sum$. The inf or 0 of $\prod(S)$ is the partition whose blocks are the one element subsets of S, and the sup or 1 of $\prod(S)$ is the partition with one block. The **lattice of partitions** $\prod(S)$ of a d–set S is the d–set of partitions of S ordered by refinement.

Note that there is a natural correspondence between equivalence relations on a d–set S and partitions of S, since the equivalence classes of an equivalence relation form the blocks of a partition and, hence, there is an induced lattice structure on the family of equivalence relations of S.

2.6 Constructed Objects: Combinatorial Systems

A **combinatorial system** consists of a unique initial label or *word* called the **axiom** of the system, and a finite d–set of strings called the **productions** of the system. Productions are the recursively generated words of a combinatorial system. The **alphabet** of the system is all the **symbols** or **letters** that occur in the axiom or productions of the system. A word of the system contains only the alphabet of the system.

Theorem 10: For every combinatorial system there exists a combinatorial system with precisely a two letter alphabet, whose decision problem is recursively solvable if and only if that for the first system is also recursively solvable.

Theorem 11: The d–set of integers generated by a combinatorial system is recursively enumerable.

Theorem 12: If the d–set of integers generated by the combinatorial system is not recursive, the decision problem for the combinatorial system is unsolvable.

Any n–form field divides all d–vector basis into two classes: those for which it is, on contraction, + and those −. This is called right- and left-handedness, respectively. If it is possible to be consistent in specifying handedness at each (not continuously here, since M is discrete) d–point P of the manifold, then M is said to be **orientable**. For every orientable manifold M, there exists an inverse to the derivate function called the **integral**.

3. MATHEMATICAL FOUNDATIONS II: ATTRIBUTE SPACE

By a **combinatorial attribute**, we mean a property of a d–sort that has been constructed by an ordering operator. In particular, for binary sequences or ensembles labeled by the generations of i ordering operators O_i, an attribute is any property which is recursively definable or computable on the ensembles of labels generated by the ordering operators. Let the sequence of labels output by each of the O_i be represented by a directed graph G_i (this is a Hasse diagram if there exist no cycles in the ordered labels produced by the ordering operator), consisting of labels as nodes and connecting arcs to represent the pairwise orderings between labels. Call the graph g_i which results from G_i by the removal of any number of arcs and/or nodes, a reduction of G_i. Consider the i collections of reductions R_i for each of the G_i. Form a new collection of reductions, consisting of no more than one reduction from each from of the R_i. If there exists an isomorphism between the all the pairs of reductions in such a collection, the reduction represents an attribute of the collection of ordering operators O_i.

Let the i ensembles of labels generated by the O_i be operated on by a new ordering operator O^*, so that the ensembles are (partially) ordered and labeled. Each generation of O^* can be classified according to whether or not there is a reduction corresponding to the underlying ensemble, which represents an instance of a particular attribute. We will call the instances of an attribute the **attribute states** of the attribute over the ordering operator O^*.

Since the output of any ordering operator O for each generation may be arbitrarily complex and have considerable internal structure (inaccessible unless constructed in the manner above), we will, henceforth, drop the notational distinction between O and O^*. The reader should, however, keep in mind the considerable structure which is implied when we refer to a combinatorial attribute or an attribute state.

A combinatorial attribute (or simply attribute where no confusion will result from the usage) is conceptually akin to a set-theoretic property, although mathematically distinct. First, attributes are constructive, whereas set-theoretic properties are not, being generally of an *a priori* nature and giving set theory that "tacked-on" look. Second, they are not "properties" of a set, but rather of a d–sort which has been constructed with an ordering operator. The definition of an attribute is thus much stronger than the definition of set-theoretic property, in that an attribute would certainly be a property but all set-theoretic properties are not attributes. Clearly, an ensemble has attributes as a set has properties, if one remembers that this similarity is metaphorical rather than mathematical.

As an example, consider the generation of the permutations of a discrete, finite, ordered collection S (noting that such a collection is a special case of a d–sort; in fact, it is a d–set). If the generator is specified via a recursive algorithm, has a unique starting ordering of the set and halts after generating all possible permutations generated in a specific order, the generator is a special (and useful) type of ordering operator. The notion of permutation so defined and used is an attribute of the ordered set P of all permutations of the original ordered collection S (i.e., it is possible to recursively define the permutations of a particular finite ordered collection and to recursively give their complete denotation), and any specific permutation P_i of the ordered collection S is an attribute state. Thus, a permutation is an attribute with respect to a reference ensemble (the starting ordering of S) and the ordering operator which generates permutations. Similarly, any specific subensemble is an attribute, with respect to the subensemble and **the identity ordering operator**—the ordering operator which, given a label as input, returns it as output. In this sense, we may search a d–set for d–subsets which are equivalent; i.e., those which have the same (identity) attribute.)

3.1 Multiple Ordering Operators

Please note that more than one attribute (indeed, more than one ordering operator which generates permutations) may be defined on the ordered set. This is an essential characteristic of d–sorts which must not be overlooked. For example, given an ordered d–set of labels of cardinality N, there are $N!$ additional distinct permutations possible. There are then $(N!-1)!$ possible ways in which the d–set of all permutations can be generated and, therefore, $(N!-1)!$ permutation attributes definable starting from the ordered d–set of labels.

In any given construction, we must explicitly state what ordering operators generate the structure, as these provide the connectivity of the elements. We may then construct the ways in which two or more operators combine or interact with each other. Suppose a d–sort of labels L of cardinality N are independently generated by i ordering operators O_i. In order to treat the d–sorts generated by O_i as a single

construction, it must be possible to demonstrate the constructive existence of a total ordering operator O' such that O' generates L. O' is said to be **decomposable** into the O_i if, to each generation of O', there corresponds one and only one O_i which generates the corresponding label. This is the first instance of the label being produced by this O_i, and all other O_i (except the O_i that generated the previous label of O' on its previous generation) generate the same label again. The O_i are said to be **serializable**.

Two or more ordering operators are said to be **intrinsically coupled**, if they generate at least one label or attribute state which is mutually indistinguishable (i.e., if the first operator cannot distinguish some label, (called the **coupling label**), generated by the second operator from a label which it generates and vice versa). Ordering operators which are serializable are **locally orthogonal**, since they are not intrinsically coupled. This does not mean that they are **globally orthogonal**, since they may be **extrinsically coupled** via a third ordering operator with which they are both intrinsically coupled.

The coupling of two ordering operators which are extrinsically coupled via a third ordering operator is said to be of **coupling degree** one; if via a third and a fourth, such that the output of the third is input to the fourth and the output of the fourth is indistinguishable from one of the labels of the first two, then the coupling is said to be of coupling degree two; the number of intervening ordering operators gives the degree of the coupling. The number of coupling labels gives the **coupling order**. Note that coupling is dependent on the specific ordering operators involved; two ordering operators may be coupled in multiple ways.

The coupling of two ordering operators O_1 and O_2 is characterized by a unique rational fraction called the **scale**, which is just the ratio of the cardinality N_1 of the labels, which may be produced by O_1 when coupled to O_2 to the cardinality N_2 of the labels, which may be produced by O_2 when coupled to O_1. The degree, order and scale, and the cardinalities and ordinalities of the ordering operators give all the information necessary to compute the probability (frequency) of one ordering operator interacting or **mixing** with another ordering operator. In a system of coupled ordering operators, the labels output by two ordering operators will be said to **superpose**.

We may relax the "set" restriction in our example: if a partial ordering is generated by an ordering operator, the collection (ensemble) is a d–sort with respect to that operator. In other words, distinguishability is meaningful only in terms of the ways (i.e., ordering operators) one has specified how to generate the ensemble. If the generator treats the order of "two" permutation states indifferently, then they are indistinguishable for that ordering operator, and we have no other means of determining distinguishability.*

* A further example may be helpful. In the d–space which is the positions on a chess board, the sequence of moves which any given chess piece takes during a game determines an ordering operator. The form or rule that specifies the legal moves that a piece can make specifies an attribute. If the ordering operator is parameterized on the attribute which defines a piece's legal moves, then each such legal position, generated in allowed sequence, is an attribute state.

It will be necessary to form quite complex attributes: attributes, then attributes of attributes, then attributes of attributes of attributes, etc. We will refer to these as **attributes of first order, second order, third order**, etc., respectively. From the definition of combinatorial attribute given above, it can be seen that the construction of attributes is potentially recursive. One can form a collection of ordering operators O_i^*, each of which generate the attribute states of an attribute; perhaps being distinct (although they need not be) in the order of generation of the attribute states. From the collection of outputs of the O_i^*, one may construct a new attribute and define a new ordering operator O^{**}, which generates the attribute states of this attribute of second order. This method of constructing higher-ordered attributes may be recursively continued, up to the point at which the only reduction possible is the graph consisting of a single node.†

3.2 Establishing a Distance Function

Based upon the definitions of ordering operator, dimension, and coordinate d–basis, a one–dimensional d–space coordinate basis behaves as a totally ordered d–set. It is convenient to represent this d–set by a sequence of binary strings; i.e., a string in an alphabet two symbols, where the order of the symbols is dictated by the ordering operator. For example, given a string composed of n unique labels, one may use Huffman encoding[21] as a way of unambiguously giving a binary representation of the sequence of labels.

Now, define **attribute distance** for a specific attribute generated by an ordering operator O_i, as the measure dependent solely upon the number of distinguishable states s between two ensembles of labels which O_i may generate, normalized by the total number of states which O_i may generate, N_i. This is equivalent to the unique number of generations of O_i required to generate the first ensemble A from the second (called the ***reference ensemble***) B, and results in a **distance function d()** on the closed interval of rational fractions [0,1]:

$$d(A, B) = s/N$$

By a reparameterization on an attribute (from the definition of reparameterization), we mean a mapping of the labels for the attribute states generated by some ordering operator O, to the tags for the attribute states generated by a second ordering operator O'. Via a reparameterization, then, we may remap $d()$ into $d'()$, defined on the closed interval of rational fractions [-1,1]. Since the cardinality of O may be

† Note that the mathematical objects of our construction are each defined relative to one or more operations. This forces an intrinsic connection between the usual static form of mathematics and the dynamics found in physics. It also precludes arbitrary identification of constructed entities. If our analysis of paradox[20] is complete, then many (we hope all) of the paradoxes which arise in logic, set theory and philosophy are not possible here.

smaller than that of O', indistinguishable attribute states for O may be induced in the mapping in order to properly map all the tags of O'[‡§] as the "maximum number [N—added for clarification] of distinguishable orientations between" two measured attribute values divided by the square root of N for normalization. This is just a measure of the distinguishability of two measurements, based on the number of values which an attribute being measured can take. Clearly, if the statistical distance between two values is zero, they are indistinguishable from an information theoretic point of view.

Wootters presents strong evidence that "statistical distance equals actual physical distance." The specific relationship is derived for the case of photon polarization measurements. We will demonstrate that this relationship is even more general than (apparently) assumed by Wootters and that the relationship serves as the basis for an extension of relativity and explains much in quantum mechanics.)

Suppose that we reparameterize $d(A, B)$. Represent a generation of O_i which decreases $d(A, B)$ as a 0 and one which increases $d(A, B)$ as a 1. The total number of 1's is simply the Hamming distance, and is defined on the interval $[0, N]$. By subtracting the number of 0's and then dividing by N, the result is a Hamming measure on the interval [-1,1] (i.e., centered on 0, which has an ordinal but not a cardinal significance), and is independent of the number of generations of O_i. In general, we will find this to be a more useful form of the attribute distance function.

3.3 Synchronization

In order to perform operations on multiple ensembles with some ordering operator O, some means must be given for establishing a label in each as the common input to O. If the ensembles are not identical, then a d–map between the ensembles will be useful. In this section, we define a particularly useful mechanism for achieving such a mapping.

Pick three ensembles, A, B and C. Let the attribute distance between A and C be zero, but with $t_C > t_A$ by some ordering operator O local to A and C, with generations parameterized by t. Let there be an ordering operator O'' local to B. Furthermore, let the attribute distance between A and B be nonzero. We say that the ordering

‡ In the chess example, the attribute distance in terms of the ordering operator which generates a specific piece's moves is just the number of moves that the piece has made, divided by the total number of moves it will make in the game. Then the distance is always some rational fraction of the total distance the piece will travel in the context of the game. Note, however, that having all the sequences of moves for all the pieces in a game does not allow us to reconstruct the game; we must know how the moves are interleaved. This can be accomplished by specifying the ordering operator in terms of the game clock; that is, for each move of the game, each piece's ordering operator must generate some attribute state. For us, whenever a piece does not move, the attribute state generated is indistinguishable from the previous attribute state. Reparameterizing the ordering operators in this way normalizes all the attribute distances, thus providing a global distance function topology on the d–space of the chess board.

§ Note the similarity between the notion of attribute distance and statistical (read with the frequency interpretation) distance, as defined by Wootters[22].

operators O and O'', with generations parameterized by t and t'', respectively, are **synchronized** if condition (1) holds, and A and B are said to be **synchronous** if the O and O'' are synchronized and conditions (2) and (3) hold:

1) $t_B - t_A = t''_C - t_B$;
2) if A is synchronous with B, then B is synchronous with A;
3) if A is synchronous with B and B is synchronous with C, then A is synchronous with C.

In other words, (1) states that the ordering operators are synchronized if there exists a binary symmetric relation between t and t'' over the specified attribute. By reason of the nonuniqueness principle (Principle IV), the property of synchrony between *ensembles* must also be reflexive and transitive as in (2) and (3), respectively. This simply means that it is possible to define a new distance function defined by T across the ensembles, which is consistent with the distance functions defined by t and t''.

Henceforth, we drop the notational difference between synchronous t and t'', since these may be replaced by a single universal ordering operator with parameter T.

3.4 The Dimensionality of D–Spaces

We are now in a position to construct a unique global property of d–spaces which have a distance function that is coordinate independent. We will begin by examining how such a coordinate independent distance function can be established, using the concepts we have defined and constructed. We will then investigate a global property of the resulting d–space.

Under the condition that the cardinality of the d–space or d–subspace precludes explicit representation of the algorithm for an ordering operator O, we are clearly faced with a severe lack of knowledge, the algorithm for O is arbitrary and the output may, therefore, may be treated as random for our purposes (Theorem 6).

Consider the labels produced by O to be represented by bit strings [i.e., strings of 1's and 0's] or, if repetition is being allowed and only two labels are allowed, to be arbitrarily treated as 1 and 0. The sequences of 1's and 0's thus produced meet the conditions of Bernoulli trials (see Figure 1): being unable to specify the algorithm used by O forces us to see successive productions of O as independent. Only after the fact, may we label the resulting output of O as representing some specific, previously known, ordering relation.

Now, let there be r such arbitrary binary number generators, $O_1, O_2, \ldots O_r$, with string (d–set) outputs $S_1(n), S_2(n), \ldots S_r(n)$ up to some maximum, R. We will refer to n as the ordinality of each string, parameterized by a counter we will call t. In the absence of other information about $O_1, O_2, \ldots O_r$, we assume them to be independent operators and, indeed, cannot discover otherwise, due to the computational complexity of the operators and our relatively impoverished (representationally) system of

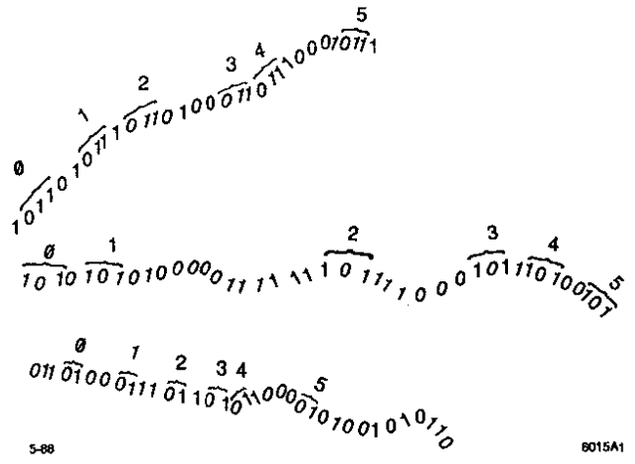


Figure 1 Independent trials.

just two symbols.* Clearly, $S_1, S_2, \ldots S_r$ are synchronized on n via t, but this provides only trivial structural information about any relationships between the strings, given our criteria for "arbitrary" and the independence of $O_1, O_2, \ldots O_r$. $S_1, S_2, \ldots S_r$ constitute a coordinate d–basis, as long as we identify the initial outputs of $O_1, O_2, \ldots O_r$ as "identical."

Look for other means of synchronizing $S_1, S_2, \ldots S_r$; pick as an attribute any sequence of length M of binary symbols (i.e., a specific substring of $S_1, S_2, \ldots S_r$); interpret the first M symbols of $S_1, S_2, \ldots S_r$ as matching this substring; then examine the output of $O_1, O_2, \ldots O_r$ for further synchronized productions of this string. As long as the algorithm governing the ordering operator which generates the strings is of sufficiently great computational complexity, the occurrence of the substrings is arbitrary and simple statistics for concurrent, independent Bernoulli trials apply.

At $L = Mn$, (i.e., the position of the $n+1^{th}$ string of length M), the probability that the number of occurrences of the substring is the same across multiple strings produced in this way is just

$$u_L = \frac{1}{2^{rL}} \sum_{k=0}^{L} \binom{L}{k}^r \tag{2}$$

where $r \in \{1, 2, \ldots R\}$.

* Proving nonindependence would require a knowledge of the ordering operator's algorithm.

Now, interpret the normalized number of occurrences of the substring as defining a discrete distance function across the R–dimensional d–space. That the normalized number of occurrences of the substring constitute an attribute distance, and, therefore, a distance function, is trivial; the identity attribute for the given substring may be satisfied (the substring may occur) only finitely many times in a string of finite length. This defines the number of possible instances of the attribute. Take as a reference ensemble, the null string (the string of zero length, i.e., the empty string). The attribute distance follows from the definition, immediately.

Call a specific generation of an r–dimensional d–space with any distance function f, a **reference frame**. We see that Eq. (2) is just the probability that the distance function has a r–independent value, given a value of R and L (i.e., the distance function is length preserving).

Theorem 13: The upper bound on the global d–dimensionality of a d–space of cardinality N with a discrete, finite and homogeneous distance function is 3 for sufficiently large N.

Argument:

Note that for $R > 3$, the terms of Eq. (2) are monotonic decreasing (i.e., Eq. (2) converges). That is, for sufficiently long strings, the probability of another synchronized occurrence of the specified substring must approach zero. For $R \leq 3$, however, the terms of Eq. (2) are typically monotonic increasing [i.e., Eq. (2) diverges]—there is always the probability of another occurrence of the specified substring. Hence, the possibility of an isotropic (Principle IV) distance function across more than three dimensions is unlikely, while it is certain if $R \leq 3$[23]. Clearly, the case of $R = 3$ contains the greatest representational power.*

Formal Argument:

Two or more attributes are mutually independent if the generating or defining ordering operators on the space of ensembles are mutually disjoint, with the exception of a single element; in which case, they define the dimensions of the space. An attribute admits a distance function if and only if there exists a total ordering on the ensembles which possess the attribute. Two or more (R) such distance functions are symmetric (do not introduce inhomogeneities into the space—Principle IV), if and only if they can be synchronized (this is equivalent to demanding that there is exist an R–way matching criteria between the productions of the R ordering operators,

* The concentration of synchronizable events for "short" attribute distances, even with $R \gg 3$ will be related to the big bang and to quantum fluctuations.

such that a match is found arbitrarily often in sufficiently long productions).†

However, this is not possible if $R > 3$. Let the productions of the ordering operators be mapped to binary strings. These strings may be treated as the results of Bernoulli trials. The probability of a specific occurrence at the n^{th} trial is given by

$$u(n) = \frac{1}{2^{rn}} \left[\binom{n}{0}^r + \binom{n}{1}^r + \binom{n}{2}^r + \cdots + \binom{n}{n}^r \right] \tag{3}$$

for n trials, with $r = R$ equal to the number of dimensional metrics. The maximal term of the binomial distribution

$$\binom{n}{k} 2^{-n} \tag{4}$$

is of the order $\sqrt{2/\pi n}$) and $< n^{-1/2}$. Therefore,

$$u(n) < n^{-(r-1)/2} 2^{-n} \left[\binom{n}{0} + \binom{n}{1} + ... + \binom{n}{n} \right] = n^{-(r-1)/2} \tag{5}$$

and so the sum of $u(n)$ converges for $R > 3$.

That it diverges otherwise is proved as follows:

Case 1: For $R = 2$, and from the normal approximation to the binomial distribution

$$u(n) = \binom{2n}{n} 2^{-2n} \simeq \frac{1}{(n\pi)^{1/2}}$$

and so $\sum u(n)$ diverges for $R = 2$. Note, however, that $u(n) \to 0$ as $n \to$ large N. Therefore, while synchronization is certain, it has a mean recurrence time on the order of $\sqrt{N}$, so that, in two dimensions, the synchronization is sparse.

† More importantly, note that these conditions are identical to those demanded by Einstein in deriving the Lorentz transformations. Namely, the demand that clocks be synchronizable is equivalent to demanding spatial homogeneity (i.e., that there is no preferred coordinate). In addition, the property of transitivity (from the definition of synchronization) implies that there exists, at least mathematically, a "universal" clock. This is ironic, in as much as special relativity is usually understood to have removed the Newtonian concept of Universal Time. In fact, Einstein did not remove the concept, but rather showed that this global time need not be accessible, as long as synchronization with transitivity was allowed. Under these conditions, local time is sufficient.

Case 2: For $R = 3$, and from the normal approximation to the binomial distribution, for sufficiently large n and

$$\frac{1}{2} n - n^{1/2} \leq k \leq \frac{1}{2} n + n^{1/2} ,$$

we have

$$\binom{n}{k} 2^{-n} > cn^{-1/2} ,$$

where c is a (small) positive constant. Therefore,

$$u(n) > 2n^{1/2} (c^3 n^{-3/2}) = \frac{2c^3}{n} ,$$

and so $\sum u(n)$ diverges for $R = 3$.

Thus, as a recurrent event, any given sequence will be shared between more than three runs only a finite number of times, and, hence, is unlikely; whereas, between three or fewer, the same sequence will be shared in position arbitrarily often for sufficiently large strings.

It follows immediately, that we cannot define a metric-homogeneous discrete space with more than three spatial dimensions. Any other space must introduce either asymmetries or inhomogeneities over the metric.

QED

It might be argued[24] that in a d-space of finite cardinality N, the theorem no longer applies. However, consider what happens if the global distance function is defined with $R = 4$; then there exist local distance functions defined on the three-dimensional d-subspaces.

Suppose that some relationship is to be defined between the local distance function and the global distance function. For large, finite N, this becomes impossible. A comparison of $u(n)$ for each distance function shows that, as N becomes large, "meter marks" for the three-dimensional d-subspace become relatively more frequent, whereas those for the four-dimensional space become less frequent. Thus, the d-map becomes impossible, unless the one-dimensional d-subspace grows more rapidly than the three-dimensional d-subspace; i.e., unless one dimension is different from the remaining three. However, by hypothesis and in keeping with Principle IV, this is not possible, since it makes the four-dimensional d-space inhomogenous.

Furthermore, there would then be a three-dimensional d-subspace, composed of the one-dimensional d-subspace and any two other dimensions, which would generate as rapidly as the four-dimensional d-space, and the difficulty of defining a relationship between the distance function on this d-subspace and the global distance function would be undiminished. Notice that this difficulty becomes apparent for relatively small runs (as soon as $n^{1/2}$ is significant), since the ratio of expectations for synchronization between a d-space and its largest d-subspace is bounded by $n^{-1/2}$.

Constructing a Coordinate System

It is important to understand how one constructs a coordinate system using Theorem 13 and the definitions that preceded it. We make explicit use of the notion of independence in order to construct an orthogonal basis, since independence is the essential constructive notion underlying orthogonality when a geometry (i.e., some notion of "angle") does not, as yet, exist. Having taken this step, we are then *required* to construct a norm which vanishes when the two arguments are orthogonal. This is, of course, trivial if the usual operations of addition and multiplication are available; but care must be taken, since we deny the need for the usual properties of closure and commutativity.

Having once identified an attribute *independently* in each of three binary strings generated as in the discussion preceding Theorem 13, computed the distance from an *arbitrarily identified* origin using the appropriate one of three distance functions (each need only be defined on one of the strings), and, finally, established synchronization across the three strings, the only quantities of interest in performing d–vector computations in this three–dimensional d–space are the "meter marks" established by synchronization. This synchronization establishes a new distance function uniquely defined in the ordering operator sense, which is independent of which of the three strings are involved in the computation.

Thus, if we now treat the three strings as generating a coordinate d–basis x, y and z, we may say that a d–vector of a certain "magnitude" has a particular "direction." In the simplest case, the direction is either "parallel" or "antiparallel" to x, y or z. In such a case, the norm which we use must give the magnitude of the d–vector, when the arguments to the norm are the d–vector and the appropriate unit d–vector in the "same" direction and the infimum of the distance function, if either of the other two unit d–vectors are used. This proscription on the construction of a norm results in a unique norm only in that all such norms will be "orthogonality" preserving.

Similar comments hold with regard to the construction of a "vector product." Great care must be taken not to assume any intrinsic notion of direction and connectivity of the d–space, such as that which is often imposed by the Pythagorean theorem (which is valid only in a Euclidean or flat–space and generally not valid in a discrete, finite space). Furthermore, great confusion and apparent contradictions result if one insists on using the distance function defined, in order to construct "meter marks" on a particular string as though it were global (i.e., useful for all three strings or identifiable with the distance functions defined by the process of synchronization).

Note that if the d–vectors are represented by binary strings, it is necessary that the independent attributes be represented consistently; thus, the attributes must be independent under the operation of discrimination (exclusive or). If the d–vectors in a d–space are represented in a manner consistent with "meter marks," we then have a means of forming the vector product. In the canonical form, the independent attribute substrings for a three dimensional d–space are just "001" (x), "010" (y) and "100" (z). These substrings form a complete representation and are independent under the operation of discrimination.

A d–vector represented by the binary string "001011110," then has an attribute distance in the x direction of 2, in the y direction of 2 and in the z direction of 1. Such a representation gives more information than just the direction and magnitudes—it contains a history of the generation of the d–vector. This explicit representation of the process nature of mathematical objects is an important characteristic of the ordering operator calculus. In order to use the usual notions of a vector space, including computation of components, this historical information must be obscured. Thus, one only considers the magnitudes and the directions, without the explicit binary representation of the d–vectors.

3.5 Characteristics of a Discrete Geometry

Having developed a d–space with a coordinate independent distance function, we may now explore certain other symmetry relations on the d–space. In particular, we will find it useful to understand the d–space equivalents of the familiar orthogonal and rotational symmetries.

It is a central point of this section that a measure of the discrete cardinality N and of the curvature of a discrete geometry in a d–space is given by the precision with which two ratios are identical in value: the ratio of the area of the maximally-sided symmetric polygon, which may be constructed in the d–space to the area of a square ($\pi(N)_{areas}$) in the d–space, and the ratio of the perimeter of that same polygon to the perimeter of the square ($\pi(N)_{perimeters}$); see Figure 2.

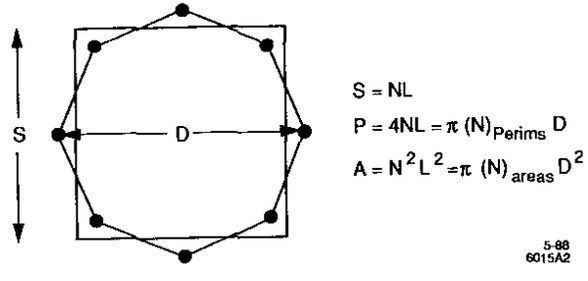


Figure 2 Relation between $\pi(N)_{areas}$ and $\pi(N)_{parameters}$.

Indeed, the relationship between these values has global significance, and we shall have need of understanding that significance in later sections, as well as being able to explicitly use one or other of the ratios thus constructed.

In what follows we construct a square and a circle, and construct an algorithm for a rational fraction ratio which plays the role of π. We begin by constructing the equivalent of a square: an orthogonal, two–dimensional coordinate patch. The only elements allowed for construction are a finite (perhaps large) number of discrete elements (essentially indistinguishable mathematical objects), ordering operators, the ability to count and the ability to label the objects through an operator.

By **nearest n neighbor** of a label e in a sequence of generations of an ordering operator O, is meant any label n such that for labels a, b, e and n, $a : a = O(e)$ (read "label a such that a is generated on input of label e to ordering operator O') or $b : e = O(b)$; then for any ordering operator O' mutually disjoint from an ordering operator O, at least one of n=O'(e), n=O'(a), n=O'(b), e=O'(n), a=O'(n) or b=O'(n) holds. Clearly, a and b are nearest neighbors of e, as well. The process of identifying nearest neighbors simply defines a new binary ordering relation between a pair n and e, if there exists a third label a for which (possibly distinct) binary ordering relations exist between a, e and n.

Before proceeding with the constructions, a comment on notation: the ordering operators used will be total ordering operators. Those differing only in a subscript will denote a d–set of mutually disjoint ordering operators having domains of equal cardinality (and ordinality by definition). The symbols O and O' will be used for ordering operators whose output is to be taken as orthogonal: that is, the generations are mutually disjoint (distinguishable) except for a single generation of each which are indistinguishable. The generations of the O and O' will be notationally distinguished by x and y, respectively. A prefixed superscript of either 1 or -1 will denote the generations of the ordering operator as coming either before or after some specified and unique label, respectively. The subscripts associated with O and O' will be carried over to the respective generations.

A Discrete Coordinate Patch

Without reference to a particular geometry or distance function, a "square" can be defined as a closed d–set having the following properties:

a) two–dimensionality;

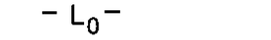
b) the edges or boundary consists of two d–sets of two mutually disjoint totally ordered d–subsets (four sides);

c) fixed center under interchange of the coordinate parameters;

d) it is possible to establish a distance function on the edges such that each of the totally ordered d–subsets is of equal length.

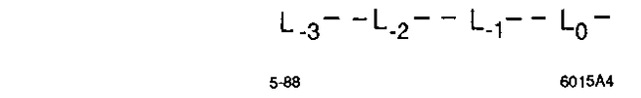
The criteria for two–dimensionality is satisfied by requiring two mutually disjoint ordering operators. The algorithm is as follows:

1) Select a label L_0; see Figure 3.



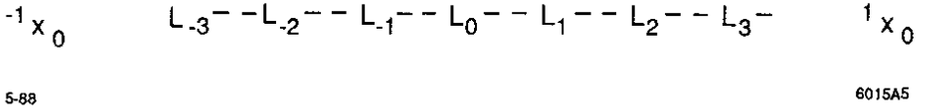
5-88 6015A3

Figure 3 A starting label L_0.



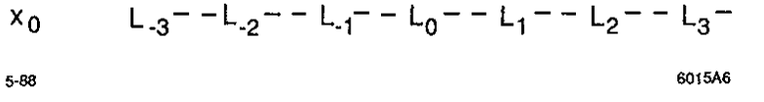
5-88 6015A4

Figure 4 The subchain of length $n = 4$, $^{-1}x_0$ with L_0 as supremum.



5-88 6015A5

Figure 5 The subchain of length $n = 4$, $^{1}x_0$, with L_0 as infimum added.



5-88 6015A6

Figure 6 The chain length of $n = 7$,x_0.

2) Establish a totally ordered d–set chain $^{-1}x_0$ of length n with L_0 as the supremum, using the ordering operator O_x; see Figure 4.

3) Establish a chain $^{1}x_0$ of length n, with L_0 as the infimum, using the ordering operator O_x; see Figure 5.

4) Call the union of $^{-1}x_0$ and $^{1}x_0$: x_0. Require that x_0 be totally ordered; see Figure 6.

5) For each label L_i of x_0, establish chains $^{-1}y_i$ and $^{1}y_i$ of length n, under the ordering operator O_y, with the selected label of x_0, as either the supremum and infimum of the chain. Require that the y_i are disjoint, as are the pairs $(^{-1}y_i, ^{1}y_i)$. This is a unique labeling or total ordering requirement on the entire construction (i.e., there must exist an ordering operator O'' such that the labels of the entire construction are totally ordered; see Figure 7.

6) Require that the n^{th} label of the y_i form chains x_i, ordered by ordering operator O_{x_i}; see Figure 8.

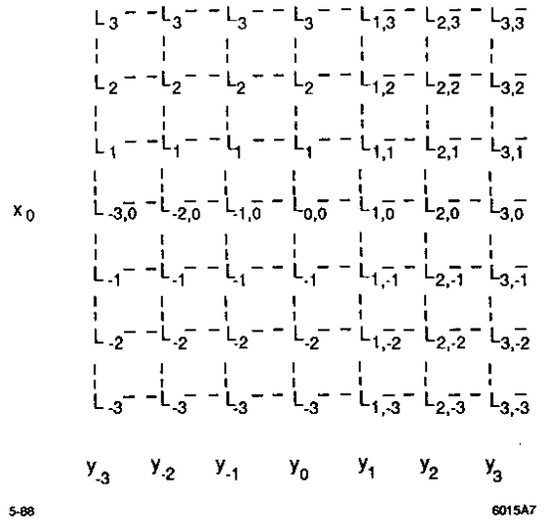


Figure 7 The chains y_i of length $n = 7$ added.

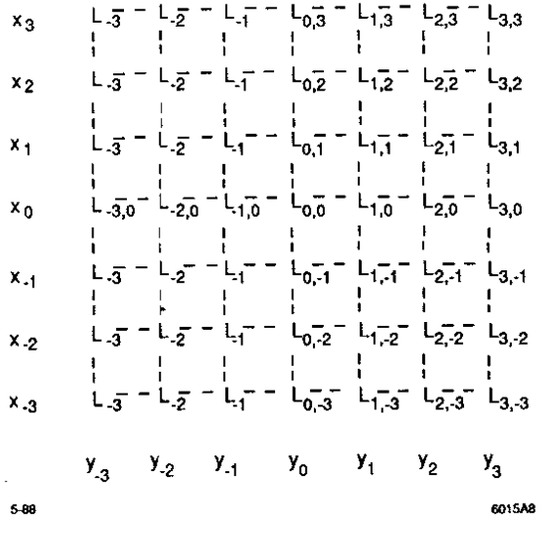


Figure 8 The chains x_i of length $n = 7$ added. Note that all labels are now subscripted twice, since they are identified as the production of two ordering operators.

7) The resulting object satisfies the requirements; it is the discretum version of a two-dimensional (square) coordinate patch. In particular, the two-dimensionality of the construction is satisfied by the definition of mutually disjoint ordering operators: at most, one label in a chain resulting from one operator will be found in a chain resulting from the other. For the given construction: at most, two operators can be used; a third would result in a partial ordering, instead of a total ordering, of the labels of the construction, and this would then represent an object which is not connected or result in an object for which "multiple" labels are doubly labeled. Thus, the ordering operators "parameterize" the object.

A Discrete Circular Patch

We can now proceed to construct an object which behaves as a discretum version of the 2-sphere. A 2-sphere (again, without reference to distance functions) has the following properties:

a) two-dimensionality;

b) every edge (boundary) label is indistinguishable from every other, under interchange of the corresponding ordering operators;

c) existence of a unique label, which remains fixed in the construction, under interchange of any two ordering operators which generate it.

The constructive algorithm is as follows.

1) Select a (square) coordinate patch with center $L_{0,0}$ and all labels uniquely subscripted. Call this patch S_0; see Figure 9.

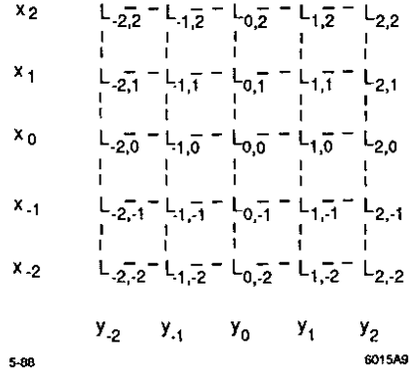


Figure 9 Select a patch, s_0.

2) Constrain the possible ordering operators (as before) to those operators which produce chains of length n and which generate from $L_{0,0}$ a nearest neighbor of $L_{0,0}$,

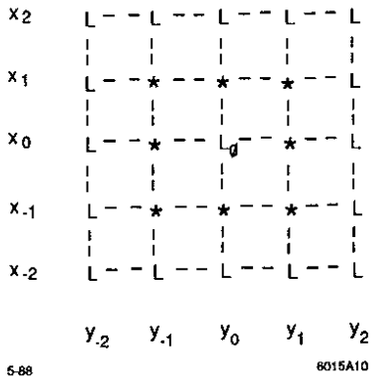


Figure 10 The nearest neighbors of L_0 are shown as 'an asterisk (*).

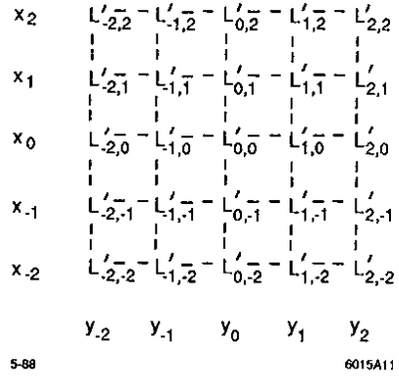


Figure 11 A new patch, P_i.

then a nearest neighbor of this label, and so on. We refer to the operators which generate these labels as **radial permutations** of the coordinate patch; see Figure 10.

3) Starting from $L_{0,0}$, construct a coordinate patch with a new pair of ordering operators which are radial permutations of the coordinate patch; see Figure 11.

4) Map the labels of this patch S_i to patch S_0, and eliminate any labels which do not have at least i subscripts; see Figure 12.

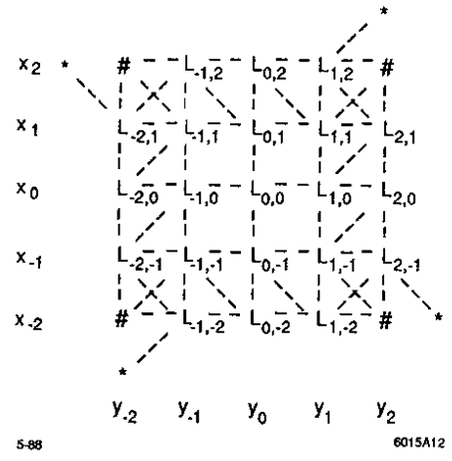


Figure 12 Mapping the new patch to the old.

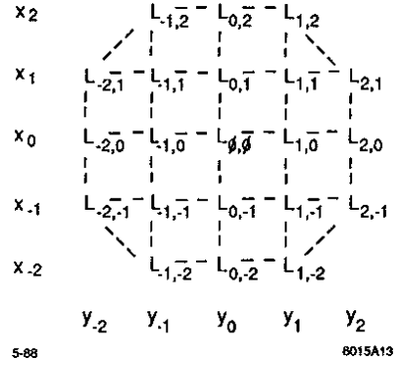


Figure 13 Elements remaining after all allowed radial permutations.

5) Repeat this process for all pairs of allowed radial permutations; see Figure 13.

The result is a discretum version of the circle, in that it has a fixed center ($L_{0,0}$) with radial symmetry (isomorphic to its radial permutations with identified center $L_{0,0}$). It has built-in bounds on "precision." The relation between the number of "sides" of the polygon formed by a set of cardinality n and the number of permutations is fixed: it gives a measure of the "size" of the circle.

Note that in Figures 3–13, the radial permutations which were not invoked would result in either the same labels of the construction being deleted, as here, or else would not maximally d–map the coordinant patch. The reader may readily demonstrate this. Also note that starting with a central label is a matter of technical convenience for the algorithms and may be circumvented.

$\pi(N)$

Given these two geometric objects, it is possible to define a ratio which plays the role of the ratio of the area of the circle to the area of the square patch from which it was formed. This number is obtained by counting the number of labels contained in the circle and the number of labels contained in the square and forming the ratio.

A second ratio is obtained from the ratio of the cardinality of the d–set of all radial permutations (obtainable by counting the labels on the perimeter of the circle) and the cardinality of the generations of one such radial permutation (e.g., the length n of the chain x_0).

In general, these ratios will be functions of the length n of the chain x_0. Furthermore, the values of the ratios will not, in general, be those obtained under Euclidean geometry. However, if one insists on isotropy, homogeneity and "density" (i.e., large n), it is easy to see that these values must be those obtained by the standard polygonal approximation to the circle. In particular, these ratios will be approximations to $\pi/4$ and π, with the appropriate precision. These constructions, and the results, are closely related to numerical and statistical "approximation" methods, as seen from within the traditional geometric paradigm. In fact, Archimedes came close to the construction used here (Measurement of the Circle). However, the definitions used here are completely constructive and general, matching the continuum definitions (which we prefer to think of as the "analytic interpolation") as desired.

CALCULATIONS

By Areas:	By Perimeters:
A (square patch) = 25	C (polygon) = 12
A (polygon) = 21	C (square patch) = 16
Ratio = π (areas)/4	Ratio = π (lengths)/4
= A (polygon)/A (square)	= C (polygon)/C (square patch)
= 21/25 = 0.84	= 12/16 = 0.75
π (area) = 3.36	π (perimeter) = 3.00

Indeed, if the cardinality of the d–space (N) is changing (evolving), then the two values of $\pi(N)$ will be changing, also. Furthermore, if the relevant discrete cardinality is related to a spatial volume, then, as this region becomes smaller, calculations

involving $\pi(n)$ cannot be treated in a naive manner. Specifically, the multiple computational definitions of $\pi(n)$ must be disassociated, if the values are different (i.e., the ratio $\pi_{areas}/\pi_{circumference}$ will not be 1). The value of *pi* can no longer be taken as a constant, independent of spatial volume. Indeed, if the d–space is inhomogenous, the value will depend on the local inhomogeneities; it will have different values depending on the "density" of the local d–space. Even more important, if the d–space is discrete and finite, and if the values of $\pi(n)$ are not related to the local spatial volume via a cardinality of the local volume, it follows that the values of $\pi(n)$ used in calculations relate only to the cardinality of the d–space. In other words, $\pi(N)$ becomes a true global discrete topological constant, and local physical properties are then immediately dependent on the global properties.*

In the remainder of this paper we will use $\pi(N)$ to refer to the combinatoric computation of π, based on the ratios of perimeters for a d–space of cardinality N. We cannot use the ratio computed from area ratios, since we will not, in general, know the "curvature" of the d–space. Note that measuring the difference between the two ratios gives a means of locally measuring the flatness of the d–space. Similarly, the curvature can be measured by examining the ratio computed on the basis of "volumes."

Radian and Trigonometric Measures

Having constructed the largest coordinate patch and the corresponding inscribed "circle," we may now pick an orientation and specify a total ordering operator which generates the sequence of attribute states constituting the perimeter as labels. We then reparameterize the generations of this ordering operator into the interval of rational fractions $[0, 2\pi(N)]$. We call this parameterization the **radian measure** on a d–space of cardinality N. Similarly, we shall refer to the cardinality or length of the total ordering generated by one of the radial permutations used in constructed any circle, the **radius** r of the circle.

A radius r and a radian measure θ then correspond to that d–point which results from a translation in coordinate distance of attribute distance r from the origin, followed by θ generations of the reparameterized perimeter ordering operator. Since every pair r and θ correspond to a unique point on the perimeter, and $\pi(N)$ is constructed from the maximal coordinate d-patch, we may regard θ as a direction and define the trigonometric computations of θ in the usual manner using the norm function. In particular, take the cosine to be the unit normal projection on the x-axis and the sine to be the unit normal projection on the y-axis. Note that this does not assume the Pythagorean theorem, unless it is already entailed in the norm function.

3.6 Properties of Evolving Systems: Attribute Velocities

Given a d–space, we require that there exist a total ordering operator on the space, so that a distance function (such as that produced by the Program Universe ordering operator) is possible. The universal ordering parameter T, on which the generation of this ordering operator is based, provides a local total ordering for the

* Applying this fact to physical phenomena, that π should then be of cosmological (global) significance is not surprising. Consider these results where the d–space is the physical Universe.

evolution of each ensemble, such that the local total orderings are isomorphic up to reparameterization. This in turn provides for synchrony.

We now define the **increment** I of an ensemble as the number of generations of some ordering operator† O needed to describe (establish local isomorphism with) the increases in attribute distance between an ensemble and a reference ensemble, with respect to T. This operator parameterizes the generation of the attribute states. Similarly, we define the **decrement** D of an ensemble as the number of generations t of the ordering operator O needed to describe the decreases in attribute distance between an ensemble and a reference ensemble, with respect to T. The **total size** S of an ensemble is defined as the arithmetic sum $+$ of the I and D. Use $[I, D]$ to denote an ensemble with increment I and decrement D and total size $I + D$. Note that the total size S is not generally the same as the maximum cardinality N since total size refers to increments and decrements of the ordering operator, and not to the cardinality of the d–sort of labels produced by the operator.

Attribute velocity v is defined as the mathematical rate of change in attribute distance of an ensemble, with respect to generations t of an ordering operator O, computed as the difference between I and D, divided by the total size S:

$$v = \frac{I - D}{S} \; . \tag{6}$$

The relative attribute velocity v' is just v computed relative to a third ensemble (reference), having attribute velocity u. The relative attribute velocity may be regarded as a discrete map which transforms an ensemble $[I, D]$ into an ensemble $[I', D']$, where I' and D' depend only on I, D and u, and where v' depends only on u and v. This is just a change in the reference ensemble. The **increment quotient** is defined as the ratio of I' to I,

$$q = \frac{I'}{I} \; . \tag{7}$$

The **attribute speed** of an ensemble is the magnitude of the attribute velocity (note that direction is given by arithmetic sign or the degenerate cosine in the one–dimensional case, a discrete version of the $x^1 - x^2$ cosine in the two–dimensional case, and a discrete version of the $x^1 - x^2$ and $x^2 - x^3$ cosines in the three–dimensional case). Finally, we define independent ensembles as those having all states generated with respect to an ordering operator O, distinguishable. We will discuss the impact of indistinguishable states in a later section. Having defined these terms, we may now prove a series of theorems regarding the properties of such ensembles.

Theorem 14: The increment and decrement are additive for independent ensembles when aggregated; that is, the number of distinguishable states and the number of

† In general, this is not the same ordering operator which generated the ensemble.

generations t relative to an ordering operator O required to describe I and D for independent ensembles is conserved.

$$[I, D] + [I', D'] = [I + I', D + D'] \tag{8} \; .$$

Argument:

As long as two states of an attribute are distinguishable over t, we are certain that a generation of O is required for each. It follows that the total number of generations T for independent ensembles (those having all states distinguishable) is given by the arithmetic sum (total count) of the generations of O, for the increment and decrement of each. Indeed, the total size of the ensemble is just $S + S'$.

QED

Theorem 15: The attribute velocity v of an ensemble $[I,\ D]$ is a function of I and D, and nothing else.

Argument:

If ensembles A and B have the same attribute speed, then the aggregate ensemble $A + B$ must also have that attribute speed. Hence, v cannot depend on total size, but only on the ratio of I to D. Let $r = I/D$; then we can write

$$v = v(r) \; . \tag{9}$$

QED

Theorem 16: v is an increasing function of the ratio r.

Argument:

Trivially, the case from the definitions.

QED

Theorem 17: If the values of I and D are reversed, then v is reversed;

$$v([D, I]) = -v([I, D]) \; . \tag{10}$$

Argument:

Inverting I and D is equivalent to counting distinguishable states from above, as compared to from below—i.e., if one counts from 0 to the maximum number of distinguishable states, one obtains the usual definition of I and D. If one counts from the maximum number of distinguishable states down to 0, consistency with the definition of additivity can be maintained if this is equivalent to a reparameterization resulting in a change of arithmetic sign.

QED

Theorem 18: If neither I nor D is 0,

$$v\left(\frac{1}{r}\right) = -v(r) \ . \tag{11}$$

Argument:

Trivially, from Theorem 4 and the supposition.

QED

Theorem 19: The attribute distance between any two ensembles has an upper and lower bound.

Argument:

Trivially, from the finitary principle (Principle I).

QED

Theorem 20: The lower bound of v CANNOT BE ZERO for independent (i.e., distinguishable) ensembles.

Argument:

If the lower bound of v were zero, the ensembles would be attribute indistinguishable and hence not independent.

QED

Theorem 21:
There is a limit to v as D approaches 0, which we can take as 1 by appropriate reparameterization; i.e., $v([I,0]) = 1$ and, hence, $v([0,D]) = -1$. We shall refer to this upper bound as v_{max}.

Theorem 22: Etters Velocity Relationship,

$$v(r) = \frac{(r-1)}{(r+1)} \tag{12}$$

holds for attribute velocities.

Proof:

Consider a d-space with distance function as previously defined. Now, examine the region between synchronization (metric marks or ticks). In this region, as we have shown, there exists a value for the isotropic distance function. Let I be the total number of 0's and D the total number of 1's generated up to n generations of the ordering operator which defines the distance function (called the **metric ordering operator**); then the total attribute "displacement" in $I+D$ generations is just $I-D$. This gives an Etters velocity relationship of $I-D/I+D$ or, if $r = I/D$,

$$v(r) = \frac{(r-1)}{(r+1)} \ .$$

QED

4. MATHEMATICAL FOUNDATIONS III: COORDINATE TRANSFORMATIONS

In order to explore the invariant properties of a system, we must have a means of expressing not only the coordinate bases defined in the previous chapter, but also transformations between coordinate bases. Of particular interest are those coordinate bases which define a reference frame. In the present chapter, we develop a series of theorems regarding transformations between reference frames.

Theorem 23: Suppose that synchronizable reference frames K, with coordinate bases x^i and k with coordinate bases y^i, i in $\{1,2,3\}$, are defined so that the origin of k has attribute velocity v in the direction x^1, with respect to the origin of K in the universal ordering parameter T; then the coordinates transform according to:

$$t' = \gamma \, \frac{t - vx^1}{v_{\max}^2} \ , \tag{13}$$

$$y^1 = \gamma \, (x^1 - vt) \ , \tag{14}$$

$$y^2 = x^2 \ , \tag{15}$$

and

$$y^3 = x^3 \ , \tag{16}$$

where

$$\gamma = \frac{1}{\left[\frac{1-v^2}{v_{\max}^2}\right]^{1/2}} \ .$$

Argument:

Select A, B and C synchronous with a distance function $d()$. Let $d(A,B)$ be the attribute distance between A and B and $d(B,C)$ be the attribute distance between B and C. Given $d(A,C) = 0$, as above; then, by symmetry (Principle IV), we require that $d(A,B) = d(B,C)$, so that for maximum attribute velocity $v_{\max}$, we have

$$v_{\max} = \frac{2d(A,B)}{t(A) - t(C)} \ . \tag{17}$$

Since $d(A,C) = 0$, note that A and C are indistinguishable, except by parameter t. Furthermore, with reference to a third ensemble with attribute velocity v,

$$t(B) - t(A) = \frac{d(A,B)}{[v_{\max} - v]} \qquad \text{and} \qquad t(C) - t(B) = \frac{d(A,B)}{[v_{\max} + v]} \ . \tag{18}$$

Now, suppose that we wish to compare the attribute distances d and d' and the operators t and t', with reference to third and fourth ensembles with attribute velocities 0 and v, respectively. Call these systems K and k. Furthermore, assume that

there exist at least two independent attribute distances (generated from mutually disjoint ordering operators, except for a single element) for K and k; call these x^i and y^i, respectively. We seek one-to-one transformations (discrete maps) between these operator values. Given (in the absence of specific cause—i.e., an ordering operator) homogeneity (Principle IV) of the system K and k in the parameters, these transformations must be linear and homogenous.

Let $x^{1'} = x^1 - vt$; then k has a system of values $x^{1'}$ independent of t. Define t' as a function of $x^{1'}$, x^2, t. Let $d'(A,B)$ be the attribute distance between A and B, and $d'(B,C)$ be the attribute distance between B and C. Given $d'(A,C) = 0$ as above; then, by symmetry, we require that $d'(A,B) = d'(B,C)$, and

$$t'(B) = \frac{1}{2}[t'(A) - t'(B)] \ ,$$

or

$$\frac{1}{2}\left[t'(0,t) + t'\left(0, t + \frac{x^{1'}}{v_{\max} - v} + \frac{x^{1'}}{v_{\max} + v}\right)\right] = t'\left(x^{1'}, t + \frac{x^{1'}}{v_{\max} - v}\right) \ . \tag{19}$$

Let $x^{1'}$ be chosen small, and use an appropriate reparameterization, so that we may use the calculus of finite differences in solving for the proper transformations. Then, taking the *finite derivates* (not the derivatives),[25]

$$\frac{1}{2}\left(\frac{1}{v_{\max} - v} + \frac{1}{v_{\max} + v}\right)\frac{dt'}{dt} = \frac{dt'}{dx^{1'}} + \frac{1}{v_{\max} + v}\left(\frac{dt'}{dt}\right) \ , \tag{20}$$

or

$$\frac{dt'}{dx^{1'}} + \frac{v}{[v_{\max}^2 - v^2]}\left(\frac{dt'}{dt}\right) = 0 \ , \tag{21}$$

and

$$\frac{dt'}{dx^2} = 0 \ . \tag{22}$$

Since t' is linear, and we can assume $t' = 0$ when $t = 0$, the solution is just

$$t' = a\left(t - \frac{v}{v_{\max}^2 - v^2}\, x^{1'}\right) \ , \tag{23}$$

where $a = f(v)$, unknown for now.

Let $v_{\max}$ be represented by the same fixed value for both K, and k by a suitable reparameterization in each reference frame. Let attribute information transfer with attribute velocity $v_{\max}$ over a positive attribute distance y^1,

$$y^1 = v_{\max} \times t' \ , \tag{24}$$

and

$$y^1 = a v_{\max}\left(t - \frac{v}{v_{\max}^2 - v^2}\, x^{1'}\right) \ ; \tag{25}$$

then, with reference to the frame K, an ensemble expressed in the system k has attribute velocity $v_{\max} - v$, or

$$\frac{x^{1'}}{v_{\max} - v} = t \ . \tag{26}$$

So

$$y^1 = a\,\frac{v_{\max}^2}{v_{\max}^2 - v^2}\, x^{1'} \ , \tag{27}$$

and

$$y^2 = v_{\max} t' = a v_{\max}\left(t - \frac{v}{v_{\max}^2 - v^2}\, x^{1'}\right) \ , \tag{28}$$

where

$$t = \frac{x_2}{[v_{\max}^2 - v^2]^{1/2}}, \qquad x^{1'} = 0 \ ; \tag{29}$$

thus,

$$y^2 = a\left(\frac{v_{\max}}{[v_{\max}^2 - v^2]^{1/2}}\right) x^2 \ . \tag{30}$$

By substitution for $x^{1'}$, we obtain

$$t' = f(v)\,\gamma\left(t - \frac{v x^1}{v_{\max}^2}\right) \ , \tag{31}$$

and

$$y^1 = f(v)\,\gamma(x^1 - vt) \ , \tag{32}$$

$$y^1 = f(v)\, x^2 \ , \tag{33}$$

where

$$\gamma = \frac{1}{\frac{1-v^2}{v_{\max}^2}^{1/2}} . \tag{34}$$

To find $f(v)$, introduce K' with coordinates x^1, $x^{2'}$ and t' in parallel translation relative to x, such that the origin of k moves with attribute velocity $-v$. Assume the origins coincident. Applying the transformations we obtain

$$t' = f(-v)\,\gamma(-v)\left(t' + \frac{vy^1}{v_{\max}^2}\right) = f(v)\,f(-v)t\ , \tag{35}$$

$$x^{1'} = f(-v)\gamma(-v)(y^1 + vt') = f(v)f(-v)x^1\ , \tag{36}$$

$$x^{2'} = f(-v)y^2 = f(v)f(-v)x^2\ . \tag{37}$$

Since the transforms from K' to K are independent of t, it follows that K and K' are relatively at rest. Therefore,

$$f(v)\,f(-v) = 1\ . \tag{38}$$

Now, let there be an attribute distance of value ℓ, given independent of x^1 and $x^{1'}$; call this x^2 and $x^{2'}$, in k and K, respectively; then ℓ in k, with reference to K, is just

$$x^2 = \frac{1}{f(v)}\ . \tag{39}$$

Since, from symmetry, attribute distance can depend only on v, and not on direction or the sense of attribute speed, it follows that the interchange of v and $-v$ does not change ℓ. Hence,

$$\frac{1}{f(v)} = \frac{1}{f(-v)} \qquad \text{or} \qquad f(v) = f(-v)\ . \tag{40}$$

Thus, from Eqs. (37) and (39), it follows that $f(v) = 1$. Therefore, we have

$$t' = \gamma\left(t - \frac{vx^1}{v_{\max}^2}\right)\ ,$$

$$y^1 = \gamma(x^1 - vt)\ ,$$

$$y^2 = x^2\ ,$$

and

$$y^3 = x^3\ ,$$

where $\gamma = 1/[1 - v^2/v_{\max}^2]^{1/2}$; these being Eqs. (13), (14), (15) and (16), respectively.*
QED

Theorem 24: If $u = 0$, then $I' = I$ and $D' = D$; that is, an ensemble with zero attribute velocity induces the identity transformation.

Argument:

Trivially, from the definition of attribute distance, an ensemble with zero attribute velocity, with respect to some reference ensemble, is indistinguishable from the reference ensemble.

QED

Theorem 25: If $u = v$ and attribute speed < 1, then $I' = D'$; i.e., if ensemble A (which we may interpret as an observer) has the same attribute velocity as ensemble B, their relative attribute velocity is 0.

Argument:

Trivially, from the definitions of attribute distance and velocity.

QED

Theorem 26: If $I = D$,

$$v' = \frac{(I' - D')}{(I' + D')} = -u\ ; \tag{41}$$

i.e., with respect to a reference ensemble A with nonzero attribute velocity, an ensemble B with zero attribute velocity is an ensemble with the same attribute speed, but with opposite sign (direction).

Theorem 27: If the reference attribute speed is less than 1, a reference ensemble A with attribute $-u$ induces the inverse of the transformation induced by changing to a reference ensemble B with attribute velocity u.
corollary 27A:

Reversing ensemble A attribute velocity sign (direction) inverts the transformation induced on the attribute velocity of ensemble B.

* Note that, although our derivation is finite and discrete, we have deliberately followed the derivation of the Lorentz transformation developed by Einstein. We wish to emphasize that, contrary to common belief, the derivation of these transformations are not dependent upon the continuum. Where Einstein used derivatives, we use finite derivates, Eqs. (20) and (21). Where he allowed for a continuum of coordinates and velocities, we are restricted to the rational fractions which suffice per Pauli and Brodsky.

Theorem 28: The relative attribute velocity in the frame of ensemble A is bounded from below by the speed of ensemble B.

Theorem 29: The limiting attribute velocities for an ensemble are invariant under the transformation induced by nonzero attribute velocity; i.e., $[I,0]' = [I'',0]$ for some number I''.

Argument:

If the sign of the relative attribute velocity is positive, this follows from lower bound. If negative, the inverse transformation corresponds to positive relative attribute velocity, so that D must remain invariant.

QED

30: The increment quotient q is a function only of u.

Argument:

$[I,0]' = [I'',0]$, where I'' depends only on I and u. However, by Theorem 14, $[I',D] = [I,0]' + [0,D]'$, hence $I' = I''$; thus $q = I'/I$ depends only on I and u. However, q cannot depend on I, since otherwise $[2I,2D]'$ would have a different attribute velocity than $[I,D]$.

QED

Theorem 31: The inverse transformation induced by an ensemble with attribute velocity $-u$ has an increment quotient of $1/q$.

Argument:

The inverse transformation is I/I'.

QED

Theorem 32: The decrement quotient is the inverse of the increment quotient:

$$\frac{D'}{D} = \frac{1}{q} . \tag{42}$$

Argument:

First, reverse I and D to get $-v$, then take inverse transformation associated with $-u$, which multiplies the increment (which is now D) by $1/q$ to get $-v'$, then reversing I and D again to get v'; Thus, D' results from multiplying D by $1/q$, and it follows that $D'/D = 1/q$.

QED

Theorem 33:

$$q = \frac{(1-u)}{\gamma} ,$$

where

$$\gamma^2 = 1 - (u^2) . \tag{43}$$

Argument:

By the definition of the decrement quotient [Eq. (42)], $D' = D/q$, and from the increment quotient [Eq. (7)], $I' = qI$, so that from the definition of $v = (I'-D')/(I'+D')$ [Eq. (41)], we can write $v == (qI - D/q)/(qI + D/q)$. Since q is a function only of u, we can choose any values of D and I that lead to an equation in q and u, and its solution will define the general functional dependency. Assume $I = D$ so $v = 0$ and $v' = -u$; then, from Eq. (41),

$$-u = \frac{(qI - I/q)}{(qI + I/q)} = \frac{(q^2)-1}{(q^2)+1} .$$

Solving for q results in the relationship to be proved.

QED

Theorem 34: Relative to the zero velocity frame v = 0, the size change δm of an ensemble with attribute velocity v' is

$$\delta m = \frac{S}{\gamma} . \tag{44}$$

Argument:

Multiplying in the first part of (43)by $(1+u)$ gives $1/q = (1+u)/\gamma$ and $D = Ix$ for an ensemble with zero attribute velocity, this follows immediately.

QED

Theorem 35: Attribute velocities combine according to

$$v' = \frac{v-u}{1-vu} . \tag{45}$$

Argument:

By definition,

$$v = \frac{r-1}{r+1}, \qquad r' = \frac{I'}{D'} = \frac{r(1-u)}{(1+u)} ,$$

and

$$v' = \frac{r'-1}{r'+1} .$$

Then, by substitution and recollection of terms, we have

$$v' = \frac{v-u}{1-vu} . \tag{46}$$

QED

Theorem 36: For an ordering operator O of cardinality N and for each run of cardinality k, the **minimal attribute distance increment** i is

$$i(O) = \frac{1}{k!} . \tag{47}$$

Argument:

Consider a sequence of productions from an unspecified ordering operator of cardinality N to be used as a coordinate basis. We can compute the minimal attribute distance increment which can be generated in a given run of cardinality k of the operator, straightforwardly: it is the ratio of the number of (order) distinguishable states C (i.e., combinations—by excluding order, we take only those states that are distinguishable under a particular ordering operator) to the number of states P (i.e., permutations—by including order, we include all states, even those which are not distinguishable under a particular ordering operator).

$$C(k;N) = \frac{N!}{k!(N-k)!} , \tag{48}$$

$$P(k;N) = \frac{N!}{(N-k)!} , \tag{49}$$

$$i(O) = \frac{C(k;N)}{P(k;N)} = \left[\frac{N!}{k!(N-k)!}\right] \times \left[\frac{(N-k)!}{N!}\right] = \frac{1}{k!} .$$

QED

In general, C gives all the possible attribute states that could produce a sequence of state ensembles of the proper cardinality k, while P gives the number of ensembles of cardinality k possible in the same total space of cardinality N. This is, of course, subject to the constraint $k < N$.

Theorem 37: The total attribute distance $d(k;I;N)$ for an ensemble of cardinality k implied by I increments of i in a total space of cardinality N is

$$d(k;I;N) = \frac{I^k}{k!} . \tag{50}$$

Argument:

Suppose that we want to generate I increments in the attribute distance; then we want to turn the crank of the ordering operator which produces each attribute state I times. In the absence of further knowledge about the specifics of the ordering operator generation, we cannot enforce sequence so that the increments are disjoint; this is equivalent to sampling k objects from a population of I objects, with repetition

allowed. Call this $R(k;I)$; then, in general, for an ordering operator to generate an attribute distance d equal to I increments from a run of cardinality k on a space of cardinality N, we have:

$$d(k;I;N) = R(k;I) \times \frac{C(k;N)}{P(k;N)} = \frac{I^k}{k!} ,$$

where

$$R(k;I) = I^k .$$

QED

Theorem 38: The sum of all values of Eq. (50) from $k = 0$ to $k = K$ approaches e^I (for any expression of I) as K becomes large. We call this **the combinatoric definition of** $e(K)$.

$$e^I \approx \sum_{k=0}^{K} \frac{I^k}{k!} = e^I(K) . \tag{51}$$

Argument:

From the identity of definition of terms of the power series for e^I and the combinatoric definition of $I^k/k!$, the result follows for all discrete, finite values of k, N and I.

QED

Theorem 39: The attribute distance, given a distance function g transformed by reparameterization from a distance function f, is just:

$$d[k;I;g(M)] = \sum_{k=0}^{K} d[k;I;f(N)] \times D(f;k;N) , \tag{52}$$

where $D(f;k;N)$ are the k^{th} derivates of f.

Argument:

Consider a reparameterization of $d(k;I;N)$ from a distance function f on a d–space of cardinality N to a distance function g on a d–space of cardinality M, where the attribute is first order for both f and g. This is given by multiplying the attribute distance increment for $D(k;I;N)$ by a conversion factor (rational fraction), D. Since the attribute distance increment is inversely proportional to N, we have:

$$d[k;I;g(M)] = d[k;I;f(N)] \times D \approx d[k;I;f(N)] \times (N/M) . \tag{53}$$

Now, examine a general distance function $f(I;N)$ defined on a d–space S. By Principles I (finiteness), II (discreteness) and III (finite computability), $f(I;N)$ may

be expressed as some ordering operator O, which generates attribute states of an attribute of some order.* Call this order K. To express the generation of O in terms of the underlying discretum of cardinality N, we must take into account the possible contributions from all orders k from 0 to K. In general, D is not constant, but is dependent on f, N and k. Thus, for a general distance function $f(I;N)$, we have:

$$d[k;I;g(M)] = \sum_{k=0}^{K} d[k;I;f(N)] \times D(f;k;N) \ .$$

Note that the $D(f;k;N)$ may be solved for by the method of difference quotients[26]. These are the k^{th} derivates of f. The series is always finite (and, hence, there is no question of "divergence" for a given evaluation of the series) since N is fixed. For sufficiently large N, the series Eq. (52) approaches the Taylor series with arbitrary precision.

QED

The Lagrange form of the remainder is of particular interest here, since it gives a measure of the deviation from the discrete form by the analytic form of the truncated Taylor series Eq. (52).

$$R_n(x) = \frac{f^{n+1}(\epsilon)}{(n+1)!}(x-a)^{n+1} \ , \tag{54}$$

where

$$x < \epsilon < a \ .$$

For sufficiently complex attributes and large N, this approaches the usual form of the exponential operator, as normally used to describe transport along a parameter. The sum may be understood as the contributions to distinguishability by successively more complex aspects of the attribute, weighted by the probability that a particular sequence that can generate the required distance is the correct one.

Theorem 40: The **incremental transport** x_0 along a basis x^i at x parameterized on t is just

$$d[k;I;f(x+x_0)] = \sum_{k=0}^{K} \left(\frac{(x_0)^k}{k!} \right) \times D(f;k;t) \ , \tag{55}$$

where $D(f;k;t)$ are the k^{th} derivate operators on x with respect to t.

* Recall that an attribute of an attribute is called an attribute of second order, an attribute of an attribute of an attribute is called an attribute of third order, etc.

Argument:

We wish to compute the incremental transport δx along a given coordinate basis x in terms of the above formulation. This is equivalent to a reparameterization from f to g, in which f and g are related as follows:

$$g(x) = f(x+x_0) \ , \tag{56}$$

with x_0 being the minimum attribute distance increment.

Since we do not know the particular ordering operator, but only the ultimate cardinality of the ensemble and the cardinality of the space, we must use the general form of reparameterization, Eq. (52). The result follows from substitution of Eq. (56) in Eq. (52).

QED

If the ordering operator produces a sequence which is of first order (linear in the ordering parameter), then the rate of change of attribute distance with respect to the ordering parameter is constant. This is, of course, just the first discrete derivative (derivate). If the ordering operator produces a sequence which is of second order, then the rate of change of attribute distance with respect to the ordering parameter is a first order function of the ordering parameter, i.e., the second derivate. Similar arguments hold for ordering operators of higher order.* In order to compute the transport along x^i from x to $x+x_0$, we must take into account the contributions of each order up to the order of the operator.

Theorem 41: Given reference frames F and F', coordinate transformations between unsynchronized events satisfy Eqs. (13), (14), (15), (16) of Theorem 23, statistically.

Argument:

Consider two reference frames, F and F', given by two sets of independent generations S_1, S_2 and S_3, and S_1', S_2' and S_3'. Again, we initially synchronize each set of three and let them go independently (Theorem 13). We count the occurrence of an attribute state which may be used as a metric mark in one of the generations as a 1, and any other attribute state as a 0, for purposes of analyzing the statistics.

Now, however, we have two ordering operators which we label O and O', global to F and F', respectively. In the absence of further information regarding the ordering operator, we will assume a normal distribution (Principle IV) of distinguishable states

* This analysis is consistent with the requirement that the k^{th} derivate may be obtained from confluent divided differences of k arguments. The k arguments are order independent and, hence, are "sampled from a population of cardinality I with repetition allowed," as previously noted.

about a metric mark in either F or F' (i.e., generated by independent O and O' as per Theorem 13).†

Consider a discrete mapping from F to F'

From the combinatoric definition of the base of the natural logarithm and the definition of the normal distribution, a sample size of two standard deviations around an attribute state, taken as the mean or center of the distribution, will consist of all the distinguishable states around the mean[28]; and, therefore, a metric mark, with a probability equal to the ratio of distinguishable states to all states, summed over all possible attribute states that might be selected in F' as a metric mark. However, this is just $1/e(N)$; thus, for a well-defined "metric mark" in F, a arbitrary transformation to F' results in a measure in F' which deviates from a metric mark by $\pm\sigma$. For a normal distribution, 2σ is just the transport for a minimum attribute distance increment. Computing the population variance σ^2 is then, for population of size N

$$\sigma^2 = \sum \frac{(U-\mu)^2}{N} , \tag{57}$$

where μ is the average of U. Suppose U is just the attribute distance in F; then the "mean attribute distance" μ is just the attribute velocity multiplied by the number of generations over which the attribute velocity has evolved. This is equivalent to giving the number of increments minus the number of decrements in terms of a global ordering operator spanning both F and F'. In other words, in the frame of the minimum of the maximum attribute velocities, the maximum number of generations for the ordering operator producing the attribute will be N, the cardinality of the universe. The normalized variable x^* corresponding to x with mean 0 and variance 1 is just

$$x^* = \frac{(x-\mu)}{\sigma} . \tag{58}$$

Here, $x-\mu$ is just the difference in the global frame between the increment and the decrement. Sigma is then the probability of obtaining an attribute increment corresponding to the ordering operator, which produces metric marks in F' relative to the ordering operator, which produces metric marks in F. Thus, Eq. (58) is equivalent to going to dimensionless (i.e., frame independent) quantities[29].

Note that for a discrete function on finite domain, this x^* is always bounded and finite; i.e., sigma is never 0 whenever $x-\mu$ is not 0. In addition, since the fluctuations in x are bounded and finite, it makes no sense to speak of specifying x beyond that discrete step length which results in the smallest fluctuation.

† In the absence of large N, we could as easily use the binomial distribution justified by the combinatorics to reflect finite N, and use the appropriate Yates adjustment in which $y_0 - \frac{1}{2}$ is substituted for y_0 in the computation of the probability $Pr(y > y_0)$, so that the unit normal variable probability $Pr(z > z_0)$ is just $z_0 = (y_0 - \frac{1}{2} - Np)/\sqrt{Npq}$, where p is the probability of a 1 and q is the probability of a 0. However, we assume here that the normal approximation is adequate in the light of the usual criteria that $N > 5$ and the absolute value of $[(1/\sqrt{N})(\sqrt{q/p} - \sqrt{p/q}]$ is less than 0.3[27].

Now, let μ be a d-velocity v multiplied by the number of generations t over which it is measured, and let x^1 be the attribute distance in F, and y the attribute distance in F'; then, from Eq. (58),

$$y^1 = (x^1 - vt)\,\gamma , \tag{59}$$

so that

$$\sigma = 1/\gamma \qquad \text{and} \qquad \mu = vt ; \tag{60}$$

then y^1 is interpretable as the normalized variable associated with x^1. Clearly, as long as β is defined as v/v_{max} as in the derivation of Theorem 23, we have recovered the coordinate transformation in the absence of synchronization. Therefore, the coordinate transformations, Eqs. (13), (14), (15) and (16), are applicable at all rational scales, for all frames and for all attributes.

QED

It is important to understand that the mean attribute distance increment computed by going to dimensionless coordinates and transformed from a metric mark in F arbitrarily to F' (i.e., in the absence of synchronization between F and F'), is, thus, identical to the minimal attribute distance increment, transformed under synchronized frames for metric marks. This result may also be taken as proof by construction that the combination of the minimum attribute distance increment and the coordinate transformation of Theorem 23 has bounded (i.e., over the range of meaningful rational fractions which may be defined by reparameterization on the d–space) scale invariant significance.*

Theorem 42: Let $P = Prob(I \rightarrow I{+}1)$ and $Q = Prob(D \rightarrow D{+}1)$ for $N \rightarrow N{+}1$. The **uncertainty** associated with a coordinate transformation satisfying Theorem 41 between meter marks is given by:

$$1 - (P-Q)^2(\Delta x)^2 = 4PQ(\Delta x)^2 > 1 . \tag{61}$$

Argument:

Now, since the variance is given by

$$1 - (P-Q)^2 = 4PQ , \tag{62}$$

and with

$$P = \tfrac{1}{2}\,(1+\beta) , \tag{63}$$

$$Q = \tfrac{1}{2}\,(1-\beta) , \tag{64}$$

⋆ This analysis shows why the random walk derivation of the Lorentz transformation, as presented by Stein, works[30].

the probabilities of I and D, respectively, for N generations, we obtain

$$\sigma = (NPQ)^{1/2} = \left[\frac{N}{4(1-\beta^2)}\right]^{1/2} , \tag{65}$$

so that

$$\sigma L = L\left[\frac{N}{4(1-\beta^2)}\right]^{1/2} = \left(\frac{L\gamma}{2}\right) N^{1/2} , \tag{66}$$

where L represents the discrete increment for the variable and

$$(P-Q)L = \beta L . \tag{67}$$

Thus, we arrive at an interpretation of the coordinate transform between reference frames and between metric marks. Note that, because N is finite, the variance is finite, i.e., bounded. This provides normalization of the transform, as well as a "maximal velocity." We have simply applied a consistency requirement to all allowed (i.e., rational) velocity frame transformations, namely bounded scale invariance.

Furthermore, because σ is bounded from below by one generation, it follows that the minimum deviation is always 1 between metric marks. Fluctuations between metric marks are thus bounded above and below. Letting Δx represent the discrete increment in x, the bound from below gives the uncertainty in the region directly from the variance:†

$$1 - (P-Q)^2(\Delta x)^2 = 4PQ(\Delta x)^2 > 1 .$$

QED

4.1 Multiply Connected Attribute Spaces

We now show how a d–space can be multiply connected, and derive some consequences of this multiple-connection. Unlike other notions of nonlocality, a multiply-connected d–space has a sequence of maximal attribute velocities.

Theorem 43: In a multiple attribute d–space, the sequence of maximal attribute velocities V_i has at least one value which is a least upper bound $V_{\max}$ and at least one value which is a greatest lower bound $V_{\min}$.

† As we will see in the physical interpretation, this fact implies that we do not require the concept of the wave function. Our "collapse" is nothing more than the attainment of more information about the specific ordering operator involved in the "evolution" of the discrete system. The uncertainty is nothing more than a quantification of the amount of detail expressible, given the selected basis having rational fraction values; i.e., as "meter marks."

Argument:

Trivially, from the definition of maximum attribute velocity, indistinguishable, attribute state and Principle I.

QED

Theorem 44: A multiple attribute d–space has relationships between attribute distance functions satisfying Eqs. (13), (14), (15) and (16), which display nonlocal correlations (i.e., require more generations than allowed by the ordering operator for v_{max}) and indeterminate relation (i.e., cannot be expressed as a function of N and the attribute states alone) to at least one of the attributes.

Argument:

Consider a discrete d–space U of cardinality N, with attributes E and P such that the number of attribute states of E is much greater than the number of attribute states of P. Further, consider d–subspaces L, R and S of U.

For a particular attribute A, we will represent the attribute distance from one d–subspace X to another d–subspace Y by $d(A : XY)$. By $V(A)$, we will mean the maximum of an attribute velocity $v(A)$ in the attribute A. By $C(X : EP)$, we will mean the minimum computational power necessary to represent the relationship between the attributes E and P in the d–subspace (or d–space) X.

Let the combined cardinalities of L, R and S be represented by M, and suppose that the number of attribute states of E in U is greater than $M+\log_2 M$ (Theorem 13); then there exist sequences of attribute states of E algorithmically producible within U, which cannot be differentiated from randomly distributed sequences of attribute states from within L, R or S, or any combination of L, R and S. Now consider the further relationship between E and P within U. Suppose that E is related to P via a function F which, by virtue of the fact that the number of attribute states of E is much greater than P, is a many-to-one d–map. It follows that the relationship F cannot be known within L, R or S, even when well-defined on U. Clearly, such a system is capable of exhibiting local "random" behavior.

Furthermore, it is clear that there must exist correlations (or anticorrelations) of P in $L+S$ and P in $R+S$, since this relationship is completely determined by F and incompletely expressible to either $L+S$ or $R+S$.

L, R and/or S are not large enough to discern the algorithmic relationship between P and E. By hypothesis, the maximum attribute velocity of P, $V(P)$, is greater than the maximum attribute velocity of E, $V(E)$. It follows that the correlation of P between L and R in U is limited by the velocity $V(P)$ rather than $V(E)$, and is thus nonlocal. Within the context of describing the system via the attribute E with maximum attribute velocity $V(E)$, these correlations appear instantaneous, based upon measurements of $d(E : SR)$ and $d(E : SL)$.

QED

Theorem 45: The attribute having the infimum of d–set of maximal attribute velocities for a maximal attribute velocity also has the smallest of the corresponding minimal attribute distance increments.

Theorem 46: The attribute having the infimum of d–set of maximal attribute velocities for a maximal attribute velocity corresponds to the attribute having the largest number of possible attribute states.

Theorem 47: The maximal range of attribute velocities over which relationships may be specified between arbitrarily selected attributes defined on some d–space is bounded from below by 0 and from above by the infimum of the d–set of maximal attribute velocities $V_{\min}$.

Argument:

Trivially, from the fact, if the zero attribute velocities are identified equal, then d–maps between attribute velocities can only be 1–1 over the interval $[0, V_{\min}]$.

QED

4.2 A Combinatoric Construction of Commutation Relations

The commutation relations as normally understood in quantum mechanics actually involve two quite distinct principles. The first is the principle that noncoordinate bases do not commute. Given a coordinate system x^i, one can adopt the derivate operator d/dx^i as a basis for the vector field. However, any linearly independent set of vector fields can serve as a basis, and one can easily show that not all of them are derivable from coordinate systems. This is because the operators d/dx^i and d/dx^j commute for all i, j, while two arbitrary vector fields do not commute.

The Exponentiation of the Derivate Operator d/dp

Theorem 48: The transport $p_0 + \epsilon$ along $x^i(p)$ may be given as

$$e(N)^{\epsilon d/dp} x^i \Big|_{p_0} . \tag{68}$$

Argument:

Let $D = d/dp$ evaluated at some point p_0 on a particular coordinate parameterization. Suppose the coordinate values $x^i(p)$ of points along the integral "curves" of a "vector field" d/dp are discrete functions of p; then the coordinates of two points with parameters p_0 and $p_0 + \epsilon$ are related by Eq. (52):

$$\begin{aligned} x^i(p_0 + \epsilon) &= x^i(p_0) + \epsilon \left(\frac{dx^i}{dp}\right)\Big|_{p_0} + \left(\frac{1}{K!}\right) \epsilon^K \left(\frac{d^K x^i}{dp^K}\right)\Big|_{p_0} \\ &= e(N)^{\epsilon\left(\frac{d}{dp}\right)} x^i \Big|_{p_0} \end{aligned} \tag{51}$$

where $e(N)$ is just the power series expansion of e truncated at the N^{th} term by the definition of $e(N)$.

QED

Discrete Geometric Interpretation of Generalized Commutation

We will use the shorthand notation for Eq. (52) developed in the previous theorem in the derivation of the discrete commutation relations which follows*

Theorem 49: The order dependence $x(B) - x(A)$ of the derivate operators d/dp, d/dq is given by

$$x(B) - X(A) = \left[\frac{d}{dp}, \frac{d}{dq}\right] + \bigcirc\left(\epsilon^3 \frac{d^2}{dp^2} \frac{d^2}{dq^2}\right) . \tag{69}$$

Argument:

Notice that by definition of a coordinate basis (orthonormality), x^1 is constant along the lines of x^2, which are the integral curves of the derivate operator d/dx^2. That is why the derivate operators d/dx^1 and d/dx^2 commute: each is a derivate along a line on which the other is fixed.

Consider a basis d/dp combinatorially produced by Bernoulli trials vis-a-vis an ordering operator. Consider a second basis d/dq similarly, but independently produced. Now consider a transformation from one basis to the other; i.e., we seek a transformation which takes us a distance ϵ from a point P to a point R in x^i, using λ for transport; see Figure 14.

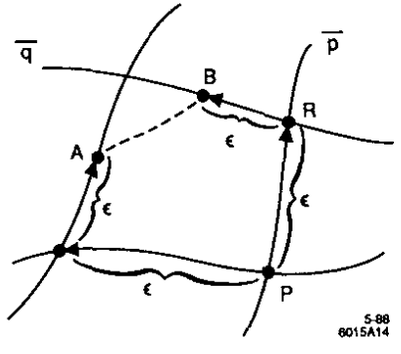


Figure 14 Transport from a point P to a point R, using d.

The two arbitrary vector fields V and W are defined by $V = d/dp$ and $W = d/dq$. Even the fact that the parameterizations look like that of a coordinate system is an

* Adapted from B. Schutz.[31]

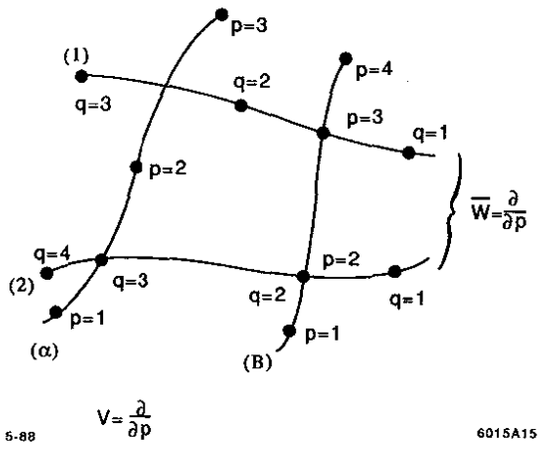


Figure 15 Relation between parameterization and transport (see text).

artifact of 2–space; in 3–space it may happen that curve 2 intersects curves a and b, but that curve 1 only intersects curve a; see Figure 15.

We obtain a picture of the vector $[V,W]$ in the following manner. Consider a starting point P, moving $dp = e$ along the V curve through P, and then moving $dq = e$ along the W curve. One winds up at A. Starting again at P and going first along the W curve, and then along the V curve, takes one to B. The vector stretching from A to B is $\epsilon^2[V,W]$, to lowest order in ϵ; see Figure 14.

The transport along x from P to R in discrete step lengths is just:

$$x(R) = e(N)^{[\epsilon d/dp]}x \quad \text{at} \quad P\,. \tag{70}$$

Now assume that we have similar relationships for d/dq. For a point A in x, ϵ distance away from P along d/dp and ϵ distance further along d/dq, the transformation is just the product of the two operators (i.e., transform along d/dp, then along d/dq).

$$x(A) = e(N)^{[\epsilon d/dp]} \times e(N)^{[\epsilon d/dq]}\, x \quad \text{at} \quad P\,. \tag{71}$$

Similarly, we may travel from P to a point B, which is located by just changing the order of the transforms. We then obtain

$$x(B) = e(N)^{[\epsilon d/dq]} \times e(N)^{[\epsilon d/dp]}\, x \quad \text{at} \quad P\,. \tag{72}$$

Now find the distance from B to A:

$$\begin{aligned} x(B) - x(A) = [e(N)^{[\epsilon d/dp]} \times e(N)^{[\epsilon d/dq]} \\ - e(N)^{[\epsilon d/dq]} \times e(N)^{[\epsilon d/dp]}\, x \quad \text{at} \quad P\,. \end{aligned} \tag{73}$$

Now we undo our shorthand notation for Eq. (52), in order to multiply out the terms explicitly, and explicitly ignore higher-ordered terms which result. Expanding, we have the right-hand side of Eq. (73) as:

$$\left[1 + \frac{\epsilon d}{dp} + \tfrac{1}{2}\frac{\epsilon^2 d^2}{dp^2} + \bigcirc(\epsilon^3) \quad , \quad 1 + \frac{\epsilon d}{dq} + \tfrac{1}{2}\frac{\epsilon^2 d^2}{dq^2} + \bigcirc(\epsilon^3)\right]. \tag{74}$$

This is just

$$= \epsilon^2 \left[\frac{d}{dq}, \frac{d}{dq}\right] + O(\epsilon^3)\,. \tag{75}$$

Thus, for two discrete operators ("vector fields") d/dp, d/dq which are not part of the coordinate d–basis x, the commutator is just the open part of an incomplete parallelogram, whose other sides are equal parameter increments along the integral curves of the vector fields. Note that the parallelogram is complete if and only if d/dp, d/dq are one to one with the coordinate d–basis; see Figure 16.

QED

It is important to understand how the operators which generate discrete distance functions might not be a part of the coordinate d–basis. Earlier, we noted that two ensembles A and B with increment and decrement I, D and I', D', respectively, were said to be independent if and only if all the defining states for A and B were distinguishable.

Theorem 50: For any two bases P and Q, the commutator of P and Q vanishes if and only if P and Q are independent; i.e., if and only if P and Q are coordinate bases.

Suppose that not all the defining states for A and B are distinguishable; then for some generation of the ordering operator, a redundant attribute state (instance) is generated. As a result, the additive law for attribute distance must fail; i.e., the sum of the total sizes for A and B does not equal $S'+S$. The sign of the deviation depends upon whether the deviation from $S'+S$ is accounted for by a deviation from $D+D'$ or by a deviation from $I+I'$ in the summation. Although the deviation can be treated as an attribute distance in its own right (indeed the inverse function of the additive law encourages us to do this), the ordering operator required to generate this deviation is clearly not independent of the generation of the two ensembles (consisting of a mixture of distinguishable and indistinguishable states), and is absolutely independent of the representation of both ensembles as being strictly independent (i.e., incorporating only distinguishable states); thus, it may be counted as a basis which behaves locally as an independent dimension.

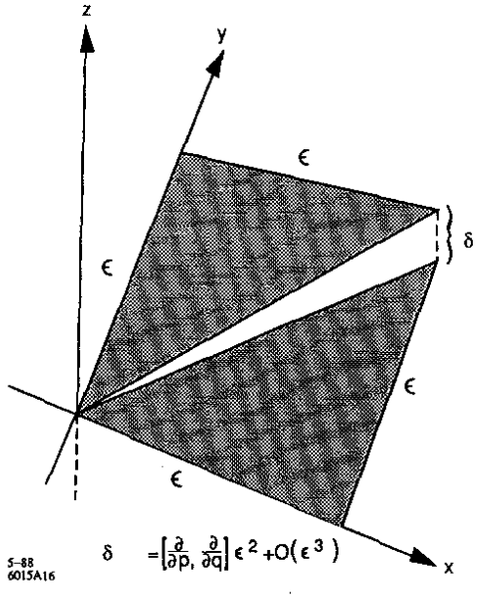


Figure 16 Incomplete closure for parameters that are not part of a coordinate d–basis.

If any distinguishable states are shared between the two coordinate parameters (i.e., one parameter is a function of the other), the product of the transports becomes order dependent: the computation of attribute distance for the first basis transport consumes the state and, thus, alters the ratio of distinguishable to total states for the second basis transport. Since the derivates for the basis are not in general the same, this results in a nonvanishing commutator. On the other hand, if the bases are independent, the commutator will clearly vanish.

Theorem 51: The commutator is bounded above and below.

Argument:

In a finite system, the commutator can clearly be no larger than the absolute maximum attribute distance representable in the dependent basis, where we assume that a dependent basis provides less information than the independent basis. Hence, the commutator is bounded. If the dependent basis has cyclicity ξ with respect to the independent basis, mapping each successive ξ distinguishable attributes of the independent basis to the same ξ attributes of the dependent basis, then the commutator is bounded by ξL (and in fact is equal to ξL), where L is the "conversion length" between bases. Based upon arguments previously given regarding dimensionality, it is clear that fluctuations of the commutator less than ξL are not consistently representable within the n–space (i.e., they occur between meter marks).

QED

Theorem 52: If $P = P(Q)$ is a first order derivate, then Eq. (75) is exact without higher-ordered terms.

Argument:

Since higher-ordered terms in Eq. (75) depend on higher-order derivates not vanishing, the theorem follows immediately.

QED

Theorem 53: For bases P and Q, if P is cyclic in Q (an angle variable), then

$$[P,Q] = \pm\frac{i\,\text{Constant}}{2\pi(N)}\,, \tag{76}$$

where $\pi(N)$ is just the discrete computation of π by the combinatoric method in a d–space of cardinality N, as given above.

Argument:

If the indistinguishable attribute states involved combine to behave as distinguishable attributes in the proper manner, this independent dimension will behave mathematically just as though it were imaginary. Suppose, as in Theorem 52, that one of the two bases P is a function of the other:

$$P = P(Q)\,. \tag{77}$$

Furthermore, suppose that $P(Q)$ describes either a closed "orbit" or a periodic function of Q. If one of the bases is cyclic, its "conjugate" basis is constant. The corresponding orbit in the QP discrete 2–space is then just a "horizontal straight line." Following Goldstein[32], the "motion" may then be considered as the limiting case of a rotation type of periodicity, in which Q may be assigned an arbitrarily long period (subject to N, of course). This is just a change of coordinates from the real coordinate P to an imaginary coordinate J in a complex discrete 2–space, following the usual practice of using complex plane to represent such a change of coordinates; see Figure 17.

Since the coordinate in a rotation periodicity is invariably an angle, such a cyclic Q always has a natural period of $2\pi(N)$. Accordingly, the length of the path in QP discrete 2–space evaluated from 0 to $2\pi(N)$ is just $2\pi(N)$ and QP becomes:

$$J = 2\pi(N) \times i \times p\,, \tag{78}$$

for all cyclic variables. Note that we evaluate π for cardinality N here. That is, we construct the combinatoric valuation of π on the global d–space of cardinality N, and not on the local d–subspace; then we map minimum Q to 0 and maximum Q to $2\pi(N)$. The value of Q measured as an angle is then discretized in increments of $2\pi/N$ from 0 to 2π by the mapping.

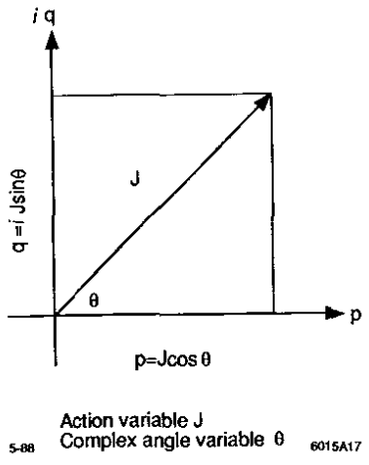


Figure 17 Relation between p, q and angle-action variables.

Given Eqs. (74) and (76), we may now express the commutation relation between J and Q:

$$[J, Q] = [2\pi(N)iP, Q] = \text{Constant} ,$$

or

$$[P, Q] = +i(\text{Constant})/2\pi(N) .$$

QED

From our earlier result, however, the general commutation to first order is just

$$[P, Q] = \frac{L}{\epsilon^2} . \tag{79}$$

If P and Q are linearly related, then the higher-ordered derivates vanish and Eq. (79) is exact. If we then take ϵ to be the minimum nonvanishing discrete value, with suitable reparameterization, we have

$$[P, Q] = L , \tag{80}$$

for the least increment in the complex angle variable.

5. A DISCRETE CONSTRUCTIVE MODELING METHODOLOGY

5.1 Definitions

Having developed the elements of a discrete, finite and computational formalism via the ordering operator calculus, we proceed to a mathematical foundation for a discrete and constructive modeling methodology. Such a methodology will allow us to use the ordering operator calculus to model various phenomena which do not have the intrinsic properties required by continuum mathematics.

We motivate the modeling methodology through a variation of a dictum issued by Bastin and Kilmister[33] in 1973 concerning the separability of syntax and semantics in a mathematical system, which we refer to here as the Separability Lemma.

> *Separability Lemma:*
> ***A system has a mathematical structure (syntax) which can be expounded separately from the interpretation of it (semantics), provided that it is understood that the mathematics describes a process which can be represented as a computer program.***

Clearly, the ordering operator calculus meets the criteria of the Separability Lemma as demanded by Principles I–IV. We are now ready to define a modeling methodology which consists of three broadly-defined structures: an epistemological framework, a representational framework and a procedural framework.

An **epistemological framework** or **E–frame** is a *d*–set of loosely-defined agreements made explicit by those engaging in the process of modeling (i.e., by injecting information into the model formulation).

1) *Agreement Upon Intent*

   The intent of the modeling effort must be agreed upon. The practice being modeled must be identified. It is also desirable to establish agreement regarding the conditions under which the effort will have been determined to fail, means of validation, the degree of accuracy required of the model (a stop rule) and rules for evaluation.

2) *Agreement On Observations*

   The ensemble of objects O, which constitutes the observations of and about the practice must be agreed upon.

3) *Agreement of Cooperative Communications*

   - commonly defined terms as fundamental

     Fundamental terms, as used in describing the practice, must be understood. They CANNOT be defined.

   - fundamental versus derived terms

     An operational distinction between fundamental and derived terms must be practiced.

- agreement of pertinence

  Engaging in attempts to communicate about the practice being modeled must be founded on an agreement to assume and attempt pertinence.

4) *Agreement of Explicit Assumptions*

   There must be an agreement to make assumptions explicit, rather than allowing them to be implicit.

5) *The Razor*

- agreement of minimal generality

  The "scope" of the modeling effort at any point in the evolution of the model should be constrained to manageable proportions.

- agreement of elegance

  The model should display a consistent and transparent structure, which minimizes the statement (size) of the model, while maximizing its explanatory (and in the event of a theory, its predictive) power.

- agreement of parsimony

  The model should contain as little as possible that is either (a) sufficient but not necessary, or (b) necessary but not sufficient in modeling the intended practice.

A **representational framework** or **R–frame** is an abstract formalism FS, consisting of a set of symbols F and a set of rules of manipulation I. It is an uninterpreted typography.

A **procedural framework** or **P–frame** is an algorithm which serves to establish rules of correspondence C between the observations O (as agreed upon in the E–frame) and the symbols of the R–frame F, and which then, through recursion, serves to modify the rules of correspondence and the E–frame and R–frame, until a sufficient level of agreement concerning accuracy is achieved or the model fails. Kuhn[34] would call such a failure a "crisis," which in the fullness of time will lead to a "paradigm shift."

Thus, we see a relationship between two d–sets being established (the O and F), with two d–sets of rules (I and C) for modification and/or information extraction.

We now cast this in terms of the ordering operator calculus and, specifically, of the finite differential geometry which we have constructed within it.

An **ob set** O is an ensemble of observations. The obs are differentiated (altered from a d–sort to a d–set) by one or more ordering operators, which serve to establish the lattice structure of the obs.

An **ob subset** is a d–set of obs, defined by at least one ordering operator. They may be multiordered and multiply-connected.

A d–**sort of formal symbols** F is an ensemble of labels which may be ordered (converted to a d–set) by a d–set of rules of manipulation I. The resulting d–set FS of formal symbols F with rules of manipulation I is called a **formalism** or **representational framework**, and may be either closed or open under the rules. Generally, this serves to form a abstract combinatorial system.

A **P–frame rule of correspondence** is a binary d–map between an element of F and an element of O. In practice the d–set of all rules of correspondence established up to some step in the modeling procedure are expressed as a *dictionary:* given an element of O one may look up a procedure for finding the corresponding element of F.

A **procedure** P is a bounded, recursive algorithm which (a) provides a recursive and exhaustive enumeration of the elements of O and the elements of F, such that there exists a smooth d–map between O and F in the sense given above, constructed from the d–set of rules of correspondence and which (b) provides a recursive reparameterization of the d–map, such that there exists a 1–1 d–map between a d–subset of O and a d–subset of F.

Ideally, the cardinality of these d–subsets increments with each recursion of the P–frame procedure, up to the cardinality of O itself.

5.2 Observation Space

We begin with a number of observations which may be clustered* (grouped into prearranged classes) into d–sets O_i. These observation d–sets are said to cover the observation d–space O in the sense that $\bigcup_i O_i = O$. Because our O must have boundaries—for any hypothetical O_u (O is a d–subset of O_u)—and is discrete, O is non–Hausdorf.

Clearly, for any finite O, there are a finite number of possible disjoint partitions of O; namely, $\sum_k k!/n!(k-n)!$, where n is the cardinality of O, and k ranges from 0 to n. However, the partitions need not be disjoint—we allow dependent observations, and any ob to be in more than one partition. Thus, the number of partitions may be as large as we wish, being determined by the bound we place on the combinatorics of repetitive sampling with replacement.

It is often convenient, in the absence of any constraints, to take a discrete version of R^N as the image space, thus allowing an analytic interpolation for functions defined on the space. We map each partition O_i of O to some subspace S_i of R^N by some d–map R_i. If each such subspace S_i of R^N is arbitrarily "labeled" with some formal symbol F_i, then the partitions O_i of O may be taken as "objects" in O and referred to by the F_i. The R_i then form rules of correspondence.

We define relationships between the O_i objects in terms of the coordinate transformations between the S_i .

* We will discuss methods of clustering compatible with the modeling methodology and the ordering operator calculus in a later paper. Note that if a distance function or a norm is definable on the O, the method of minimal distances may be used as a clustering algorithm for the partitioning. Methods based upon a general attribute distance function are closely connected to a general theory of computational measurement, in development.

Note that our definitions tell us immediately that there is no *a priori* parameterization on S which gives a preferred reference frame. In fact, there is no structure at all on S without a parameterization. There exists no metric, only local topology induced on O by F via R. The global topology is given by the cardinalities of O and F and by the partitioning of O, as well as by coordinate transformations between partitions induced by requiring that the formal rules of manipulation I_i map isomorphically to O, via the rules of correspondence R_i, giving connectivity to the topology. The image in O under R of I may leave invariant certain attributes of O, the study of which provide an understanding of the structure of the formal model of O.

5.3 The Modeling Methodology Algorithm, Models and Theories

We now give a specific P–frame algorithm, which meets the criteria established in the preceding section, and which establishes and guides the evolution of the model.

1. Choose the ob set O with n elements. This is a recursively enumerable d–set with cardinality n.
2. Partition the ob set O.
   a. Define the n obs (labels) by partitioning the d–set O into disjoint d–subsets.
   b. Choose a set of symbols O_i for these partitions, labeling them.
3. Select or develop an abstract formalism FS meeting the criteria of an R–frame.
4. Choose a set of rules of correspondence R between the symbols O_i of O and the formal symbols F.
5. Map to some space such as R^N. (We can always choose our discrete version of R^N locally for the d–map, although we must then define the obs on open d–sets.)
6. Determine relationship between obs vis-a-vis the formalism. In particular, determine the image of the I in O under R.
7. Establish a set of coordinate transformations and determine the induced structural invariances, in order to identify the interpreted global properties of the model.
8. We say that this procedure establishes a **model**, if the cardinality of the O_i is the same as that of F and if R is an isomorphism between O and FS. If the isomorphism fails, we call the result a **theory**, in that it has predictive power. In empirical practice, we will rarely obtain a formal model.
9. If a model is not established because the isomorphism fails, then recursive application of the P–frame procedure is required to evolve the model. While no deterministic algorithm may be given which prescribes how the model should be altered, given a certain failure of the isomorphism, the P–frame procedure P allows one to develop heuristic knowledge about the modeling practice and how best to proceed in modifying the model. This heuristic knowledge may be made explicit within the E–frame from the outset and, indeed, becomes a part of the E–frame via P–frame recursion.†

Keep in mind that through P–frame recursion, one has many options: we may alter the partitions of O, the range of the maps R, the coordinate parameterizations on O, the d–maps R, the rules I, and so on. Each recursion of the procedure P modifies one and only one such aspect of the model; in so doing, the entire model must be reexamined for consistency and completeness of the representation, as each change alters the definition of one or more ordering operators. These modifications are necessarily inductive, and therefore have unpredictable consequences.

The revolutionary step is taken based on an inductive decision that a Kuhnian crisis[35] has developed. This is largely based upon *subjective* criteria concerning the viability of the model and, in some sense, an *intuitive* measure of the relative benefits of proceeding, starting over or opting for a radical revision. It is important to note that such criteria can be agreed upon as part of the E–frame; namely, agreeing in advance how much and what kind of deviation from the required isomorphism will be tolerated and how the validity of the modeling effort will be judged.

We halt the classical infinite regression of analysis of terms in modeling by recognizing the effect of the epistemological framework. We deny the validity and the value of any attempt to analyze "theory-laden"[36] language as used in the E–frame. Such an analysis lies outside the purported task of generating a specific model, and would require us to generate a model containing the specific model, as an instance. In particular, analysis of fundamental terms involves treating these terms as the ob set for a modeling effort. In keeping with the agreed upon intent of the modeling practice and our methodology, we can not engage in such analysis. The practice would necessarily involve nonconstructive methods: the analyst would have to work from the specific model by generalization, having failed to construct the general model first. The transition from the specific to the general is not only inductive in nature, but not recursively definable, and constitutes a revolutionary redefinition of the modeling effort as specified in the agreement of intent.

Note the implication here that it is possible to work from the general to the specific. It is possible to constructively "model the model" or even the modeling process. Indeed, part of the power of our modeling methodology lies in the constructive and recursive nature of the process.

In practice, we always bootstrap into the modeling process with a set of loose agreements and definitions (we don't really know what we are talking about), but the ordering operator calculus gives us a consistent mechanics of typography and the procedural framework gives us a recursive method of evolving toward an acceptable model and definitions. Once the process has begun, each pass through the P–frame may generate a modified, but nonetheless well-founded and well-defined, E–frame and R–frame. Constructively, we may keep records of our efforts and review these at will. On starting the effort, we have no record of earlier effort and no way of (re)constructing one; we may make no constructive claims regarding either the earlier

† We call the process of exercising P *recursion*, rather than *iteration*, because it operates on itself, as well as the model. In some sense, P, together with the modeling participant operating on the model, constitute a self-organizing system.

effort or the results of that earlier effort. In some sense, we, thus, have a "fixed past and uncertain future," but with a fixed starting point.

5.4 HIERARCHICAL MODELS

We will frequently have cause to deal with hierarchical structures. For this reason, we give a P–frame algorithm for constructing **hierarchical models** as a constructive definition.

1. Start with a model.
2. Specify a many-to-one d–map from the formalism F to ob labels O_i.
3. Redefine the partitioning via the process of refinement, mapping from the image in O to the representation d–set F with new mutually disjoint partitions, using inverse d–map of the R_i. This insures consistency for next step.
4. Remap the formalism from new partitions induced in F under the inverse of the R_i to the image space O, using old mapping R.
5. Keep in mind the constraints of a many-to-one d–map. This d–map provides inclusion relations on the d–set F; thus partitions contain partitions or parts thereof, forming a lattice of partitions.

Theorem 54: For each model with multiple partitions mapped to a representational framework without disjoint refinement, there exists a hierarchical model with an equivalent local topology.

6. AN INTERPRETATION: LABORATORY PHYSICS

6.1 ESTABLISHING THE E–FRAME

We start on the route to physical interpretation by adopting the constructive modeling methodology developed in the previous chapter. We must, therefore, state explicitly the E–frame, the R- frame and the P–frame. Within the E–frame, we adopt as our agreed upon intent the modeling of the current practice of physics. We take as fundamental the commonly defined terms of laboratory physics, treating terms denoting nonobservables as derived or theoretical terms. Our understanding of the current practice of laboratory physics is guided by the "counter paradigm" [37].

Any **elementary laboratory event**, *under circumstances which it is the task of the experimental physicist to investigate, can lead to the firing of a counter.*

In this context, by "can lead to the firing of a counter," we implicitly allow for any measurement apparatus which involves discrete and finite measures, i.e., counting. Inasmuch as all laboratory measurements are normally viewed as bound by limitations of precision and resources—which bounds for us are evidence of the intrinsic finite and discrete character of the practice—few, if any, laboratory measurements are excluded by the counter paradigm; one must make the connection to counting explicit. We take laboratory events as a sufficient set of observations to be modeled, without requiring the standard theoretical interpretation. We take as understood that an experimental (laboratory) measurement may encompass many acts of observation and, thus, that our obs may be complex (e.g., multiply-connected). In other words, we are not committed to accept the how and why of the observations, only the

observations themselves, operationally understood.* If the internal structure of an act of measurement is to be examined, then there must exist a finite procedure for carrying out the measurement (i.e., the measurement must be operational), so that the internal structure is transparent. Otherwise, we are required by Principle I to plead ignorance of the apparent internal structure.

We have now satisfied the requirements of establishing an E–frame, inasmuch as the requirements have to do with making explicit various aspects of the modeling effort. As to whether or not we are faithful to the other strictures of the E–frame, we shall leave it to the reader to decide, this being the very nature of consensual validation of the value of our effort.

6.2 ESTABLISHING THE R–FRAME

As our R–frame formalism, we adopt the ordering operator calculus. Inasmuch as quantum events, as understood within the current practice of physics, are unique, discrete, irreversible, nonlocal and yet indivisible, the principles upon which development of the ordering operator calculus was based make this an appropriate formalism.

6.3 ESTABLISHING THE P–FRAME

As our P–frame procedure, we select the algorithm given in the preceding chapter. We note in advance that some detailed aspects of the model are evolving. In particular, we are in the process of refining the specification of the d–space generator required by our formalism. This will have consequences regarding the detailed specification of any distance function on any attribute we identify. In addition, any global invariants are likely to be affected. Thus, the detailed identification of physically conserved quantities within the theory is tentative, though their existence is not.

As noted in Section 5.1, the rules of correspondence may now be elucidated in the form of a dictionary. If we establish rules of correspondence between obs from the E–frame and symbols in the R–frame, any relationship between the symbols in the R–frame must reflect relationships within the context of the E–frame, whether known at this time or not. We, therefore, adopt rules of correspondence which are more useful than current practice in relating observations to the R–frame, and then see how the practice of *discrete physics* will differ from the current wisdom. In other words, we hope to see how the E–frame (and perhaps the R–frame) should be modified. Bridgman tried long ago to get rid of the representational framework by "operational" rules of procedure that reflected directly back into the E–frame. We expect that it would be conceded by most physicists that this heroic effort failed in its initial intent, and even Bridgman was led to modify it by including "mathematical operations" within the allowed procedures. One related effort was to reduce everything in physics to "pointer readings." Our methodology is even stricter in that sense, since we require every E–frame procedure and every R–frame construct to be reducible, at least in principle, to *counting* and finitely computable algorithms. We hope to have accounted for the philosophical and technical problems which led to the failure of Bridgman's operationalism.

---

* Note the distinction between E-terms and R-terms. Von Neumanns "observation" is, at best, only an R-term. Criticism of von Neumann's representation of quantum mechanics can start there, because his R-term is not necessarily consistent with Schrodinger continuity.

Spatial Distance

For us, an attribute distance is the only thing in the R–frame that can correspond to a datum (E–frame) achieved by an experimental measurement within the practice of physics (E–frame). From the R–frame, however, we see that attribute distance has no computational meaning or significance outside the context of a particular reference frame, or without some ordering parameter (R–frame symbols). We do not make an absolute rule of correspondence between attribute distance and spatial distance; spatial distance will be a particular attribute distance. For us, however, any quantifiable experimental measurement must correspond to some attribute distance.

Cosmological and Proper Time

As noted above, we take the notion of sequence and counting in the laboratory as fundamental, so that the very character of observation in time (E–frame) is bound to the R–frame notions of counting, synchronization and both local and global ordering. We establish a rule of correspondence between laboratory proper time (E–frame) and the ordering parameter t_i (R–frame), associated with the generation of any particular reference frame F_i, via an ordering operator O_i. Similarly, we must establish a rule of correspondence between cosmological time and the global ordering T, associated with the generation of all reference frames within the model. That the global ordering may be specified in terms of the R–frame synchronization of attributes identically to the E–frame synchronization of events, establishes a requirement that events be specifiable as some particular kind of attribute. A significant portion of this section will be devoted to establishing the required nature of event attributes.

Three Dimensional Physical Space

As seen in Theorems 43 and 44, for any attribute space, no matter how simple or complex, there is some attribute which has the greatest number of attribute states of all the attributes which may be defined on the d–space. From Theorem 46, it is also clear that the corresponding attribute velocity for this attribute will be the infimum of the d–set of maximal attribute velocities. Finally, from Theorem 44 and by definition, this maximum attribute velocity will be the first bound encountered in any function involving more than one attribute. For these reasons, we identify this unique attribute velocity with the (E–frame) speed of light c, and the corresponding attribute states with the points or "4–positions" of physical space. Note that these points are events in the sense of the geometric view of general relativity.

As demonstrated in Theorem 13, for any attribute distance function, there are at most three independent runs of the ordering operator which generates these attribute states, if the global character of the d–space so generated is that it not have a preferred coordinate. Thus, the d–dimensionality of the attribute space is three, and we establish a rule of correspondence with the three–dimensionality of laboratory space.

The Global Structure of d–Space Generator

The next rule of correspondence must specify an ordering operator U, which generates the coordinate d-bases and a reference frame (R–frame) suitable for identification with the spatio-temporal reference frame (E–frame). This ordering operator U must provide the appropriate global invariances, if the identification is to be successful.

The relevant E–frame global invariances include the fundamental constants, the scale constants and the quantum numbers. For these invariances to be generated via a discrete algorithm suggests a hierarchical structure with a stop rule. For further justification of these requirements, see Bastin, 1966[38] and Bastin, 1956[39]. We may interpret the generators of each level of such a hierarchy to be coupled ordering operators; then the coupling scale may be calculated by definition, together with probabilities of coupling between the levels, which must be the coupling constants of laboratory physics.

We allow multiple, independent, but synchronized, runs of the U in order to generate a discrete space, without a preferred axis, and preserving translational invariance (i.e., having a homogenous distance function). By Theorem 13, the dimensionality of this d–space will then be three; that is, we need only three independent runs of U or any other generator of the d–space, as additional runs will not produce additional global structure. The unobservable universal (cosmological) and locally consequential (proper) time will then* be given by the universal ordering parameter associated with U.

As noted previously, an ordering operator U may be understood as generating bit strings, instead of labels which we take as abstract representations of physical attribute states. Here, we invoke the principles requiring that any specified attributes of a finite and discrete ensemble can be mapped onto an ordered sequence of 1's and 0's, by asking whether they are present or absent in a reference ensemble. Such an ordered sequence is called a **bit string**, and may combine with other sequences of the same bit length by an operation such as XOR ("exclusive or"), symmetric difference, addition (mod 2), $+_2$ .... When Noyes treats the symbols "0," "1" as bits and/or as integers, the more general *discrimination* operation "$\oplus$" defined by

$$S^a \oplus S^b \equiv (ab)_n \equiv \left[\ldots , \left(b_i^a - b_i^b \right)^2 , \ldots \right]_n$$

$$= (\ldots , b_i^a +_2 b_i^b, \ldots)_n ; \; b_i^\ell \in 0, 1 ; \; i \in 1, 2, \ldots , n ; \; \ell \in a, b, \ldots ,$$

is used. Note that discrimination meets succinctly the requirements for combining serializable ordering operators, if the bit strings are linearly independent; i.e., there is no information loss regarding the distance function on discrimination, if the resultant bit string is given a dual Hamming measure—one counts the 0's instead of the 1's—and discrimination is then a length preserving operation. For us, this is a required property of U.

U is further required, by the definition of ordering operator, to consist of an incompletely specified (though, in principle, specifiable) part, and a completely specified part. The incompletely specified part must not have an effect on the global structure, nor on the combinatoric complexity of its generation. As long as the structure generated, and the order in which it is generated, is compatible with the knowable Universe (E–frame), the unspecified part can be any algorithm, whatsoever.

* This result was anticipated conceptually by E. W. Bastin[40].

Non–Local, Discrete Events

What is now required is an R–frame definition of event. For us, this definition must follow the geometrodynamic point of view, in that the existence of an event depends upon an operation defined on strings (or similar representation of attribute states) and a distance function defined on these strings, which satisfies the so-called "triangle inequality." That is, the distance function must be a norm (see Section 2.4). Note that the definition of a norm requires a minimum of three independent strings. We establish a rule of correspondence which identifies the satisfaction of these conditions on the points of our physical space representation with the unique, nonlocal, yet indivisible and irreversible, events of quantum mechanics, since they meet the minimal conditions for nonlocalized operations on localized d–points which have a norm.

Defining $k_i^x(n) = \Sigma_{i=1}^n b_i^x$, $x \in a, b, c$, with n being the number of generations of the ordering operator, from the definition of a norm (Section 2.4), it is easy to see, for any three strings $(a)_n\,(b)_n\,(c)_n$ which satisfy the constraint $(abc)_n = (0)_n$ where $b_i^0 = 0; \in 1, \ldots, n$, that $|k^a - k^b| \leq k^c \leq k^a + k^b$ (cyclic on a, b, c) for any event. Thus, k, the number of "1's" in a string, can serve as a discrete distance function; in fact, this is just the Hamming distance. Note that the our definition of events necessarily will make them nonlocal. That is, a minimum of three independent and distance function ordered bit strings is required, although some attribute distance exists between them.

In order to locate the required reference frame "origin" (which in the R–frame corresponds to a reference ensemble) of our metric symmetrically in the finite and discrete interval allowed, we define an attribute distance $q_a \equiv f(k, n, \lambda_a)$—a linear function of k, n, λ_a, where λ_a has the dimensions of attribute distance and is identified via a rule of correspondence with a physical length. At each generation of the ordering operator, q_a changes by $\pm\lambda_a$, which we associate via a rule of correspondence with the minimum attribute distance increment of the R–frame, with the sign + or − being determined by whether a "1" or a "0" is concatenated with the extant string; i.e., whether the distance is increasing or decreasing, with respect to the reference ensemble. Note that if perfect synchronization is possible, λ_a is just $1/n$. This factor serves to normalize the distance on the [-1, 1] interval.

If we define the local event time (proper time) as a linear function of the ordering parameter $t = n\Delta t$, we see that we can define a velocity $v_a \equiv f(k, n, \lambda_a)V_x = \beta_a V_x$ where $V_x = \lambda_a/\Delta t$ is a *maximal velocity* of magnitude identified with the speed of light c, achieved when all the steps have the same sign (i.e., are in the same direction) and $f(k, n, \lambda_a)$ is a linear transformation of the Hamming distance k. We also have an *event horizon* that grows with the number of steps the generating operator has taken.

Lorentz Invariance

It is clear that q satisfies the definition of an attribute distance and satisfies β, as required in Theorems 23 and 41. We formally establish a rule of correspondence between that β and the usual β of special relativity. The specific dependence of λ on the generation of attribute states in the sequence given by the ordering operator is

unknown, and, for our purposes, not required, as long as sufficient variety is produced. From Chapter 3 (and independent of the particular generator of the d–space), we have immediately a 3+1† discrete—and locally flat—space with distance function, which is invariant with respect to the coordinate transformations of Theorems 23 and 41 and with the previously stated rule of correspondence that the maximal attribute velocity for this "position" attribute corresponds to the velocity of light c; i.e., to the minimum of the maximal attribute velocities. We now identify the coordinate transformations of Theorems 23 and 41, *when applied to the position attribute*, as the Lorentz Transformations.

That the definition of velocity is indeed a first derivate of the position q is obvious. If q is linear in t, then we have $(q/n) \times (\lambda/\Delta t)$, where $\lambda/\Delta t$ is just the "slope." If q is not linear in t, then λ is a function of t, so that we obtain $(q/n) \times (\Delta\lambda(t)/\Delta t)$, which (evaluated at some q and t) give the "instantaneous" velocities. Furthermore, not only these velocities, but any attribute velocities, thus satisfy Theorem 35, which is now identified as the relativistic composition law for velocities.

Persistence Effects and de Broglie Wavelengths

By evolution of a system, we mean that some attribute states are invariant under some transformations on the system, and nothing more. When such attribute states are jointly identified and are invariant together, we say that they constitute an "object" which persists or is stable. We now note that if we consider a system that evolves with constant velocity—i.e., by a linear d–map, $\beta_0 \equiv f(k_0, n_0, \lambda_0)$—strings which grow subject to this constraint—i.e., $n = n_T n_0$, $k = n_T k_0$, $1 \leq n_T \leq n/n_0$—will have a periodicity $T \equiv n_T \Delta t = n_T \lambda / V_x$, specifying the events in which this condition can be met. Hence, in more complicated situations, where there can be more than one "path" connecting strings with the same velocity to a single event, this event can occur only when the paths differ by an integral number of attribute distance increments. We, therefore, establish a rule of correspondence between λ and the "de Broglie wavelengths." Thus, our construction already contains the seeds of "interference" and an explanation of the "double slit experiment."

The Relativistic Doppler Shift

From Theorem 33, and independent of the particular d–space generator, we obtain the relativistic doppler shift, as required from the laboratory evidence.

Supraluminal Correlations

Because the derivations in the development of the ordering operator calculus do not depend upon any particular interpretation, particularly those which could be read as referring to "physical distance," it is clear that the principles and axioms suffice to imply relativistic and quantum effects which could be identified with physical characteristics other than distance.

† Our use of 3+1, here, is meant only to emphasize the evolution of the ordering operator which locally distinguishes the ordering parameter, and not to deny the validity of the 4–space geometric view which is globally valid after the generation has taken place.

On the face of it, this is a surprising conclusion. However, for us, it demands that we treat the universe as a multiply-connected attribute space. If it is not the case that nonspatial attribute distances behave as does the spatial attribute distance, then either conventional or discrete theory must supply some reason for this difference. To our knowledge, making such a distinction has yet to be motivated in current analyses. Clearly, not all the attributes which may be generated in a discrete space will satisfy the precise definition given for q. Therefore, regardless of the generator of the d–space, we must conclude that the d–space is multiply-connected, with the consequences derived in Theorems 43–46. We show in this section that the theory encourages us to accept as "obviously possible" the disturbing facts demonstrated by the laboratory experiments of Clauser, Frye, Aspect and others[41]. Indeed, the theory predicts that such results could be obtained for quantum attributes other than spin and polarization. These results are predicted in the following way.

Theorem 43 describes the essential character of Aspects EPR experiments, where E is electromagnetic and P is polarization, S represents the source, L the left detector and R the right detector systems. The time-of-flight experiment does not alter the model, since this only serves to verify the "instantaneous" character of the anticorrelations. The results of such experiments are readily understood in this context.

Note that supraluminal communication is not allowed, since the connection between E and P is not 1:1 and is, in fact, locally "random." Furthermore, the theory is not a hidden variable theory, nor is it a nonlocal theory in the usual sense in which these are understood. We do not provide hidden variable extensions to quantum mechanics or to special relativity in order to understand the correlations: we provide a theory which reduces to quantum mechanics or special relativity under certain restricted interpretations (e.g., the existence of the continuum). We do not postulate an absolute nonlocal quantum multiple-connectedness, as is implied, for example, by Bohm's implicate order. Neither is the multiple-connectedness like that proposed by the branching universe of Wheeler and DeWitt. Rather, we postulate a topology which admits multiple, usually independent, distance functions and metrics.

For Aspect's experiments in particular, the global relation between polarization angle and electromagnetic propagation must be identified as some cosine-squared function. This function must be independent of the electromagnetic attribute distance identified as q, but dependent upon the polarization attribute distance—i.e., the difference between the polarization angles—by hypothesis. Since the least increment for polarization angle is defined by the event horizon N (i.e., from a computation of $\pi(N)$ via the method given in Chapter 1), we may expect that the number of spatial attribute states is approximately the square of the number of polarization attribute states. This suggests that the correlations seen by Aspect will fall off as the time for propagation of changes in the optical switches approaches the square root of the propagation delay for light.

We are led by the formalism to predict that there is a correlated rate of change of the optical switch, which destroys the correlation between the arms; namely, $V(P)$. That is, when the time T between switching in one arm versus switching in the other arm is short compared to $d(P : LR)/V(P)$, the correlation should be destroyed by our analysis. An examination of the correlation with T would show stronger correlation as T approaches $d(E : LR)/V(E)$ from below. One might reasonably expect the distribution to be exponential. Unfortunately, T is likely to be extremely short for any practical distance $d(E : LR)$.

The global topology of the discrete finite attribute space is multiply-connected. There is a unique attribute which serves to define a global metric; in our case, conventional 3–space as provided by the electromagnetic attribute. Globally, our d–space is necessarily limited to 3–space. However, locally a nonisotropic n–space may be defined. That is, if we no longer require translational invariance, there is no preferential coordinate, or if synchronization is not required locally in transforming between reference frames, one may define more than three independent, short runs of the parameterized bases which will behave (locally) as coordinates.

This topology, together with the fact that events as defined have intrinsic quantum interference properties, leads one to suspect that superluminal correlations should display quantum interference; namely, the "measurement" in the right and left detectors constitute events in both E and P attribute space. Suppose that the events are arranged in such a way that they are separated in E–space, but not in P–space. Furthermore, suppose that in P–space the events have wavelengths such that interference can occur. This interference should then modulate the correlation in E–space. Such a "correlation interference pattern" would be striking evidence of the proposed topology, since this cannot occur in distant (in E) events in the conventional theories.

Computer Models

We may model our system with the required topology on a computer* In particular, the violation of Bell's Inequalities and related effects may be demonstrated in the computer model, since our formalism is strictly computable. Care must be taken in establishing the functional connection between E and P in the computer model, however. The connection must be sufficiently complex computationally to lead to the appearance of local (i.e., restricted memory) "random" behavior. This is just the problem of precision in computer modeling, used in reverse to establish certain statistical properties of the model. Indeed, it would appear that the model may be set up to demonstrate physical supraluminal correlations between physically separated computer systems in a distributed processing, shared memory environment.†

⋆ As has been partially done for a particular 3–space generator.[42]

† Related Work: The relationship between this model and cryptographic techniques is interesting as well. A recent paper by Goldreich[43] considers a constructive approach to random bit strings based on computational complexity which is similar, though more specific and restrictive than that introduced in the present paper. In particular, the authors introduce programs that run in polynomial time and which lead to identical results when fed with either a set S of strings or elements randomly selected from the set of all strings.

Such poly-random collections can be shown to enable many parties to share efficiently a random function f in a distributed environment, by which we mean that if f is evaluated at different times by different parties on the same argument x, the same value $f(x)$ will be obtained. Such sharing can be achieved by selecting k-bits to specify a function in a poly-random collection. These k-bits are then communicated to and stored by each party. No further messages need be exchanged between parties to share f. It is a trivial matter to make the sharing either correlated or anticorrelated if $f(x)$ is two-valued. The physical communication of the k-bits may be dispensed with in a multiply-connected attribute space, as the k-bits may be "local" through some particular attribute. Thus, the k-bits are always available in "local" shared memory.

Mass and the Law of Relativistic Mass Change

We can associate a parameter m with the total size S [Eq. (44)] of the ensemble, and establish a rule of correspondence which identifies m (R–frame) with *mass equivalent* or *energy* (E–frame). Note that we differentiate between the mass and the energy. For a bit string in an evolving system to have an invariant mass at constant attribute velocity, the mass may be defined as the energy divided by some normalization factor, which depends on the cardinality of the attribute states which might be generated, and on the cardinality of the attribute Universe (R–frame). In this way, adding a distinguishable state (a '1') to the Universe and to the bit string do not alter the "mass" parameter in a measurable way, and results in a statistically invariant mass. For consistency with our finite principle, we must require $0 < k < 1$; thus, no massive event can lie on the event horizon.‡ Independent of the particular generator of the d–space, Theorem 34 is interpreted as showing that the definition of this parameter follows the law of relativistic mass change.

Momentum Conserving Events

We require the existence of a norm for an attribute which can be identified with momentum, and in this way obtain momentum conservation. Once we have shown that the attributes of position and momentum can be identified (or equivalently, position, velocity and an invariant mass), and a norm in each of these spaces defined for a configuration which we identify as a quantum mechanical event, the generator of the d–space can be any algorithm whatsoever.

Defining $p_a \equiv m_a v_a = m_a \beta_a V_x = \beta_a m_a \lambda_a / \Delta t$ and establishing the rule of correspondence which identifies this as momentum, we see that $|p_a - p_b| \leq p_c \leq p_a + p_b$, provided only (as is required for consistency) $m_a \lambda_a / \Delta t$ is any finite constant independent of a. Thus, there is a norm in momentum attribute space. As Noyes would put it, the "triangle" thus closes in "momentum space," as well as "configuration space." Our d-events can now be interpreted as 3–momentum conserving, 3–particle scattering events in the zero momentum frame, with the "center-of-mass" of laboratory physics at rest.

These results are, of course, familiar in terms of so-called public key encryption systems. Here, a public key is distributed for encryption of messages to the key distributor. Although the encryption key is public, the cryptographic function does not allow decryption without access to the private key. And the number of possible private keys is too large to be determined by trial and error.

Actually, the entire scheme of shared random number generators has been put into effect. One can purchase a plastic card which contains a microprocessor. This processor produces an apparently random sequence of bit strings. When interrogated by a system which shares the random function, a match is produced and thus the card serves as a "key." Each card contains a k-bit code for the particular function and this serves to identify the particular user. Clearly, two cards with the same k-bit code would be perfectly correlated regardless of separation and yet would produce apparently random output.

‡ From the definition of maximal attribute velocity, we should be led to the mass conversion law.

Zitterbewegung

We have already seen that any system with "constant velocity" (i.e., at those generations of the ordering operator when events can occur) evolves by discrete increments $\pm\lambda$ in q *between* d–events. These steps occur in the *void* where space and time are undefined. Since $\lambda/\Delta t = V_x$, each step occurs forward or backward with the limiting velocity. Thus, we deduce a discrete *Zitterbewegung* from our theory. If we think of this as a "trajectory" in the traditional pq phase space, each time step induces a step $\pm\lambda$ in q correlated with a step $\pm mV_x$ in p. Even in the case of a particle "at rest," this must be followed by two steps of the opposite sign to return the system to "rest;" see Figure 18.

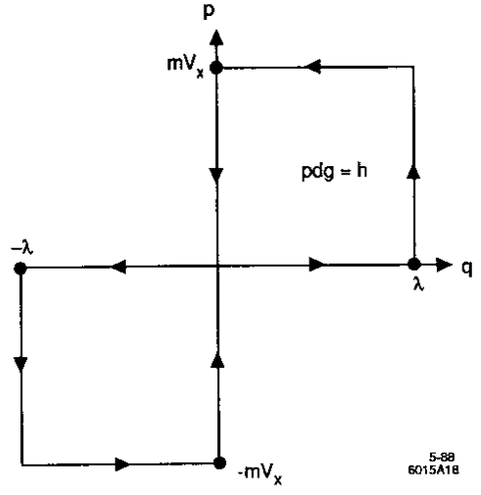


Figure 18 Zitterbewegung in phase space for a particle "at rest."

Thus, there is, minimally, a four-fold symmetry to the "trajectory" in phase space, corresponding to the generation periodicity we discovered above.

Commutation Relations, Uncertainty, Planck's Constant

From the E–frame definitions of the obs corresponding to p and q, and consistent with the present example, we see that p and q are not independent. It follows from Theorem 50 that p and q do not commute, and from Theorem 42 that there is an uncertainty associated with the product of the variances in p and q. We establish a rule of correspondence between the constant in Eq. (76) and Plancks constant. By definition, the least step in p is just mc, since this step occurs at the maximal attribute velocity. Once again, these results are independent of the particular d–space generator chosen.

Since the least change in the product of the variances is h by the rule of correspondence, it follows that the least step in q is appropriately identified as just $L = h/mc$. To go on to the commutation relations, we take the usual step in the geometrical description of periodic functions, of taking the qJJ plane to be the complex plane $(q, 2\pi ip)$; then the steps around the cycle in the order $qpqp$ are proportional to $\pm 2\pi(1, i, -1, -i)$, where $\pm$ depends on whether the first step is in the positive or negative direction or, equivalently, whether the circulation is counterclockwise or clockwise.

We have now shown that $qp - pq = \pm i\hbar$ for free particles; this result holds for any theory which uses a discrete free particle basis.

The Angular Momentum Commutation Relations

Going to three dimensions, the commutation relations for angular momentum (as usually defined) follow immediately. Following T. F. Jordan[44], we may now derive the angular momentum commutation relations. Suppose we have P and Q in a discrete 3–space (i, j, k), related by a basis vector L, which we will call the angular momentum:

$$L = Q \times P \ ,$$

which is shorthand for three equations

$$L^i = Q^j P^k - P^j Q^k \ ,$$

with i, j, k taking all values from 1 to 3, and not equal to each other.

From the previous derivation of the P, Q commutation relations, we have

$$Q^j P^j - P^j Q^j = \frac{ih}{2\pi(N)} \ ,$$

$$L^i \times L^j = \frac{ihL^k}{2\pi(N)} \ .$$

For example,

$$\begin{aligned} L^1L^2 - L^2L^1 &= (Q^2P^3 - Q^3P^2)(Q^3P^1 - P^1Q^3) - (Q^3P^1 - Q^1P^2)(Q^2P^3 - Q^3P^2) \ , \\ &= Q^2P^3Q^3P^1 + Q^3P^2Q^1P^3 - Q^3P^1Q^2P^3 - Q^1P^3Q^3P^2 \ , \\ &= Q^1P^2(Q^3P^3 - P^3Q^3) + Q^2P^1(P^3Q^3 - P^3Q^3) \ , \\ &= (Q^1P^2 - Q^2P^1)\frac{ih}{2\pi(N)} = \frac{ihL^3}{2\pi(N)} \ . \end{aligned}$$

Similar results follow for each of the relationships involving other coordinates (Q^1, Q^2, Q^3, P^1, P^2, P^3, L^1, L^2, L^3).

We have now shown that

$$L^i \times L^j = \frac{ihL^k}{2\pi(N)} \ ,$$

for free particles; this results holds for any theory which uses a discrete free particle basis.

Complete Identification of Laboratory Units

Now that we have shown, once given a specific generator of the 3+1–space, how to compute two ($\hbar$ and c) of the three dimensional constants needed to connect a fundamental theory to experiment in the 3–space in which physics operates, and which we have proved must be the asymptotic space of our theory, all that remains is to determine a unit of mass. Theorem 34 allows us to specify that the mass of an object is just the size S, although it does not tell us what object determines the fundamental unit. This can only be done once a specific generator of the d–space has been selected.

Scattering Range Computation

Once a unit of mass has been identified, we can show how to compute the classico-quantum scattering range from attribute distance. Note that for $h/(2m_pc)$, from the existing rules of correspondence for c, m_p and h, one obtains the following. Define an attribute such that the minimum attribute distance increment is I, with the following definitions holding: $h = I^2$ (minimum possible "area" in "phase space," $m_p = S = I + D$, and $c = v_{max} = I - D/I + D$. Thus, $h/(2m_pc) = I^2/2(I - D)$ when $v = c = v_max$; i.e. when $D = 0$. Therefore, we have $I^2/2(I) = I/2$, where I is just the minimum attribute distance increment for the attribute corresponding to the ensemble A with invariant size (mass) $I+D$. Clearly, since the 3–space is homogenous, we may interpret I as a diameter. Suppose a second ensemble B "approaches" with the first. Take two cases for the (generalized) attribute distance between them; $r > I$ and $r \leq I$. If $r > I$, then ensemble A may "travel" a distance I without any states in the generalized attribute distance being shared with the attribute states of B. If, however, $r \leq I$, then there exists the possibility of shared generalized attribute states between A and B, and thus nonindependence, exactly as described in the explanation of commutivity.

Note further, that if ensembles A and B do not have the same size and the same attribute distance definition for velocity computation, then the minimum interaction distance is not just the minimum of the "minimum attribute distance increments" for A and B, as compared via the generalized attribute distance. This is because the operation of addition is no longer well-defined: ensembles A and B are no longer independent, and this alters the generalized attribute distance definition.

Along other lines, we also have indications of how to compute transition probabilities from the ratios of the number of ensembles in given states, as determined by the combinatorial hierarchy. Let there be two attributes, such that the enumeration of states generated by the corresponding ordering operators are just the inverse enumeration of each other; that is, the last state generated by one is indistinguishable from the first state generated by the other, the next to the last state indistinguishable from the second state, etc. Further, let the representation of the states be duals ('0' in one represents the same thing as '1' in the other, and vice versa). From our rules of correspondence, these then correspond to particle and antiparticle.

This geometry suggests that zero mass particles are anomalous: no photon can be observed without both emission and absorption, and the path length in the photon frame is zero.* In the rest frame of the photon, any point on the photon trajectory can be treated as an electron/positron pair without violating relativity or the conservation laws. It would appear that photon emission/absorption is then modeled in our formalism as an electron emitted by the "emitter" and a positron (i.e., electron traveling backward in time) emitted by the "absorber," so that the photon can be treated as a virtual particle. From the reference frame of the photon, this exchange, and the evolution of the corresponding state vector, takes place atemporally. It is outside of time, happening everywhere along the photon path "at once." There is a difference in the energy of the two ends of the trajectory which is given by the torsion of the space—this being related to the constant identified above as Planck's constant, and to the minimal attribute distance increment exactly as in the (five-dimensional) Kaluza–Klein model. Thus, there is an apparent "transfer of energy" in the electron/positron pair exchange. This structure can *not* be detected locally. A similar argument holds for massless particles, in general.

6.4 Related Results: The Combinatorial Hierarchy and Program Universe

Bastin, Kilmister, Amson, Noyes and Parker–Rhodes have shown that there exists a unique finite hierarchy, combinatorially generated, which constructs at least some of the properties we require. This structure is referred to in the literature as the **combinatorial hierarchy**. Without developing the details here, we point out the essential features which make this structure interesting. First, the cardinalities of the primary objects (discriminately closed subsets) at each level of the structure are identifiable with the number of (E–frame) objects which may participate in the fundamental forces: to first order (which in our terms assumes first degree coupling only), they are the scale constants of laboratory physics (which we would identify computationally with the coupling scale of the relevant ordering operators). Second,

* This is just the Wheeler–Feynman rule, as was pointed out to us by H. P. Noyes. Indeed, the work of Cramer's transactional interpretation is in full agreement, and is an extension of the Wheeler–Feynman interpretation. That such an interpretation results in a time-symmetric, self-renormalizing QED with no singularities or second-quantization problem is indeed encouraging[43].

Parker–Rhodes has shown that the construction leads to an amazingly accurate *computation* of the ratio of the mass of the proton to the mass of the electron (consistent with the present work). Third, Noyes et al. have developed a particular algorithm, known as **Program Universe**, for generating the combinatorial hierarchy, and have shown that the quantum numbers may be specified in such a manner as to make appropriate identification with the first generation of leptons and quarks. We refer to this algorithm as *PU*.

This last algorithm is of particular interest for our purposes, since it has all the characteristics of an ordering operator, including the fact that it is too complex to be deterministicly knowable from partial generation. The algorithm has two degrees of freedom, that is, two points at which an appeal to an arbitrariness generator is necessary. These two steps in the algorithm do not affect the global structure of the combinatorial hierarchy thus ultimately produced. Rather, the specifics of these steps will determine the dynamic evolution of the structure and the statistics during this evolution. Once the structure has been completely generated, the statistics are no longer affected.

For these reasons, we point out that PU[46] is, as an algorithmic definition, exemplary of the type of ordering operator which will generate the three-dimensional d–space, as described in Theorem 13. We caution the reader, however, to keep in mind that PU, the specific distance functions which are defined on it and the related derivations *are simply an example* of how we may proceed in detail. We identify U with *PU*, subject to falsification and subsequent modification. We are not dependent upon these details for the results presented here, which deal primarily with a physical interpretation of the ordering operator calculus. Nonetheless, we believe that either the details are valid, or that these aspects of the model can evolve smoothly (via the P–frame) to become valid.

For example, PU generates a universe of such strings which grows, sequentially, in either number (SU) or length (NU). The main program starts with PICK, an arbitrariness generator that picks two arbitrary strings from memory and discriminates them. This is one of the degrees of freedom mentioned. If this produces a novel string, an operation called ADJOIN results, which adjoins the string to the universe (SU:=SU+1). If the string produced by PICK is already in the universe, an arbitrariness generator called TICK is triggered which increases each string independently, by concatenating it with one arbitrary bit (NU:=NU+1). After either ADJOIN or TICK, the algorithm then recourses to PICK. The arbitrariness which occurs in selected strings from memory (in PICK) or in selecting bits to concatenate (in TICK), serves to guarantee that the algorithm represented by Program Universe is incompletely specified (though in principle specifiable) and, hence, we may treat the output as a Bernoulli trial (as required by Theorem 13), and PU as an ordering operator. If these are fully specified in an algorithmic sense, PU becomes deterministic, and the full evolution of the cosmology becomes known. However, much of the phenomena of laboratory physics arises specifically because we do not have the information. Indeed, we claim that the finite system represented by laboratory physics lacks the space complexity required to fully represent such an algorithm. Thus, some free parameters in the algorithm may not be determined from the recorded output of PU to date. At best then, PU represents a class of algorithms, each of which is sufficient, but not necessary, to account for the phenomena of laboratory physics. We propose that the class encompasses the necessary conditions.

That PU meets the conditions outlined in previous chapters for an ordering operator which is a metric generator, is easy to see. When the operation TICK of PU occurs, there will be three strings connected with the generation process which satisfy the conditions

$$S^a \oplus S^b \oplus S^c = (0, 0, ..., 0)_{NU} \ .$$

When NU is large, these conditions will be satisfied by many combinations. We can now identify the free function $f(k, n, \lambda)$, presented in the discussion entitled Nonlocal, Discrete Events in defining the attribute distance q_a, subject to a possible scaling factor. For PU, $f(k, n, \lambda) \equiv [2k^a(n) - n]\lambda_a$, and the conditions required in the preceding paragraphs are satisfied automatically. Therefore, PU is consistent with—and can legitimately appeal to—the results presented in this paper, without further derivation.

In an earlier work by Noyes et al.[47], a propagator for relativistic quantum scattering theory was derived. Now that we have shown how to explicitly construct the commutation relations, the interpretation or use of complex notation and how to construct the exponentiation operator, we claim that this work is well founded in all its detail.

Noyes has subsequently shown how to provide the interaction terms of the theory, by identifying our 3–momentum conserving events as "Yukawa vertices." Additionally, a tentative identification has been given of the first three levels of the hierarchy with (1) chiral electron-type neutrinos, (2) electrons, positrons and photons and (3) up and down quarks in a color octet, and with level four to provide weak-electromagnetic unification, with weak coupling to the first three levels.

That the overall mass scheme should come out right, is clearly suggested by the success of the Parker Rhodes calculation:[48] $m_p/m_e = 137\pi/[(3/14)[1 + 2/7 + (2/7)^2](4/5)] = 1836.151497 \ldots$, which was later reformulated by Noyes to be consistent with the present theory. As Noyes has pointed out[49], the cosmology of Program Universe appears to have a charged lepton and a baryon number consistent with current observation, and, hence, with a locally flat space. These results can be understood as following immediately from establishing rules of correspondence between laboratory practice in high energy physics and performing the appropriate computations. Indeed, this author believes that there are few degrees of freedom available in establishing that interpretation, and perhaps none whatsoever. For example, if PU is selected, we must compute the largest to the smallest mass ratio; but this has already been done for us by the combinatorial hierarchy result $2^{127} + 136 \simeq 1.7 \times 10^{38} \simeq \hbar c/Gm_p^2 = (M_{Planck}/m_p)^2$, which tells us that we can either identify the unit of mass in the theory as the proton mass—in which case we can calculate, to about 1% in this first approximation, Newton's gravitational constant—or, if we take the Planck mass as fundamental, calculate the proton mass.

CONCLUSIONS

The ordering operator calculus has provided a formalism compatible with, and having explanatory and predictive power regarding, the current practice of physics. Indeed, a discrete and unified model of quantum mechanics and special relativity has been made possible.

Much work remains to be done. Not only is considerable effort required in establishing and validating the rules of correspondence, but extensions of the ordering operator calculus to other domains of mathematical investigation are desirable—we have mentioned some of these efforts along the way—and, of course, we would clearly like to incorporate a discrete version of general relativity in our theory. We have laid the foundation for doing so with the definitions of manifolds, neighborhoods, one-forms and other relevant mathematical objects. The reader should note that ours is always a "locally Lorentz invariant" theory and that local frames are, by construction, "inertial," meaning that the geometry is locally flat and exhibits no accelerations. Indeed, accelerations can only arise between the kinds of events we have constructed nonlocally, via the global topology (the connection), even though any dynamics are completely determined from the local geometry. Also in keeping with the geometric picture, our coordinate space has been constructed (from the beginning) from attribute "events," which locate an event by "what happens there," the ordering operator calculus being context sensitive. We already have some indication that "local (gravitational) distortion" of our distance function by a mass can be shown,and work we have recently encountered in the domain of cellular automata is relevant to, our corresponding notion of a field.

A number of experimental predictions have been made. According to P. Suppes[50], there are many generalized inequalities concerning joint probabilities, among which Bell's Inequalities are but a specific example. We have suggested a means of using these inequalities to test whether the nonlocality which violation of the inequalities demonstrates is absolute (along the lines of Bohm's Implicate Order), or, in fact, due to a multiply-connected topology.

We also suggest several other tests of the topology. Our theory predicts that the correlation in Aspect's time-of-flight experiments must be sensitive to the time between changes in the randomly shifted Brewster mirrors, and that the correlation will disappear for data taken arbitrarily close in time to one or the other shift. We should also be able to calculate the shape of an expected distribution curve for the fall-off in correlation, and might be able measure the slope experimentally. These experiments will be quite difficult because of the accuracy in measurement required.

Finally, we have suggested that this phenomena is NOT necessarily microscopic, or limited to spin and polarization quantum variables. The theory is sufficiently general that macroscopic violations of Bell's Inequalities should be constructable. Certainly, the effect can be modeled on computers and, indeed, is used today in publicly key encrypted security (access) cards.

As pointed out in the introduction, the ordering operator calculus is intended as a formalism for modeling diverse phenomena, and not just physical phenomena. Work along these lines is proceeding, and as yet unpublished applications to computational linguistics and computer science have been quite successful.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of the following, without whom this work could not have progressed: Tom Etter, C. W. Kilmister, Ted Bastin, John Amson, M. J. Manthey, C. Gefwert and, finally, our deceased colleague, F. Parker–Rhodes.

REFERENCES

1. Godel, K., *On Formally Undecidable Propositions of Principia Mathematica and Related Systems I*, in Von Hiejenoort, J. ed., *From Frege to Godel: A Sourcebook in Mathematical Logic, 1879–1931*, Harvard University Press, Cambridge, MA (1967), pp. 592–616.
2. Dirac, P. A. M., *Principles of Quantum Mechanics*, Clarendon Press (1981), p. 48.
3. Bishop, E. and Bridges, D., *Constructive Analysis*, Springer–Verlag (1985), p. 2.
4. Bishop, E., op. cit., pp. 4–13.
5. Bishop, E., op. cit.
6. Bishop, E., op. cit.
7. Bishop, E., op. cit.
8. Beeson, M., *Foundations of Constructive Mathematics*, New York, Springer–Verlag (1985).
9. Parker–Rhodes, A. F., *Theory of Indistinguishables*, D. Reidel, Dordrecht, Holland (1981).
10. Kilmister, C. W., *The Mathematics Needed for Self-Organization*, in *Disequilibrium and Self-Organization*, C.W.K., ed., (Reidel, 1986), pp. 11–17.
11. Noyes H. P., C. Gefwert and M. J. Manthey, in *Symposium on the Foundation of Modern Physics*, June 16–20, 1985, Joensuu, Finland; Lahti, P. and Mittelstaedt, P., eds., pp. 511–524 (World Scientific, Singapore); see also *Toward a Constructive Physics*, SLAC–PUB–3116 (1983) and references therein.
12. Parker–Rhodes, A. F., op. cit., p. 66.
13. Blizzard, W. D., *Theory of Multisets with Indiscernables*, Oxford University (1987).
14. Birkhoff, G., Lattice Theory, Amer. Math. Soc., Providence, Rhode Island (1967).
15. Conway, J. H., *On Numbers and Games*, Academic Press, London (1976).
16. McGoveran, D., *The Importance of Being Discrete: Four Alegories* (1984) unpublished.
17. Kolmogorov, A., *Three Approaches to the Concept of "The Amount of Information."*, Prob. Inf. Trnsm. 1, 1 (1965).
18. Chaiten, G. J., *On the Length of Programs for Computing Finite Binary Sequences*, J. ACM 13, 4 (1966), pp. 547–570.
19. Milne–Thomson, L. M., *The Calculus of Finite Differences*, Chelsea Publishing Co., New York, 2nd ed. (1981).
20. McGoveran, D., *Getting Into Paradox*, 1981, in Proceedings, ANPA West (1985).

21. Huffman, D., *A Method for the Construction of Minimum Redundancy Codes*, Proc. IRE, vol. 40 (1952) pp. 1098–1101.
22. Wootters, Willian K., *The Acquisition of Information from Quantum Measurements*, Thesis, University of Texas at Austin (1980).
23. Feller, W., *Probability Theory and Its Applications*, Vol. 1, Wiley (1950) p. 247.
24. Stapp, H., to Noyes, H. P., private communication (1987).
25. Milne–Thomson, L. M., op. cit.
26. Milne–Thomson, L. M., op. cit., p. 23.
27. Box, G., Hunter, W. and Hunter, J., *Statistics for Experimenters*, Wiley Interscience, New York (1978) pp. 43–50 and 123–131.
28. Box, G., Hunter, W. and Hunter, J., op. cit., pp. 43–50.
29. Feller, W., op. cit., p. 179.
30. Stein, I., seminar at Stanford (1978) and papers at ANPA 2 (1980) and ANPA 3 (1981), King's College, Cambridge.
31. Schutz, B., *Geometrical Methods of Mathematical Physics*, Cambridge University Press (1980) pp. 43–47.
32. Goldstein, H., *Classical Mechanics*, Addison–Wesley, 2nd ed., (1980); see the discussion on cyclic coordinates and angle variables in physical systems, pp. 445–471.
33. Bastin, Ted and Kilmister, C. W., *Fields and Particles in a Discrete Physical Model*, Work paper (1973).
34. Kuhn, T. S., *The Structure of Scientific Revolutions*, University of Chicago Press (1962).
35. Kuhn, T., op. cit.
36. Suppe, F,. ed., *The Structure of Scientific Theories*, University of Illinois Press (1979) pp. 191–192.
37. Noyes, H. P., Gefwert, C. and Manthey, M. J., *A Research Program with No Measurement Problem*, in *New Techniques and Ideas in Quantum Measurement Theory*, Greenberger, D. M., ed., Annals of the New York Academy of Sciences, Vol. 480 (1986) pp. 553f.
38. Bastin, E. W., *On the Origin of the Scale-Constants of Physics*, Studia Philosophica Gandensia, v. 4 (1966) pp. 77.
39. Bastin, E. W., *A General Property of Hierarchies*, Proc. Royal Camb. Soc. (1956).
40. Bastin, E. W., op. cit.
41. Clauser, J. F. and Shimony, A., Rep. Prog. Phys., v. 41 (1970) pp. 1881f.
42. Manthey, M. J., McGoveran, D. and Noyes, H. P, in preparation.
43. Goldreich et al., *How to Construct Random Functions*, Journal of the Association for Computing Machinery, **33**, No. 4 (1986) pp. 792–807.

44. Jordan, T. F. *Quantum Mechanics In Simple Matrix Form*, John Wiley and Sons, New York (1986) p. 161.

45. Cramer, J. G, *The Transactional Interpretation of Quantum Mechanics*, Rev. Mod. Phys. Vol. 58, No. 3 (1986).

46. Noyes, H. P., Gefwert, C. and Manthey, M. J., in *Symposium on the Foundation of Modern Physics*, June 16–20, 1985, Joensuu, Finland, Lahti, P and Mittelstaedt, P., eds., World Scientific, Singapore, pp. 511–524; see also *Toward a Constructive Physics*, SLAC–PUB–3116 (1983) and references therein.

47. Noyes, H. P., Gefwert, C. and Manthey, M. J., *A Research Program with No 'Measurement Problem'*, Annals of the N. Y. Acad. Sci. **480** (1986) pp. 553–563.

48. Parker–Rhodes, A, F,m op. cit., pp. 184–185

49. Noyes, H. P., McGoveran, D. O. et al., *A Discrete Relativistic Quantum Physics*, SLAC–PUB–4327 (1987).

50. Suppes, P. and Zanotti, M., *New Bell-type Inequalities for $N > 4$ Necessary for the Existence of a Hidden Variable*, submitted to Phys. Rev. Letters (1987).