

# PURE DATA SPACES

SAUL YOUSSEF  
Department of Physics  
Boston University

August 15, 2025

## **Abstract**

Pure data is a proposed axiomatic foundation for mathematics and computing based on *finite sequence* as the foundational concept rather than based on *logic* or *types*. The basic theory of pure data is reviewed. Within this theory, minimal requirements to have set-like collections of mathematical objects leads to the definition of a *pure data space*. A theory of pure data spaces and their morphisms is developed where a space is best understood as a semiring of endomorphisms with homomorphisms representing structure, and with subspace endomorphisms representing new and possibly very different substructures. As an exploration, and in order to illustrate these concepts, pure data spaces are “grown organically” from the substrate of pure data with the minimal combinatorial definitions which follow from the basic concept of a finite sequence. Familiar objects from classical mathematics emerge this way, including natural numbers, integers, rational numbers, boolean spaces, matrix algebras, Gaussian Integers, integer Quaternions, and even non-associative algebras like the Integer Octonions and Sedenions. Insights from these examples suggests new directions in theory and new exploration.

## 1 Introduction

In *Pure Data Foundation of Mathematics and Computing*[?], we propose a framework based on *finite sequence* as the foundational concept, rather than based on *logic*, as in Set Theory or based on *type*, as in dependent type theories. In this framework, pure data acquires meaning via definitions as explained below. A collection of definitions defines equality of data which, therefore, relates mathematical quantities to each other. This makes for a drastically simple relationship between proof and computation. Any sequence  $A_1=A_2=\dots=A_n$  is both a valid proof of  $A_n$  given  $A_1$  and a valid computation of  $A_n$  starting with  $A_1$ .

Reference [?] mainly developed consistency, logic, and the resolution of paradoxes. A goal of this work is to begin the exploration of classical mathematics through the lens of pure data. In particular, doing mathematics in general seems to require considering collections of objects as a set or as a category. In the pure data context, we find that minimal requirements for data to contain a collection of other data leads to a simple conceptual definition of a “space.” A space is any data which is “associative.” Although a space is like a set in the sense that it is a minimal collection, we surprisingly find that spaces have a rich internal structure which is shared by everything from a single atom to the entire space of pure data. Understanding spaces appears to be the key to better understanding the relationship of this framework to classical mathematics.

In Sections 1-4, we review the axiomatic framework and establish a suitable minimal set of starting definitions. Section 6 develops the theory of spaces, morphisms of spaces, semirings, semialgebras and fields. Sections 7-14 are an exploration of the spaces which “organically” emerge

from pure data, and how they relate to familiar mathematical objects. Section 15 summarizes and suggests research directions.

## 2 Foundation

In [?], we argue that mathematics and computing can both be captured within a small axiomatic framework, based on **finite sequence** as the foundational concept. Assuming that finite sequences are understood, *data* and *coda* are defined by

**Definition 1.** **Data** is a finite sequence of **codas**, where each **coda** is a pair of **data**.

The two foundational operations defined on data are 1) concatenation of data  $A$  and data  $B$ , written ' $A B$ ', and 2) pairing of data  $A$  and data  $B$  as a coda, written with a colon ' $A : B$ '. By Definition 1, the empty sequence, written ' $()$ ', qualifies as data and, therefore,  $()$  paired with itself is a coda, written ' $((():()))$ ' or ' $(():)$ '. Any finite sequence of codas is data, so, for example,  $(():)(:(():))((():)(:(():)))$  is data consisting of a sequence of three codas. We can think of this as *pure data* since it is "data made of nothing." By convention, the colon binds from the right first and binds less strongly than concatenation, so that  $A : B : C$  is defined to be  $(A : (B : C))$  and  $A : B C$  is defined to be  $(A : (B C))$ . Data is typically written with upper case, and codas are typically written in lower case. To indicate the left and right data of a coda, we sometimes use L/R superscripts so, for any coda  $c$ ,  $c = (c^L : c^R)$ .

All meaning within the system is determined by a chosen partial function from coda to data called a *context*. Contexts are partially ordered by inclusion, so if  $\delta$  and  $\delta'$  are contexts,  $\delta \leq \delta'$  if  $\delta$  and  $\delta'$  are equal on the domain of  $\delta$ . Given a context  $\delta$ , equality of data is the equivalence generated by  $c \sim \delta(c)$  for any coda  $c$ , and by compatibility with concatenation and colon in the sense that if  $A \sim B$ , then  $(A X) \sim (B X)$ ,  $(X A) \sim (X B)$ ,  $(A : X) \sim (B : X)$  and  $(X : A) \sim (X : B)$  for any data  $X$ . This relation  $A \sim B$  is denoted  $A \stackrel{\delta}{=} B$ , or simply  $A = B$  when the context is unambiguous. It follows that if  $\delta \leq \delta'$ , then  $A \stackrel{\delta}{=} B$  implies  $A \stackrel{\delta'}{=} B$ . We say that if  $A$  and  $B$  are equal then they are "always equal," thinking of moving from  $\delta$  to  $\delta'$  as the passage of "time." The fixed points of a context  $\delta$  are *atoms*. Note that if  $c$  is an atom in context  $\delta$ , and if  $\delta \leq \delta'$ , then  $c$  is still an atom in context  $\delta'$ . Thus, atoms are "permanent" and if  $c$  is an atom,  $c$  is "always" an atom.

**Definition 2.** Within a given context  $\delta$ , coda  $c$  is an **fixed point** if  $\delta : c \mapsto c$ . A coda equal to a fixed point is an **atom**. If data  $A$  contains an atom in its sequence,  $A$  is **atomic** data. Data  $A$  is **invariant** if every coda  $a$  in the sequence of  $A$  is an atom and if  $a^L$  and  $a^R$  are both **invariant**.

It follows from the definition that empty data is invariant and atoms  $a$  and  $b$  are equal if and only if  $a^L = b^L$  and  $a^R = b^R$ . If  $a$  and  $b$  are atoms,  $A$  and  $B$  are data, then  $(a A) = (b B)$  and  $(A a) = (B b)$ , if and only if  $a = b$  and  $A = B$ . If  $A = B$  and  $A$  and  $B$  are invariant, then  $A$  and  $B$  are identical as pure data.

## 3 Genesis

In the proposed framework, all mathematical objects are pure data and all meaning is defined by a chosen context  $\delta$  and its corresponding equivalence. A **proof** in  $\delta$ , for example, is just a sequence  $A_1 \stackrel{\delta}{=} A_2 \stackrel{\delta}{=} \dots \stackrel{\delta}{=} A_n$  which can be viewed either as proving  $A_n$  given  $A_1$  or as a computation  $A_1 \mapsto A_n$ . The immediate issue is how to choose  $\delta$ .

In the beginning, there are no definitions, and the corresponding context is the empty partial function  $\delta_0$ . Since the domain of  $\delta_0$  is empty, it has no fixed points and, therefore no atoms. Thus, the empty sequence is the only invariant,  $X \xrightarrow{\delta} X$  is the only valid proof, and the only valid computations do nothing  $X \mapsto X$  for any data  $X$ . To define a non-empty context  $\delta_0 \leq \delta$ , we need, at least, an invariant specification of the domain of  $\delta$  as a partial function. Since the empty sequence is the only invariant, the only way to specify if a coda  $(A:B)$  is to be in the domain of  $\delta$  is to require either A or B or both to be the empty sequence. In each of these cases, the coda  $(:)$  is within the domain of  $\delta$ , and so we must, in any case, decide on what  $(:)$  maps to. There are three possibilities;

choice 1:  $\delta : (:) \mapsto ()$ ,

choice 2:  $\delta : (:) \mapsto (:)$ , or

choice 3:  $\delta : (:) \mapsto \text{anything other than } () \text{ or } (:)$ .

Since pure data is “made of  $(:)$ ”, choice 1 trivially causes all data to be equal to the empty sequence. On the other hand, choice 3 would mean that for any non-empty data  $A$ , the number of  $(:)$  atoms in  $A$  grows without limit. This would mean, for instance, that no computation could produce a final result. Thus, we are constrained to choice 2, where  $(:)$  is an invariant atom in  $\delta$ , and, therefore an atom in any context  $\delta \leq \delta'$  as well. It is convenient to generalize choice 2 and to let  $\delta:(:X) \mapsto (:X)$  for all data  $X$ , since this provides a supply of invariant atoms  $(:)$ ,  $(:(:))$ ,  $(:(:)(:))$ , ... which can be used to expand the domain of  $\delta$ .

We adopt a convention for expanding the domain of  $\delta$  while preserving context partial ordering. A context of the form

$$\delta_a : (a \ A : B) \mapsto \delta_a(A, B) \quad (1)$$

for some invariant atom  $a$  is called a **definition**. We conventionally restrict ourselves to contexts  $\delta \cup \delta_a \cup \delta_b \cup \dots$  where  $a$  is an invariant atom in  $\delta$ ,  $b$  is an invariant atom in  $\delta \cup \delta_a$ ,  $c$  is an invariant atom in  $\delta \cup \delta_a \cup \delta_b$  and so forth. By requiring  $a, b, \dots$  to be disjoint, we maintain the context partial ordering.

## 4 Definitions

Before proceeding to mathematical exploration, we need to add enough expressive power to the original context  $\delta$ . To do this, we introduce minimal definitions using the mechanism explained in the previous section.

1. Define “bits”, “bytes” and “byte sequences” so that words are single atoms, so that  $a$  can be a word in future definitions  $\delta_a$ .
2. Define all of the naturally occurring combinatoric operations given two finite sequences  $A$  and  $B$ . For example  $\delta_{\text{ap}}$  maps  $(\text{ap } A : b_1 \ b_2 \ \dots \ b_n)$  to  $(A:b_1) \ (A:b_2) \ \dots \ (A:b_n)$ , so that  $\text{ap}$  “applies  $A$  to each atom of input.”
3. Define a minimal internal language  $\delta_{\{\}}$  so that a coda  $(\{\text{language expression}\} \ A : B)$  is mapped to data as described by the expression and possibly using data  $A$  and data  $B$ .
4. Include definitions giving direct access to context-level values. This include a Boolean definition  $\delta_{\text{bool}}$  to determine if data is empty,  $\delta_=$  to test if two data are equal in  $\delta$ , and  $\delta_{\text{def}}$  to add definitions via Equation 1.

These are meant to be a minimal collection of definitions, including only the minimal inevitable finite sequence concepts and, more generally, the combinatorics of Equation X. There are intentionally no definitions which presume underlying arithmetical operations. A natural number, ‘123’, for instance is a single atom byte sequence and there is no definition which implies interpreting this as a number in the usual way. The idea is to allow natural numbers and whatever else to emerge “organically.” There are approximately fifty definitions needed to create a basic system. We define some here for orientation. More details can be found in the Glossary.

## 4.1 Bits, Bytes and Byte sequences

If ‘ $a$ ’ is an invariant atom, a new atom with domain  $(a \ A : B)$  is defined by  $\delta_a : (a \ A : B) \mapsto (a \ A : B)$ , so that  $\delta_a$  is a fixed point on its domain. We may refer to these as “ $a$ -atoms.” For readability convenience, we define  $(:)$ -atoms,  $((:))$ -atoms and  $((:)(:))$ -atoms to hold bits, bytes and words respectively so that text strings are invariant atoms within the context  $\delta \cup \delta_{(:)} \cup \delta_{((:))} \cup \delta_{((:)(:))}$ .

## 4.2 Combinatorics

Given that text strings are invariant atoms, we can add definitions with readable names. Define the identity ‘pass’ so that  $(\text{pass} : X = X)$  for all data  $X$ .

- $\delta_{\text{pass}} : (\text{pass } A : B) \mapsto B$

and define ‘null’ so that  $(\text{null} : X = ())$  for all  $X$ .

- $\delta_{\text{null}} : (\text{null } A : B) \mapsto ()$

It is convenient to define “getting the A and B components of a coda” like so

- $\delta_{\text{left}} : (\text{left } A : B) \mapsto A$
- $\delta_{\text{right}} : (\text{right } A : B) \mapsto B$

We proceed to define the minimal combinatorics involving of A and B as finite sequences. These don’t have commonly understood names, so we are forced to invent new names. For example the name ‘ap’ is meant to suggest the concept “apply A to each atom of B” is expressed by the definition.

- $\delta_{\text{ap}} : (\text{ap } A : b \ B) \mapsto (A : b) \ (\text{ap } A : B)$
- $\delta_{\text{ap}} : (\text{ap } A : ()) \mapsto ()$

The domain of  $\delta_{\text{ap}}$  is defined with the assumption that  $b$  is an atom, so the domain of the first branch of  $\delta_{\text{ap}}$  is exactly codas  $(\text{ap } A, B)$  where  $B$  starts with an atom. By definition, if we list multiple “branches” as in this case, the first branch in the domain applies.

## 4.3 Language

Axiomatic systems like ZFC[?] or dependent type theories[ref] are formal languages. There is an alphabet, special symbols, and syntax rules for defining valid expressions. Our approach avoids this by defining a language as just one more definition like any other. The basic idea is to define minimalist language, mainly giving access to the two foundational operations via text blank space (for concatenation) and text colon (for pairing to data as a coda). So, the idea is

- $\delta_{\{\}} : (\{x\} A : B) \mapsto (\{x\} A:B) (\{y\} A:B)$
- $\delta_{\{\}} : (\{x:y\} A : B) \mapsto (\{x\} A:B) : (\{y\} A:B)$

where  $x$  and  $y$  are byte sequences as defined above. As written, this is ambiguous since  $x$  and  $y$  may contain space and colon characters, but these ambiguities can be resolved just by choosing and order, thus forcing the language into one definition  $\delta_{\{\}}$ . This definition includes

- $\delta_{\{\}} : (\{A\} A : B) \mapsto A$
- $\delta_{\{\}} : (\{B\} A : B) \mapsto B$

so that  $A$  and  $B$  have special meaning in the language. As a result, for example,  $(\{B\}:1\ 2\ 3)=1\ 2\ 3$ ,  $(\{B\ B\}:1\ 2\ 3)=1\ 2\ 3\ 1\ 2\ 3$ , and  $(\{A\ B\}\ a\ b : 1\ 2)=a\ b\ 1\ 2$ .

Given a language expression in the form of byte sequence data  $L$ , the corresponding data is, simply,  $(L:)$ . Every sequence of bytes is a valid language expression, so there is actually no need to define or check for syntax errors. Similarly, an *evaluation* of  $(L:)$  is merely some sequence  $(L:) = D_1 = D_2 = \dots = D_n$  producing an “answer”  $D_n$ . Typically, this is done with a simple strategy based on applying  $\delta$  whenever possible or up to time or space limits. This also means that there is also no such thing as a “run time error.”

## 4.4 System

The most basic question to ask about data is whether it is empty or not. This is captured by the definition  $\delta_{\text{bool}}$ , defined so that  $(\text{bool}:B)$  is  $()$  if  $B$  is empty (“true”) and  $(\text{bool}:B)=(:)$  if  $B$  is atomic (“false”).

- $\delta_{\text{bool}} : (\text{bool } A : B) \mapsto ()$  if  $B$  is empty,  $(:)$  if  $B$  is atomic.
- $\delta_{\text{not}} : (\text{not } A : B) \mapsto (:)$  if  $B$  is empty,  $()$  if  $B$  is atomic.

The equality defined by the context  $\delta$  is available within  $\delta$  via the following definition:

- $\delta_{=} : (= A : ()) \mapsto A$
- $\delta_{=} : (= (): A) \mapsto A$

and if  $a$  and  $b$  are atoms,

- $\delta_{=} : (= a A : b B) \mapsto (= a:b) (=A:B)$
- $\delta_{=} : (= A a : B b) \mapsto (=A:B) (= a:b)$

so, if  $(A : B)$  is empty,  $A$  and  $B$  are “always” equal and if  $(A : B)$  is atomic,  $A$  and  $B$  are “never” equal. The full language has a little bit of syntactic sugar, so one can write  $(A=B)$  instead of  $(= A:B)$  [ref].

New definitions can also be added to context via

- $\delta_{\text{def}} : (\text{def } \text{name } A : B \mapsto ())$

which is defined to be in domain as a partial function if `name` is not already in the current context, and, in that case, it adds the following definition to context.

- $\delta_{\text{name}} : (\text{name } A':B') \mapsto (B A':B')$

so, for instance (def first2 : first 2 : B), means that (first2 : a b c d) is interpreted as ({first 2:B} : a b c d) which is equal to the data (a b).

We take this as the “organic” base of naturally occurring definitions plus the language will be a starting point in searching for the “spaces” defined in section 5. Further definitions used in the text can be found in the Glossary. Examples, tutorials, a complete definition of the language, and software can be found in reference [x].

## 5 Global Structure

Pure data has a global structure in the sense that the foundational operations ( $A \cdot B$ ) and ( $A : B$ ) are defined for any data  $A, B$ . This suggests defining a corresponding associative product ( $A \cdot B$ ) and associative sum ( $A + B$ ) by

$$(A \cdot B) : X = A : B : X \quad (2)$$

$$(A + B) : X = (A : X) (B : X) \quad (3)$$

for all data  $X$ . This product distributes over the sum, from the right only, so we have

$$(A + B) \cdot C = (A \cdot C) + (B \cdot C) \quad (4)$$

for any data  $A, B$ , and  $C$ . Since  $(\text{pass} \cdot A) = (A \cdot \text{pass}) = A$  and  $(A + \text{null}) = (\text{null} + A) = A$  for any  $A$ ,  $(\text{pass})$  and  $(\text{null})$  are the units of data composition and data addition respectively. Following standard terminology, we say that  $A$  is *idempotent* if  $A \cdot A = A$ , is an *involution* if  $A \cdot A = 1$ , and  $A$  has an *inverse* if  $A \cdot A' = A' \cdot A = 1$  for some data  $A'$ . In the case of a *product*  $A = A_n \cdots A_1$ , we say that  $A$  starts with  $A_1$  and ends with  $A_n$ .

Data  $A$  is called *algebraic* or *commutative* if  $A : X Y = A : Y X$  for all  $X, Y$ . Data  $A$  is *distributive* if  $A : X Y = (A : X) (A : Y)$  for all  $X$  and for all  $Y$ . Data  $A$  defines a binary operator on data via  $(X * Y) = (A : X Y)$ . Since  $*$  is associative if and only if  $(A : X Y) = (A : (A : X) Y) = (A : X (A : Y))$  for all  $X, Y$  we say that data  $A$  is *associative* if it has this property.

Algebraic and distributive data foreshadow a theme where these two properties make data mathematically interesting for complementary reasons. Algebraic data has, in a sense, “transcended finite sequence,” the foundational concept of the system and has a platonian existence in that sense. Distributive data, on the other hand, is maximally sequence dependent, but it may also be mathematically interesting since such data is a “morphism of finite sequence.”

## 6 Spaces

A basic requirement for useful mathematics is to have a way to define and refer to collections. Within the framework of pure data, however, there is only one “substance” to work with. All mathematical objects are merely data and all collections of mathematical objects are also merely data. Thus, we are lead to ask, given some data  $S$ , how could  $S$  represent a collection of other data? There are a couple of plausible ways to do this.

- The collection corresponding to  $S$  would be the collection of data  $(S : X)$  for any data  $X$ ;
- The collection corresponding to  $S$  would be the collection of all of the fixed points of  $S$ .

These ideas coincide if we require  $S$  to be idempotent, so that  $(S : X)$  is automatically a fixed point. It is natural to also expect compatibility with sequences in the sense that if  $(S : X)$  is in the collection

and  $(S:Y)$  is in the collection, then  $(S:X)$   $(S:Y)$  is also in the collection. This is guaranteed if we require

$$(S : X \cdot Y) = S : (S : X) \cdot (S : Y) \quad (5)$$

for all data  $X, Y$ . Note that data  $S$  is idempotent and satisfies Equation 4 if and only if  $S$  is associative, and so we are lead to a simple definition.

**Definition 3.** Data  $S$  is a **space** if  $S$  is associative.

If  $S$  is a space, we say that any data  $(S:X)$  is *in*  $S$ . The data  $(S:)$  is called the *neutral data* of  $S$ .

Examples of spaces can be found in the basic definitions, including ‘pass’, ‘null’ and ‘bool’. More examples come from noting that any data which is both idempotent and distributive is a space. Thus, if  $J$  is idempotent, then  $(\text{ap } J)$  is idempotent and distributive, and is, therefore, a space. If spaces  $S$  and  $T$  commute, then  $S \cdot T$  is a space.

## 6.1 Morphisms

A product  $F$  that starts with space  $S$  and ends with space  $T$  is a *morphism* from  $S$  to  $T$  and can be written  $S \xrightarrow{F} T$ . Since spaces are idempotent,  $F = (F \cdot S) = (T \cdot F)$ . Since  $T \cdot S$  is a morphism from  $S$  to  $T$ , morphisms always exist. Since the product is associative, morphisms can be composed, so that if  $S \xrightarrow{F} T$  and  $T \xrightarrow{G} U$ , then  $S \xrightarrow{G \cdot F} U$ . If  $F$  and  $G$  are morphisms from  $S$  to  $T$ , we can also define a *sum of morphisms* by  $F \oplus G = T \cdot (F + G)$ , so that  $S \xrightarrow{F \oplus G} T$ . A morphism from a space to itself is called an *endomorphism*.

Since  $T \cdot X \cdot S$  is a morphism from  $S$  to  $T$  for any data  $X$ , a morphism can define any function. Special morphisms which “preserve the structure” of  $S$  and  $T$  can be defined as follows. A morphism  $S \xrightarrow{H} T$  is a *homomorphism* if

$$H \cdot (f \oplus_S g) = (H \cdot f) \oplus_T (H \cdot g) \quad (6)$$

for all endomorphisms  $f$  and  $g$  of  $S$ . Compare with morphisms that happen to also be a space. Morphism  $S \xrightarrow{U} T$  is a space if and only if

$$U \cdot (f \oplus_S g) = U \cdot ((U \cdot f) \oplus_T (U \cdot g)) \quad (7)$$

for all endomorphisms  $f$  and  $g$  of  $S$ . Thus, any idempotent homomorphism is a space. Note that homomorphisms are closed under composition, but the composition of two spaces is a space only if the two spaces commute.

## 6.2 Spaces are Semirings

Consider the endomorphisms of a fixed space  $S$ . Since  $S$  itself is a product starting at  $S$  and ending at  $S$ ,  $S$  qualifies as an endomorphism of itself. Since  $(S \cdot f) = (f \cdot S) = f$  for any endomorphism  $f$ ,  $S$  is “its own identity morphism.” The endomorphisms of  $S$  are closed under both composition and addition, with addition defined by  $f \oplus g = S \cdot (f + g) \cdot S = S \cdot (f + g)$ . Thus, the endomorphisms of  $S$  are a ‘semiring’ with composition and addition and with  $1 = (S \cdot \text{pass} \cdot S) = S$  and with  $0 = (S \cdot \text{null} \cdot S) = (S \cdot \text{null})$ .

**Definition 4.** A **semiring** is a set with associative addition and multiplication (denoted  $f \oplus g$  and  $f \cdot g$ ), with additive and multiplicative identities **0** and **1** satisfying  $f \oplus 0 = f = 0 \oplus f$  and  $f \cdot 1 = f = 1 \cdot f$  for all  $f$  in the semiring, and where  $(f \oplus g) \cdot h = (f \cdot h) \oplus (g \cdot h)$  for all  $f, g$ , and  $h$  in the semiring. A subset of a semiring  $S$  which contains 0 and 1 and is closed under addition and multiplication is a **subsemiring** of  $S$ .

Note that the semiring of space  $S$  has  $1=0$  if and only if  $S$  is a constant space equal to  $(\text{const } K)$  for some data  $K$ . It is easy to verify that the semiring of the space (pass) is exactly the global algebra defined in Section 5. Since isomorphic spaces have isomorphic semirings, understanding the semiring of a space is mainly what we aim for in the examples which follow. Before going to examples, however, it is helpful to identify semiring features of interest.

Consider the semiring of a fixed space  $S$ .

- **Subspaces.** If an endomorphism  $s$  of  $S$  happens to also be a space, we say that  $s$  is a *subspace* of  $S$ . The name “subspace” is justified because every data contained in  $s$  is also contained in  $S$ , and every endomorphism  $s \cdot X \cdot s$  of  $s$  is also an endomorphism of  $S$ . Subspaces partially distribute from the left as with  $s \cdot (f \oplus g) = s \cdot ((s \cdot f) \oplus (s \cdot g))$ . Every space has subspaces  $1$  (the whole space), and  $0$ , the constant neutral space. If subspace  $s$  is not equal to  $S$ ,  $s$  is a *proper* subspace.
- **Constants.** An endomorphism  $k$  of  $S$  satisfying  $k \cdot 0 = k$  is a **constant** subspace of  $S$ . There is one such constant for each data in  $S$ . Since  $k \cdot f$  and  $k \oplus l$  are constants for constant  $k$  and  $l$  and for any endomorphism  $f$ , the constants of  $S$  are a left-ideal of the semiring of  $S$ .
- **Homomorphisms.** By Equation 5, the homomorphisms of  $S$  are exactly the endomorphisms which distribute over addition from both the left and from the right. The endomorphisms  $0$  and  $1$  are homomorphisms. If  $f$  and  $g$  are homomorphisms of  $S$ , then  $f \cdot g$  is a homomorphism. If  $S$  is algebraic, the homomorphisms are a subsemiring of the semiring of  $S$ .
- **Group of units.** The collection of endomorphisms with multiplicative inverses is called the *group of units* of the space  $S$ .
- **Central endomorphisms.** Endomorphisms which commute with the group of units are *central endomorphisms*. The endomorphisms  $0$  and  $1$  are central. If  $f$  and  $g$  are central endomorphisms, then  $f \cdot g$  is a central endomorphism. If  $f$  and  $g$  are *central homomorphisms*, then  $f \oplus g$  is also a central homomorphism. Thus, the central homomorphisms of  $S$  are a subsemiring of the semiring of  $S$ .
- **Idempotents.** Define a relation  $\leq^i$  on the endomorphisms of  $S$  by  $f \leq^i g$  if  $(f \cdot g) = f$ . The equivalence classes of the corresponding symmetric relation  $f \sim^i g$  contain all the idempotent endomorphisms, since  $f \sim^i g$  implies that both  $f$  and  $g$  are idempotent. In this *idempotent order*,  $1$  is the maximum endomorphism and the constant endomorphisms are minimal.
- **Neutral spaces.** If addition in  $S$  allows left and right cancellation ( $f \oplus g = f \oplus h$  implies  $g = h$  and  $g \oplus f = h \oplus f$  implies  $g = h$ ),  $S$  is a *neutral* space. The neutrality of  $S$  guarantees  $h \cdot 0 = 0$  for homomorphism  $h$ , and guarantees that if  $f \oplus f = f$ , then  $f = 0$ .
- **Semilattice spaces.** If  $S$  is algebraic and  $f \oplus f = f$  for all endomorphisms  $f$ , then  $S$  is called a *semilattice space*.

In the case where  $S$  is distributive,  $(f \oplus g) = (f + g)$  for  $f, g$  endomorphisms of  $S$ . This means that the homomorphisms of  $S$  are just the distributive endomorphisms of  $S$ . If the group of units of  $S$  includes permutations, then the central homomorphisms are algebraic. This illustrates the generalization of the earlier remark about algebraic and distributive data being mathematically interesting for complementary reasons. Central endomorphisms are the generalization of algebraic data and homomorphisms are the generalization of distributive data. This suggests that the central homomorphisms of a space will be maximally mathematically meaningful.

### 6.3 Isomorphic Spaces have Isomorphic Semirings

The fact that homomorphisms compose suggests adapting definitions from Category Theory. We say, for instance, that spaces  $S$  and  $T$  are *isomorphic* if the diagram of Figure 1 commutes for homomorphisms  $h$  and  $h'$ , where we have used  $S$  and  $T$  instead of respective “identity morphisms.”

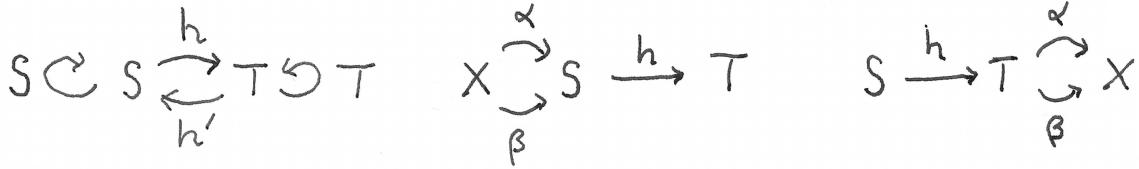


Figure 1: Since homomorphisms compose, standard definitions can be adopted from Category Theory where each space is its own “identity morphism.”  $S$  and  $T$  are isomorphic if the diagram on the left commutes for homomorphisms  $h$  and  $h'$ . A homomorphism  $h$  is a monomorphism if the center diagram implies  $\alpha = \beta$  and  $h$  is an epimorphism if the right diagram implies  $\alpha = \beta$ .

Given an isomorphism as in Figure 1, the bijection  $f \mapsto h' \cdot f \cdot h$  is a semiring isomorphism, preserving product, sum, 1 and 0, so that isomorphic spaces have the same features defined in the previous section, since these are all defined with semiring operations. A homomorphism  $h$  where  $h \cdot \alpha = h \cdot \beta$  implies  $\alpha = \beta$  is a *monomorphism*, and if  $\alpha \cdot h = \beta \cdot h$  implies  $\alpha = \beta$ , then  $h$  is an *epimorphism* [?].

### 6.4 Spaces can be Semialgebras

The semiring of a space  $S$  is close to the structure of an “algebra” in standard mathematics. There is a binary operation  $f \oplus g$ , so constants of the space can be added. Homomorphisms have the property that they map constants to constants and are distributive with  $h \cdot (k \oplus l)$  equal to  $(h \cdot k) \oplus (h \cdot l)$  so homomorphisms of  $S$  are like a module over the constants of  $S$ . This would be an algebra in the ordinary sense except that homomorphisms might not be constants. This suggests a definition.

**Definition 5.** A space with a one-to-one mapping from constants to homomorphisms is a **semialgebra**. If the homomorphisms in the image of the mapping are central, the semialgebra is a **central semialgebra**.

Since pure data is meant to include all mathematical objects, we might expect that important algebras of standard mathematics could appear as semialgebras.

### 6.5 Subspaces are Substructures

Suppose that space  $S$  has subspace  $s$  and space  $T$  has subspace  $t$ , and suppose that  $F = t \cdot X \cdot s$  is a morphism from  $s$  to  $t$ . Since  $(t \cdot X \cdot s) = (T \cdot t \cdot X \cdot s \cdot S)$ ,  $F$  is also a morphism from  $S$  to  $T$ , with the extra property that  $F$  respects the structure of  $s$  and  $t$  in the sense that  $t \cdot F = F \cdot s$ . If a space has multiple subspaces, each commuting pair of subspaces is a new subspace with compatible structure.

Consider a subspace  $s$  of a space  $S$ . If a homomorphism  $h$  of  $S$  is also an endomorphism of  $s$ , then  $h$  is also a homomorphism of the subspace  $s$ . Similarly, if a unit  $u$  of  $S$  is an endomorphism of  $s$ , then it is also a unit of  $s$ . Such homomorphisms and units *descend* into the subspace  $s$ .

For any morphism  $S \xrightarrow{h} T$ , there is a guaranteed factorization

$$h = m \cdot e \tag{8}$$

where  $m$  is a monomorphism and  $e$  is an idempotent of  $S$ . If  $h$  is a homomorphism, we can substitute this into the definition of a homomorphism, use  $h \cdot e = h$ , and, cancelling  $m$  on the left, find

$$e \cdot (f \oplus g) = e \cdot (e \cdot f \oplus e \cdot g), \tag{9}$$

and conclude that  $e$  is a subspace of  $S$ , which we call a *quotient* of  $S$ , and denote  $e$  by  $S/h$ . Note that  $S/h = S$  if and only if  $h$  is a unit as well as a homomorphism.

## 6.6 Fields

Since spaces have subspaces, which may themselves have subspaces, a descending sequence of subspaces can only terminate with a constant space, since constant spaces are the only spaces with no proper subspaces. From the semiring perspective, constant spaces have no variety. All constants spaces are isomorphic and they all have  $0=1$ . On the other hand, there is a variety of spaces one step above constants. These are the ‘fields’.

**Definition 6.** A space where all proper subspaces are constants is a **field**.

In classical ring theory, a field is guaranteed to have multiplicative inverses. An analogous result nicely combines homomorphisms, constants, units and subspaces [?].

**Theorem 6.1.** A space  $S$  is a field if and only if every non-constant homomorphism of  $S$  is a unit.

*Proof.* If  $S$  is a field, every non-constant homomorphism  $h$  has  $S/h=S$  and, thus, is a unit. Conversely, suppose that  $S$  is not a field. Then  $S$  contains a proper non-constant subspace  $s$  and  $\sigma = S \cdot \text{ap } s \cdot S$  is, therefore, a non-constant homomorphism of  $S$ . Since  $\sigma : X$  is in  $s$  for any data  $X$ ,  $\sigma$  is not a unit.  $\square$

The obvious examples of fields are null, bool and any constant space. In the examples, we will see familiar fields such as the natural numbers modulo a prime or the field of rational numbers. In these examples, zero will be the only constant homomorphism and since semialgebra multiplication is application by a homomorphism, the guarantee of an inverse for non-zero elements comes from a guarantee that non-constant homomorphisms are units.

## 7 Plan

Although the motivation for the concept of a space is merely to define a general collection of data, and a space is merely any associative data, we find that spaces and their endomorphisms have a rich internal structure including subspaces, a group of “symmetries,” a class of endomorphisms which qualify as homomorphisms, and central endomorphisms which commute with the symmetries. This is a common structure shared by all collections ranging from a space containing only empty data (null) to a space containing one atom (`const (:)`) to a space containing all mathematical objects (pass). The scope of these results means that pure data spaces can be thought of as a general point of view about mathematical objects, analogous to sets with structure or Category Theory. We highlight some of the theoretical differences before going on to examples.

- A set is defined by its contents, but a space is not. Spaces  $(\text{ap } \{a\})$  and  $(\text{is } a)$ , for instance, contain the same data, are idempotent equivalent  $(\text{ap } \{a\})^i \sim (\text{is } a)$ , are isomorphic as spaces, and have the same semirings, but they are not the same space. For example,  $\text{ap } \{a\}:(:) = a$ , but  $(\text{is } a):( :) = ()$ . Since every space contains its neutral data at least, it is not possible for a space to be empty.
- Spaces have morphisms between them which compose associatively, but, unlike Category Theory, morphisms and spaces are “made of the same substance.” They are both merely data. For instance, every data  $A$  of pass is also a morphism  $\text{pass} \cdot A \cdot \text{pass}$  from pass to pass. In contrast with categories and objects, a morphism  $S \xrightarrow{T \cdot S} T$  exists between any two spaces  $S$  and  $T$ , so there are no closed categories where morphisms do not cross category boundaries. Unlike Category Theory, a morphism can be morphisms between multiple spaces at once. For instance, if  $S$  and  $T$  are commuting spaces, any fixed morphism  $(S \cdot T) \xrightarrow{F} U$  means that  $F$  is a morphism from  $S \cdot T$  to  $U$ , from  $S$  to  $U$  and is also a morphism from  $T$  to  $U$ .
- Category Theory is a theory of structures with corresponding morphisms. In the theory of spaces, each space comes with both structure preserving “homomorphisms” and not-necessarily-structure-preserving morphisms, packaged together in a coherent semiring.
- In classical mathematics, one expects sets to have subsets, groups to have subgroups, rings to have ideals, and, in general, structures to have similar substructures. Thus, it is not surprising that spaces can have subspaces, but unlike the classical situation, subspaces of a space can have completely different mathematical structures from the ambient space. For example, we shall see that the space  $(\text{is } a \ b \ c \ d)$  has various free monoids, natural numbers, integers, the algebra of 4x4 natural number matrices and Gaussian Integers as different subspaces. The extreme version of this is the space  $(\text{pass})$  which contains *every* space as a subspace. One can think of an algebraic space  $S$  as having “horizontal” features coming from the subsemiring of homomorphisms and “vertical” features coming from general endomorphisms and subspaces.

Since this is a rich but unfamiliar viewpoint, the plan of the paper is to clarify the situation with example spaces, proceeding from the simplest, most “organic” spaces first, starting with foundational definitions which happen to also be spaces, as shown in Figure 2. The pure data view of spaces gives us a criteria for what is “most organic” and gives a way to systematically search for all spaces up to a specified data width and depth. Although we mainly use the pure data perspective to choose spaces of interest, what we say about spaces will be put in semiring language whenever possible. Aside from the organic and semiring themes, we will take special notice of algebraic data following the intuition that algebraic spaces are most likely to be mathematically interesting since they have, in a sense, earned a platonic meaning by transcending the foundational finite sequence structure of the system. In semiring terms, we will thus focus on the central homomorphisms of a space as the most likely features of interest.

## 8 The Space of all data

As a first example and for orientation, let’s consider the space of all pure data. Since  $(\text{pass}:X)=X$  for all data  $X$ , the distributive space  $(\text{pass})$  contains all pure data. Since spaces are fixed points on their contents, pass is the unique space containing all data. Every data  $X$  is both in pass and is also an endomorphism of pass, since  $X$  is equal to  $\text{pass} \cdot X \cdot \text{pass}$ . Thus, the semiring of pass is also the collection of all data, with addition  $(X \oplus Y)=(X + Y)$ , and multiplication  $(X \cdot Y)$ . The units of the semiring are  $1=\text{pass} \cdot \text{pass} \cdot \text{pass}=\text{pass}$ , and  $0=\text{pass} \cdot \text{null} \cdot \text{pass}=\text{null}$  respectively.

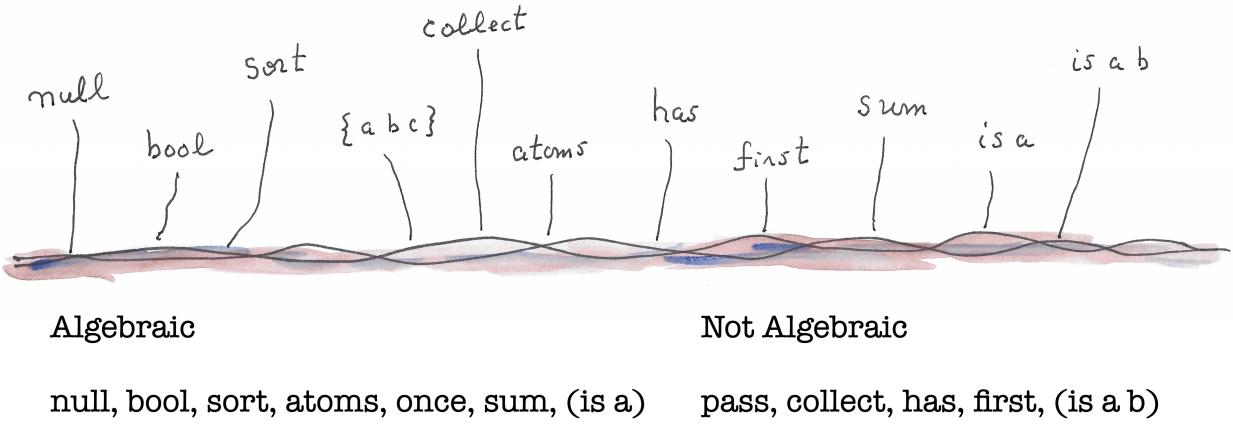


Figure 2: *The simplest spaces emerging “organically” from the space (pass) of all pure data. The algebraic spaces are commutative and, in a sense, transcend foundational sequence concept, making them more “platonic” and more mathematically interesting than non-algebraic spaces.*

Consider the properties defined in section 6.

- **Subspaces.** Since every data is an endomorphism of pass, every space is a subspace of pass. The constants of pass are the data  $K$  such that  $K \cdot 0 = K$ , in other words, the constant spaces ( $\text{const } K$ ) are the constants.
- **Homomorphisms.** Since pass is distributive, the homomorphisms of pass are the distributive data in pass, including spaces, such as  $(\text{ap } \{a\})$  and non-spaces such as  $(\text{ap } \{B\ B\})$ .
- **Group of units.** The units of pass are permutations such as  $(\text{swap } 1\ 2)$  and  $(\text{rev})$ . The central constants of pass are the data which are invariant under all permutations. These are the sequences of identical atoms such as  $(:) (:) (:)$  or  $(a\ a\ a\ a\ a)$ . These are interpreted as “organic natural numbers.”
- **Central endomorphisms** The maximally platonic, maximally significant content of pass are the central homomorphisms, which include subspaces such as  $(\text{ap const } a)$ ,  $(\text{atoms})$ ,  $(\text{is } b)$ . These are isomorphic spaces, each representing natural numbers with natural number addition. This and the direct connection of bool with the underlying pure data context means that both the natural numbers and bool are starting points for exploration.
- **Semialgebra** In pass, there is a bijection between the homomorphisms of pass and the constants of pass given by  $X \leftrightarrow (\text{ap } X)$ . Thus, pass is a semialgebra as well as a semiring with  $X \star (Y + Z)$  equal to  $(X \star Y) + (X \star Z)$  where  $X \star Y$  is defined to be  $(\text{ap } X) \cdot Y$ .
- **Idempotents** The idempotent equivalence classes of pass includes all idempotent data where  $(\text{pass})$  is the unique maximum idempotent and where all constants are in the minimal equivalence class. In the idempotent order,  $(\text{pass})$  is the unique maximum.
- **Semilattice** Semilattice spaces include  $(\text{bool})$ ,  $(\text{once})$  and  $(\text{const } K)$  for any data  $K$ . The space  $\text{bool}$  is special in several ways. It is a semilattice, a field, a central endomorphism, and is isomorphic to any space with exactly two data.

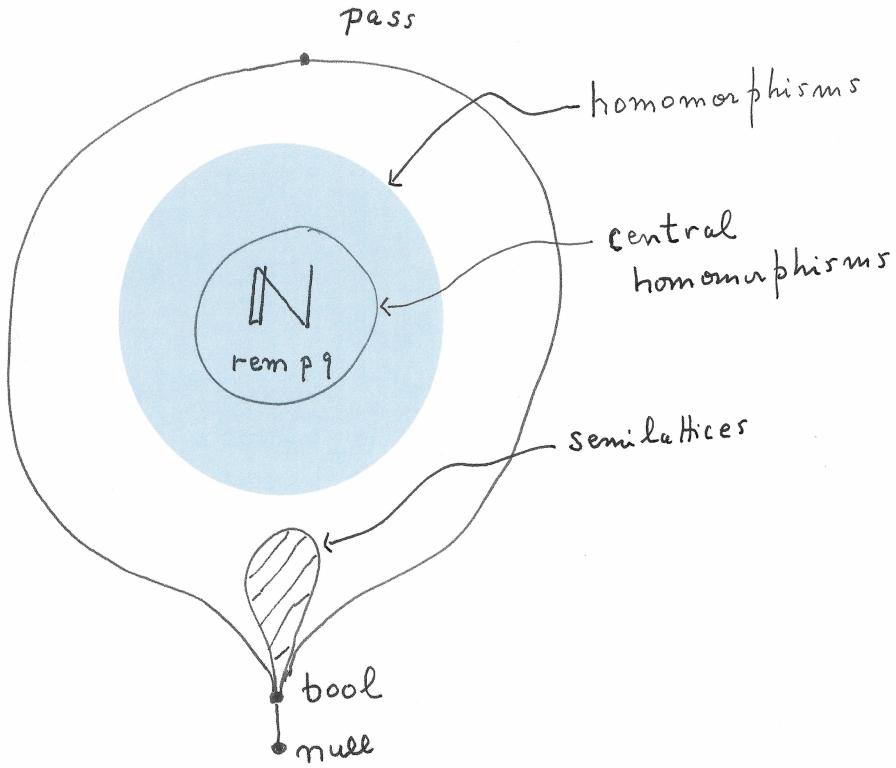


Figure 3: As a collection,  $\text{pass}$  contains all pure data; that is, all objects with mathematical meaning as well as all computations. As a space, it has a specific structure which is typical of other examples in the text. In the idempotent partial order,  $(\text{pass})$  is the unique maximum and  $(\text{null})$  is minimal as are all constants. Each data in  $\text{pass}$  is also an endomorphism of  $\text{pass}$ . The homomorphisms of  $\text{pass}$  are the distributive data. The central homomorphisms of  $\text{pass}$  are a single isomorphism class equivalent to the “organic” natural numbers. Spaces that are fields includes  $\text{bool}$ ,  $\text{null}$ , and the  $\text{rem}(p,p)$  subspaces of  $\mathbf{N}$ , equal to  $\mathbf{F}_p$  for prime  $p$ . The space  $\text{bool}$  is the unique space with two data and is a semilattice as well as a field.

In classical mathematics, the concept of a Set intentionally gives no information about its contents. Thus, the set of all sets is vast, but not interesting. In our case,  $(\text{pass})$  is vast, but also has particular features which will be common to all of the spaces considered here. Figure 3 gives a visual overview of  $\text{pass}$ , highlighting the features of interest and points for exploration in the following sections.

## 9 Null

$\text{Null}$  is the distributive algebraic space defined by  $(\text{null}:X)=()$  for all  $X$ . Since  $\text{null} \cdot X \cdot \text{null} = \text{null}$ ,  $\text{null}$  is the only endomorphism of  $\text{null}$ , and so  $1=0$  in the semiring, as is true for all other constant spaces. Since  $0 \cdot (0 \oplus 0) = (0 \cdot 0) \oplus (0 \cdot 0) = 0$ , the one endomorphism is a homomorphism as well as a subspace.  $\text{Null}$  is minimal in the idempotent order.

## 10 Organic Numbers

We have seen that “organic natural numbers” appear as sequences of identical atoms where we identify  $(:) (:) (:)$  or  $(a a a)$  with “3,” just depending on a conventional choice of atom. The spaces where these numbers are data are the maximally platonic central homomorphisms of pass. For instance, the spaces  $(\text{atoms})$ ,  $(\text{ap const}(:))$ ,  $(\text{ap const } a)$ ,  $(\text{ap } \{\#\})$ ,  $(\text{is } a)$  all are central homomorphisms and each represents natural numbers with natural number addition. Since these spaces are isomorphic and, therefore, have the same semirings, we can choose any one for investigation. For convenience we will choose  $(\text{is } a)$  as the particular space to start with. For convenience, denote  $n$  concatenations of data  $A$  by  $A^n$  so that “5” represented by  $(a a a a a)$  and can be written as  $a^5$ . The spaces  $(\text{is } a)$ ,  $(\text{is } a b)$ ,  $(\text{is } a b c), \dots$  are *organic numbers*.

### 10.1 is a

Start with  $N = (\text{is } a)$ , the *organic natural numbers*. The first thing to notice about  $N$  is that  $N$  contains one data  $a^n$  for each natural number  $n$ , and addition in  $N$  is natural number addition, since  $(N:X Y)$  is the natural number sum of  $(N:X)$  and  $(N:Y)$ . Since homomorphisms are distributive, a homomorphism  $h$  of  $N$  is determined by  $(h:a)$ , so that  $h$  is natural number multiplication by some natural number, for instance,  $(h = \text{ap const } a a a)$  is multiplication by 3. There is one such homomorphism for each data in  $N$ , and so  $N$  is a semialgebra and is identical to the standard semiring of natural numbers with multiplication distributing over addition.

Subspaces of  $N$  are indicated by subspace endomorphisms. As always, the subspace 1 is the whole of  $N$  and the subspace 0 is the neutral data of  $N$ , which, since  $N$  is distributive, is the empty sequence. It is easy to guess some subspaces of  $N$ .

1. **Saturation subspaces** An endomorphism  $(\text{min } a a)$  is a subspace which effectively computes the minimum of 2 and its argument. Thus  $(\text{min } a a)$  contains the data  $()$ ,  $(a)$ , and  $(a a)$ .
2. **Modular arithmetic subspaces** An endomorphism which removes three  $(a)$  atoms while possible, such as  $(\text{while remove } a a a)$  is a subspace which computes the sum modulo 3. Note that  $(\text{while remove } a a a)$  contains the same data as  $(\text{min } a a)$ , but they have a different additive structure.

These unsurprising subspaces have a slightly more surprising generalization. For  $p, q \geq 1$ , let  $\text{rem}(p,q)$  be the endomorphism defined by

$$\text{while } n \geq p, q : \text{remove } p \text{ from } n \tag{10}$$

which results in  $\text{rem}(p,q)(n) = \min(n \bmod p, q - 1)$ , specializing to addition modulo  $p$  for  $\text{rem}(p,p)$ . This completes the analysis of  $N$ .  $N$  is a semialgebra identical to the standard semiring  $\mathbb{N}$ . The only subspaces of  $N$  are 1, constants and  $\text{rem}(p,q)[?]$ . The subspaces are all mutually commuting. The non-trivial fields of  $N$  are  $\text{rem}(p,p)$  for prime  $p$ .

### 10.2 is a b

The second simplest organic number is  $(\text{is } a b)$ , which we can refer to as  $N_2$  for brevity. Since  $N_2$  is distributive, the homomorphisms of  $N_2$  are the distributive endomorphisms. Consider subspaces  $N_2$ .

- The subspaces 1 and 0 give the whole of  $N_2$  and just the empty sequence as subspaces, respectively. Projections  $N_2 \cdot (\text{is } a) \cdot N_2$  and  $N_2 \cdot (\text{is } b) \cdot N_2$  produce two  $N$ -isomorphic subspaces. The absolute value subspace  $(\text{ap const } a)$  is also isomorphic to  $N$ .

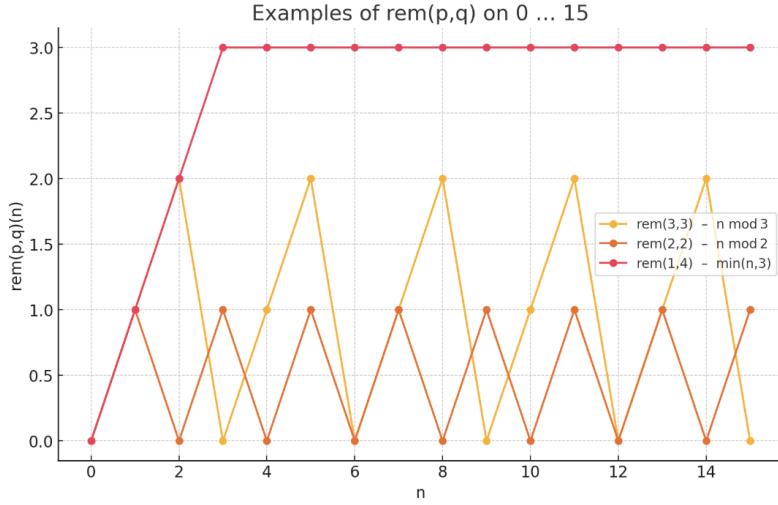


Figure 4: Subspaces of (is a) or  $N$  are  $N$  itself, constants and  $\text{rem}(p,q)$  subspaces. These specialize to natural numbers mod  $p$  and ceiling functions  $\min(q)$ .

- Let  $\text{reduce}$  be the endomorphism of  $N_2$  which removes (a b) and (b a) subsequences until saturation. Thus, every data in  $\text{reduce}$  is either  $a^n$  or  $b^n$  for some natural number  $n$ . A homomorphism  $h$  satisfies  $h : X Y = \text{reduce} : (h : X) (h : Y)$ , so  $h$  is determined by  $(h:a)$  and  $(h:b)$ . If  $h$  is a central homomorphism,  $(h:b)$  is  $(h:a)$  followed by an  $a \leftrightarrow b$  swap. Thus, if we interpret  $a^n$  as the integer  $n$  and  $b^n$  as the integer  $-n$ , then applying  $h$  is standard distributive integer multiplication by the integer  $(h:a)$ . Since there is one central homomorphism for each data in the subspace  $\text{reduce}$ ,  $\text{reduce}$  is a central semialgebra isomorphic to the standard ring  $\mathbf{Z}$  of integers.
- If  $\text{sort}$  is the endomorphism which does lexical sorting of  $N_2$  data, then  $\text{sort}$  is also a subspace of  $N_2$  where the data of  $\text{sort}$  can be written  $a^m b^n$  for  $m, n \geq 0$ . As in the previous case, a homomorphism  $M$  of  $\text{sort}$  is defined by

$$M : a \mapsto a^{m_{11}} b^{m_{21}}, \text{ and}$$

$$M : b \mapsto a^{m_{12}} b^{m_{22}}$$

for some choice of  $\begin{pmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{pmatrix}$  so the action of  $M$  is standard matrix multiplication. Thus, the data of  $\text{sort}$  are pairs of natural numbers and the homomorphisms are  $2 \times 2$  matrices with natural number entries. The central homomorphisms of  $\text{sort}$  commute with the involution  $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  and so the central homomorphisms are matrices  $\begin{pmatrix} m & n \\ n & m \end{pmatrix}$  for natural numbers  $m$  and  $n$ . Thus, there is exactly one central homomorphism for each data in  $\text{sort}$ , and  $\text{sort}$  is, therefore, a central semialgebra, equivalent to the standard matrix semiring  $\text{Mat}_{2 \times 2}(\mathbf{N})$ .

- Lexical sort followed by reducing (b b) to (b) results in a subspace equivalent to  $\mathbf{N} \times \{0, 1\}$  with addition defined by  $(n, \alpha) + (m, \beta) = (n + m, \alpha \vee \beta)$ .
- Lexical sort followed by reducing  $a^n b^m$  to  $a^{(n/g)} b^{(m/g)}$  where  $g = \text{GCD}(n, m)$  is the greatest common divisor. This is the space of positive rational numbers, but with non-standard

'median' addition where  $(n/m) + (n'/m')$  is defined to be  $(n + n')/(m + m')[?]$ .

### 10.3 is a b c d...

It is clear that in going to higher organic numbers, some of the features of  $N$  and  $N_2$  will repeat.  $N_n$  sorted will always result in a space with homomorphisms  $\text{Mat}_{n \times n}(\mathbf{N})$ . Sorting followed by cancelling atoms in pairs will result in  $\text{Mat}_{(n/2) \times (n/2)}(\mathbf{Z})$ . For example, in (is a b c d) with subspace (a b)=(b a)=(c d)=(d c)=(), homomorphisms are  $\text{Mat}_{2 \times 2}(\mathbf{Z})$ . Central homomorphisms of  $N_4$  sorted and reduced must commute with  $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$  and are thus matrices of the form  $\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$  and we have the central semialgebra equivalent of the Gaussian Integers  $\mathbf{Z}[i]$ . Figure 5 shows similar results for some of the more notable subspaces. Notice that this is a success from the organic gardening perspective. Important mathematical structures are emerging quite naturally. Even though both semiring operations are associative, a non-associative algebra such as the integer Octonions appears in the same way as the other spaces with homomorphisms implementing the non-associative octonion multiplication[?].

#### Organic Numbers

Space	Subspace	Data Elements	Central Homomorphisms	Central Semialgebra
is a	—	$\mathbb{N}$	Multiply-by-k	$\mathbb{N}$
is a b	$ab = ba = ()$	$\mathbb{Z}$	Multiply-by-k	$\mathbb{Z}$
is a b	sort	$\mathbb{N}^2$	Symmetric $2 \times 2$ matrices	$\text{Mat}_{2 \times 2}^{\text{sym}}(\mathbb{N})$
is a b	sort, $bb = ()$	$\mathbb{N} \times \mathbb{Z}_2$	Block-diagonal maps	$\mathbb{N} \times \mathbb{Z}_2$
is a b	sort, gcd-reduced	Coprime $(m, n)$	Identity/constants only	None
is a b c d	$ab = ba = (), cd = dc = ()$	$\mathbb{Z}^2$	Complex scalars	$\mathbb{Z}[i]$
is a <sub>1</sub> ...a <sub>8</sub>	pair reductions	$\mathbb{Z}^4$	Quaternion scalars	$\mathbb{H}_{\mathbb{Z}}$
is a <sub>1</sub> ...a <sub>16</sub>	pair reductions	$\mathbb{Z}^8$	Octonion scalars	$\mathbb{O}_{\mathbb{Z}}$
is a <sub>1</sub> ...a <sub>32</sub>	pair reductions	$\mathbb{Z}^{16}$	Sedenion scalars	$\mathbb{S}_{\mathbb{Z}}$

Figure 5: *Classical algebras appear as organic numbers.*

Although many interesting spaces appear as organic numbers, some simple spaces appear to be missing. It's noticeable, for instance, that standard rationals do not appear. We would like to understand why.

## 11 Rationals

In the naive version of the organic theme, we would define a context with the basic combinatoric definitions and then search the space of pure data, up to, say, some specified width and depth. Then, for each data, we would test if it is a space or not. Ideally, all of the most important mathematical structures would all be found this way. They would appear “organically.” This has actually worked fairly well up through the organic numbers (Figure 5) but there we noticed that only a mediant-addition version of the rational numbers appeared as an organic integer. The theory of fields makes the reason for this situation clear. Any non-constant subspace of an organic number space ( $(is\ a\ b\ c\dots)$ ) will have proper non-constant subspaces formed by projecting onto a subset of ( $a\ b\ c\dots$ ). This means that the only non-trivial fields in the organic numbers are  $\mathbf{F}_p$ , the subspaces  $\text{rem}(p,p)$  for prime  $p$  of ( $is\ a$ ).

As a result, we take a more active design approach to creating a space that represents the standard rationals, while still aiming to be as organic as possible in the sense of only using basic combinatoric definitions. For illustrative purposes, we follow a broadly applicable plan for defining a space corresponding to some algebra of interest. The plan is as follows.

- (a) Define suitable atoms to represent the objects of interest.
- (b) Create a base “free construction” space, typically a distributive sequence of a) atoms.
- (c) Invent a sequence of mutually commuting subspaces of the base space, until the sum in the subspace product agrees with the desired sum in a mathematical sense.
- (d) The homomorphisms of the resulting product subspace are then define a module-like product. If there is a one to one mapping from constants to homomorphisms, one has a semialgebra or a central semialgebra.

In the case of rational natural numbers, we proceed as follows.

- (a) We will represent single rational numbers with  $q$ -atoms such as  $(q\ a\ a:\ a\ a\ a)$  representing  $2/3$ . We could define  $q$  direct as a fixed point via  $(q\ A\ B:B) \mapsto (q\ A\ B:B)$ . It is more convenient and more illustrative to instead define a “fancy”  $q$ -atom that is not a direct fixed point, but is nevertheless equal to a fixed point.
  - $(q\ A\ B : B) \mapsto (q\ A:B)$
  - $(q\ A : A\ B) \mapsto (q\ A:B)$

The point of this definition is that the guaranteed fixed point guarantees that the  $A$  and  $B$  components of each  $q$ -atom are coprime natural numbers.

- (b) For the base “free construction” space, we can simply choose the distributive space of  $q$ -atoms, modified only by filtering to guarantee that “numerator” and “denominator” are sequences of  $a$ -atoms, and excluding  $q$ -atoms with empty denominators.
- (c) Define mutually commuting subspaces of  $Q$  approaching the desired properties of rationals.
  - ADD is the subspace defined by  $(q\ X_1 : Y_1)\ (q\ X_2 : Y_2)$  maps to  $(q\ (\text{pr}\ X_1 : Y_2)\ (\text{pr}\ X_2 : Y_1) : (\text{pr}\ Y_1 : Y_2))$ , where  $\text{pr}=(\text{ap const}\ a)\cdot\text{ar}$  is the natural number product.
  - NORM is the subspace of  $Q$  defined by  $(q\ A:A) \mapsto (q\ a:a)$ .
  - NORM0 is the subspace of  $Q$  defined by  $(q:B) \mapsto ()$ .

The subspace  $S = (\text{ADD} \cdot \text{NORM} \cdot \text{NORM0})$  thus contains the empty sequence (representing the zero rational) and a single  $q$ -atom (representing non-zero rationals). The sum in  $S$  is standard rational number addition in  $\mathbb{Q}^+$ .

- (d) As for the nicest of the organic numbers, for each rational in  $S$ , there is a corresponding homomorphism  $h_s$  where  $h_s : q$  is multiplication of  $s$  and  $q$  as rational numbers.  $S$  is a central semialgebra and a field.

We find that the standard rational natural numbers do appear to have a representation in this framework which is natural in the sense that it is a simple construction from basic combinatorial definitions. It is, however, created via a bespoke intentional process steering towards an intended result, rather than as a space appearing with essentially no effort.

## 12 Number Sequences

Given the organic naturals  $N = (\text{is } a)$ , from Section 9.1, we can go from a space containing single natural numbers to a space containing sequences of natural numbers. Let  $\mathbb{N}$  be

$$\text{ap}(\text{put } n) \cdot N \cdot (\text{get } n) \quad (11)$$

where  $n$  is a pre-defined atom. Since  $(\text{put } n) \cdot N \cdot (\text{get } n)$  is idempotent,  $\mathbb{N}$  is a distributive space containing sequences of  $n$ -atoms containing  $N$ -data, such as

$$T = (n : a a a) (n :) (n : a a) (n : a) (n :). \quad (12)$$

Unlike the case of  $N$ , the action of  $\mathbb{N}$  is merely to concatenate sequences. It's easy, however, to identify natural number sum again as one of several subspaces

- $\text{sum}:T = (n:a a a a a a)$
- $\text{sort}:T = (n:) (n:) (n:a) (n:a a) (n:a a a)$
- $\text{min}:T = (n:)$
- $\text{first}:T = (n:a a a)$

All but the last are algebraic subspaces and, predictably, have established names. Some endomorphisms of  $\mathbb{N}$  are inherited from  $N$  as follows. Define  $\text{inner}:F$  to be

$$\mathbb{N} \cdot (\text{put } n) \cdot F \cdot (\text{get } n) \cdot \mathbb{N} \quad (13)$$

so that  $(\text{inner}:F):X$  means letting  $F$  act on the concatenated  $N$ -contents of  $X$ , returning the result in a single  $n$ -atom. The sum endomorphism above, for instance, is equal to  $\text{inner}:N$ . The construction of Equation 7 works for any space, and so we can define a “functor”  $\text{Seq}$  to be

$$\{\text{put } A\} \cdot B \cdot (\text{get } A) \quad (14)$$

Then  $\mathbb{N}$  is equal to  $(\text{Seq } n:N)$  and given any atom  $s$  and any space  $S$ ,  $(\text{Seq } s:S)$  is the space of  $S$ -values stored in  $s$ -atoms. Similarly, if we define  $\text{inner}$  to be

$$S \cdot \{\text{put } s\} \cdot B \cdot (\text{get } s) \cdot S \quad (15)$$

then  $\text{inner}:f$  is the inner version of an endomorphism  $f$  of  $S$ . Many general constructions of this type are clearly possible. For example, since the endomorphism  $\text{last2} = (\text{inner}:(\text{last } 2))$  sums the last two atoms of a data of  $\mathbb{N}$ ,  $(\text{series } \text{last2})$  generates the fibonacci sequence.

## 13 Sets

We have seen that familiar mathematical structures organically appear as pure data spaces rather than as sets with additional algebraic structure. It is natural to ask whether ordinary sets also appear organically.

Consider  $S = \text{first} \cdot (\text{is } a \ b \ c)$ .  $S$  is a distributive space containing  $\{(), a, b, c\}$  as data. Endomorphisms of  $S$  are  $S \cdot X \cdot S$  for some data  $X$  and, thus, can be any ordinary function from  $\{(), a, b, c\}$  to  $\{(), a, b, c\}$ . Homomorphisms of  $S$  are endomorphisms which map  $()$  to  $()$ . Thus,  $S$  basically is the set  $\{a, b, c\}$  but with slightly awkward complications due to the forced presence of  $()$ . Thematically, we attribute this awkwardness to the “flaw” that even though  $S$  has the key property that  $f \oplus f = f$  for all endomorphisms,  $S$  is not algebraic and, thus, we expect it not to be perfectly mathematical.

Hasse diagram of  $S = \text{sort} \cdot \text{once} \cdot (\text{is } a \ b \ c)$

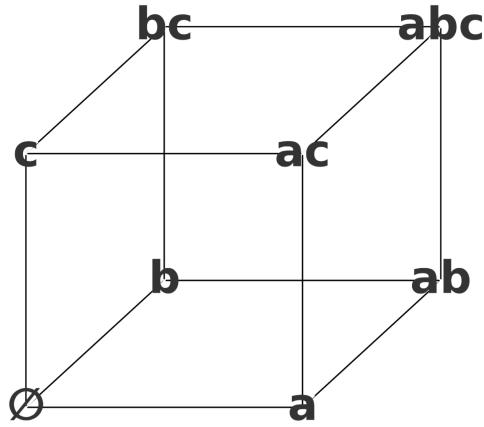


Figure 6: Hasse diagram of the constant endomorphisms of  $S = \text{sort} \cdot \text{once} \cdot (\text{is } a \ b \ c)$ . The data of  $S$  are subsets of  $\{1, 2, 3\}$  ordered by inclusion.

The partial success of  $\text{first} \cdot (\text{is } a \ b \ c)$  suggests considering an *algebraic* space where  $f \oplus f = f$ . These are the *semilattice* spaces. The space  $S = \text{sort} \cdot \text{once} \cdot (\text{is } a \ b \ c)$  is an example. Instead of containing the elements of  $\{a, b, c\}$ ,  $S$  now contains the  $2^3$  subsets  $\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$  represented as sorted non-repeating sequences. The endomorphisms of  $S$  are general endofunctions of  $\mathcal{P}(\{a, b, c\})$ . The semilattice of constants of  $S$  is the standard semilattice of subsets of  $\{a, b, c\}$  ordered by the set union operation. The unit endomorphisms are the permutation group  $S_3$  permuting  $a$ ,  $b$  and  $c$ . The central endomorphisms are  $S_3$ -equivariant functions, and the homomorphisms are morphisms of the standard set union, as shown in Figure 6. In other words,  $S$  has everything that one mathematically expects, without extraneous features or unexpected constraints. Thus, the intuition that algebraic spaces will be more platonic and mathematically perfect is encouraged.

## 14 Bool

As mentioned already, the space  $(\text{bool})$  is likely to be mathematically interesting, both because of its direct connection to the underlying pure data context and, because it has a number of special properties. Although  $\text{bool}$  is not a homomorphism of  $\text{pass}$ , it is algebraic, a central endomorphism, a semilattice, a field, and  $\text{bool}$  is the only space containing exactly two data. Because  $\text{bool}$  only

$f \cdot g$	ID	TRUE	FALSE	NOT	$f \oplus g$	ID	TRUE	FALSE	NOT
ID	ID	TRUE	FALSE	NOT	ID	ID	ID	FALSE	FALSE
TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	ID	TRUE	FALSE	NOT
FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
NOT	NOT	FALSE	TRUE	ID	NOT	FALSE	NOT	FALSE	NOT

Table 1: *Product and sum of the four morphisms ID, TRUE, FALSE, NOT of the space bool.* Note that ID is the unit of multiplication and TRUE is the unit of addition, and we have  $f \oplus g = g \oplus f$  and  $f \oplus f = f$  since bool is a semilattice space.

Table 2: The eight inner morphisms of  $\mathbb{L}_2$  slightly generalize the eight standard symmetric binary boolean operators.

$e_i \in \mathbb{L}_2$	0	T	F	TT	TF	FT	FF	standard	generalized	description
$e_1$	T	T	T	T	T	T	T	TRUE	<b>always</b>	always true
$e_2$	T	T	T	T	T	T	F	OR	<b>any</b>	any are true
$e_3$	T	T	F	T	F	F	T	XNOR	<b>even</b>	even (:)s
$e_4$	T	T	F	T	F	F	F	AND	<b>all</b>	all are true
$e_5$	F	F	T	F	T	T	T	NAND	<b>notall</b>	not all are true
$e_6$	F	F	T	F	T	T	F	XOR	<b>odd</b>	odd (:)s
$e_7$	F	F	F	F	F	F	T	NOR	<b>none</b>	none are true
$e_8$	F	F	F	F	F	F	F	FALSE	<b>never</b>	never true
Number of (:)s	0	0	1	0	1	1	2			

contains: () for *true* and (:) for *false*, there are only four endomorphisms: the identity ID=bool, two constants: TRUE=const (), and FALSE=const (:), and one involution, NOT=bool-not·bool. Table 1 shows the bool semiring. The only proper subspaces of bool are the constants TRUE and FALSE, so bool is a field. The group of units is ID and the involution NOT, so ID is the only central endomorphism.

### 14.1 Boolean Sequences

Repeating the sequencing operation from Section 11, we can let  $\mathbb{L}=(\text{Seq } b:\text{bool})$ , be the distributive space of bool-valued data stored in pre-defined  $b$ -atoms, so a typical data in  $\mathbb{L}$  is

$$(b:) (b:) (b:(:)) (b:) (b:(:)) (b:(:)) \quad (16)$$

We write such sequences replacing  $(b:)$  with T,  $(b:(:))$  with F and the empty sequence with 0, so that the above is written TTFTFF. Consider the shortest subspaces of  $\mathbb{L}$  first. The subspace  $(\mathbb{L} \cdot \text{first} \cdot \mathbb{L})$  contains the  $\mathbb{L}$  sequences of length less than or equal to 1 and similarly,  $(\mathbb{L} \cdot \text{first } 2 \cdot \mathbb{L})$  is the subspace of sequences with length less than or equal to 2. Lets refer to these spaces as  $\mathbb{L}_1$  and  $\mathbb{L}_2$  respectively. Thus,  $\mathbb{L}_1$  contains data {0,T,F} and  $\mathbb{L}_2$  contains data {0,T,F,TT,TF,FT,FF}. For  $\mathbb{L}_1$ , we can specify an endomorphism  $f$  of  $\mathbb{L}_1$  by listing the values of  $f$  on 0,T,F in standard order. So, for example, 0TF denotes the identity endomorphism of  $\mathbb{L}_1$ . Since  $\mathbb{L}_1$  has 27 morphisms, we can be completely explicit about this space. Figure 4 lists the 27 endomorphisms and indicates endomorphisms with special properties. Even with relatively small space like  $\mathbb{L}_1$ , all of the classes of endomorphisms appear in a nontrivial way.

Moving to  $\mathbb{L}_2$ , there are already  $7^7 = 823,543$  endomorphisms; far too many to be as explicit

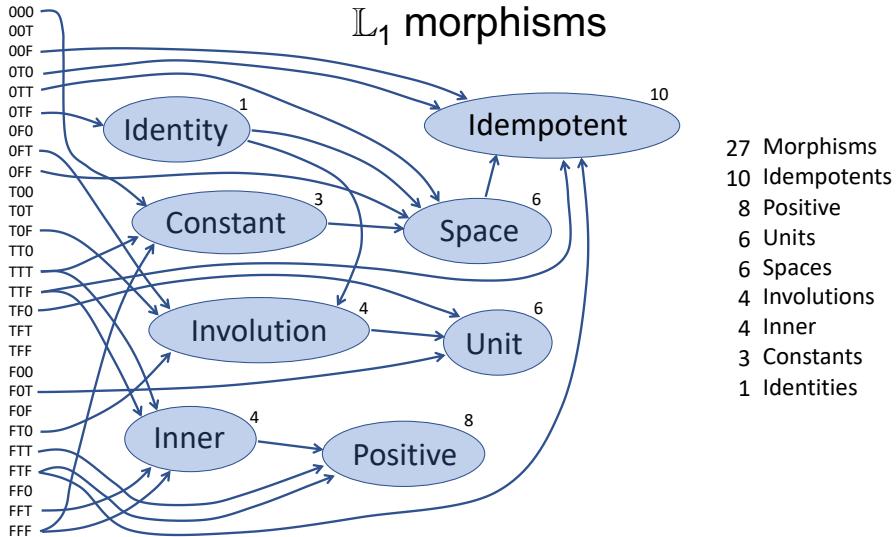


Figure 7: The endomorphisms of  $\mathbb{L}_1$  are maps from  $\{0,T,F\}$  to itself. These are specified as the three function values in the column on the left. Even with a small space of this size, the  $3^3 = 27$  all of the classes of endomorphism identified in the text appear in a non-trivial way.

as in the case of  $\mathbb{L}_1$ . We can, however, at least examine the eight inner endomorphisms

$$\mathbb{L}_2 \cdot (\text{put } b) \cdot X \cdot (\text{get } b) \cdot \mathbb{L}_2 \quad (17)$$

These depend only on the total number of  $(:)$  values in the  $b$ -atoms of input, and each endomorphism returns exactly one  $b$ -atom. The inner endomorphisms of  $\mathbb{L}_2$  are all central homomorphisms, so we have reason to expect them to be mathematically interesting. Indeed, Table 2 shows the eight endomorphisms and their values on  $\{0,T,F,TT,TF,FT,FF\}$ . Looking at the values for the last four entries  $\{TT,TF,FT,FF\}$  we recognize that these are slight generalizations of the eight standard symmetric binary boolean operators. The values of the eight morphisms on  $\{0,T,F\}$  suggests that the standard binary operator names are not perfectly descriptive in this slightly larger context.

Further analysis of  $\mathbb{L}$  is beyond the scope of this paper, but we note that the obvious isomorphism between  $(\text{is a } b)$  and  $\mathbb{L}$  means that all of the semiring structures and subspaces found in  $(\text{is a } b)$  have copies in  $\mathbb{L}$ .

## 15 Discussion

Within the axiomatic framework of pure data, one needs a concept of collections of mathematical object analogous to sets in Set Theory or categories in Category Theory. We find that data representing a collection of other data needs to satisfy an intuitive and simple condition. Such data must be an associative *space*. The morphisms corresponding to spaces are then just products of data which start and end with a space. A space, together with the endomorphisms of the space are a *semiring* which may be *semialgebras*, *central semialgebras* or *fields*. This surprisingly rich structure is shared by all collections ranging from a single atom  $(:)$  to the entire space of all pure data (Figure 3).

After working out the basic theory, we develop a set of examples, partly to illustrate the theory, and partly as an exploration of the general space of mathematical results from the pure

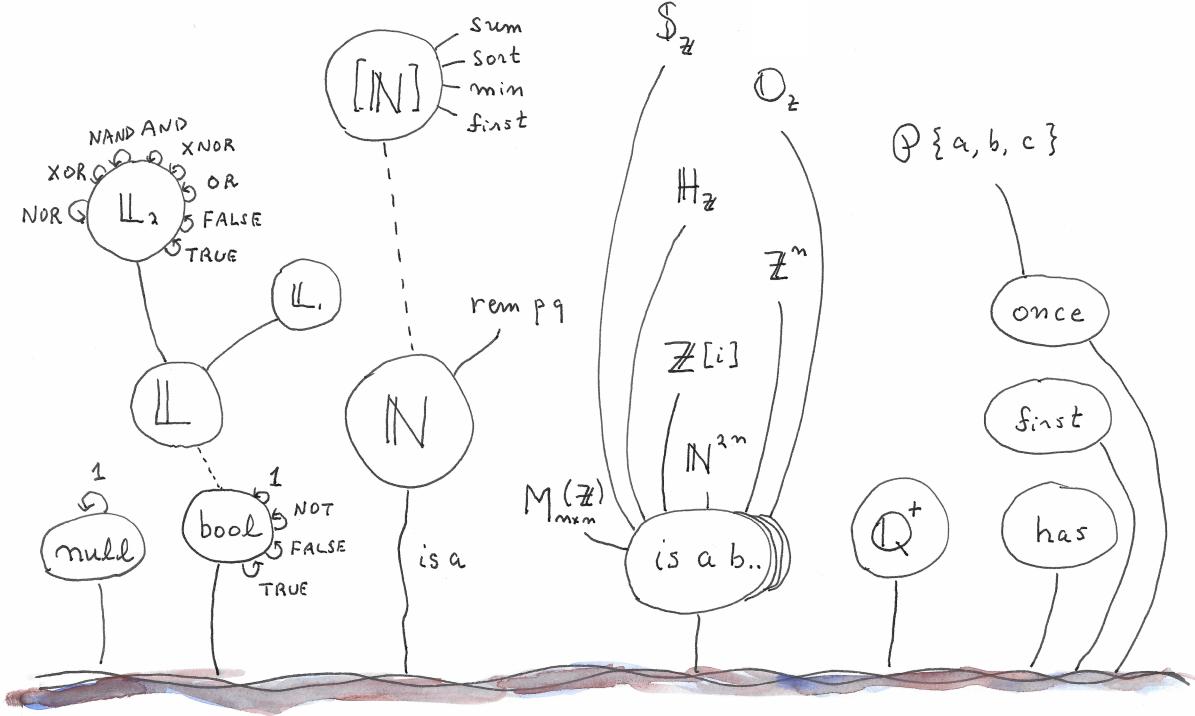


Figure 8: *Spaces and subspaces discussed in the text, viewed as emerging “organically” from (pass), the space of all pure data. Organic numbers are subspaces of (is a..) are shown in the center with bool and descendants on the left, sets and the bespoke  $\mathbb{Q}^+$  construction to the right of the figure. Dashed lines indicate the action of Seq functors.*

data perspective. We adopt the theme of letting spaces emerge “organically” from the simplest combinatorial definitions which are inevitable from the foundational finite sequence concept. A second major theme is to focus on algebraic data which “transcends” finite sequencing, and so might be expected to have platonic meaning that transcends the system. The more sophisticated version of this is focusing on algebraic spaces and on central homomorphisms. Figure 8 illustrates the spaces explored with these considerations in mind. We find that the purely organic approach is at least partially successful. A substantial fraction of natural combinatoric definitions are spaces already, and, for instance, the *organic numbers* show that a substantial fraction of the resulting spaces are important classical algebras. Of course, this is only scratching the surface. Surely many areas of mathematics can be approached from this angle.

Section 10, describing the rational numbers, is an example of a more intentional approach where one has something mathematical in mind from the beginning. This process is interestingly different from what one does in classical mathematics. The rational numbers must be a space, since, basically, everything must be a space. This means that rather than inventing axioms, one starts with a space large enough to encompass what one wants and then try to coax it towards the desired space by passing into subspaces. In this process one is only defining the main addition operation of a space, hoping or trusting that the resulting homomorphisms will end up doing what you intended from the beginning. This seems much more constrained than standard mathematics. On the other hand, one does seem to end up with the intended structure.

One advantage of the pure data approach is that spaces are “born with computation.” If you can define a space, it is also available computationally. Spaces are also theoretically born with a built

<b>width \ depth</b>	0	1	2	3	4
0	1	1	1	1	1
1	1	2	5	26	677
2	1	3	91	$6.9 \times 10^7$	$2.2 \times 10^{31}$
3	1	4	4369	$7.0 \times 10^{21}$	$1.1 \times 10^{131}$
4	1	5	406901	$7.5 \times 10^{44}$	$1.0 \times 10^{359}$

Table 3: The number of pure data within maximum specified width and depth.

in proof infrastructure. In general, data THM is a *theorem* if it is “never false” in the sense that no context greater than the current context causes THM to be false data. For instance, the existence of an isomorphism between space  $S$  and space  $T$  can be expressed as  $\text{ISO} = ()$  for some suitable data ISO. With infinite computing resources, one could search up to say, a specified depth and width, test all data to see if they are spaces and thus find all mathematical structures within some given range of simplicity. Nothing like this, however, has been attempted in this paper. The proofs are human comprehensible and not machine generated or even machine validated. The main reason for this is that even to approach doing proofs computationally requires sophisticated evaluation where one is cleverly moving through the context partial ordering towards a goal. The current software [?] is by no means capable of this. Even with this sort of advance, the resulting computational problems may be difficult. Table 3 illustrates how drastically impractical it is to naively extend even the “organic” search for spaces. Although major advances are needed to make deeper searches and automated proofs practical, continuing to investigate other areas of mathematics in the bespoke style of Section 10 seems quite promising.

## 16 Glossary

### 1. Basics

- (a)  $\delta_{\text{const}} : (\text{const } A : B) \mapsto A$
- (b)  $\delta_{\text{put}} : (\text{put } A : B) \mapsto (A:B)$
- (c)  $\delta_{\text{get}} : (\text{get } A:B) \mapsto \text{ap } (\text{get0 } A:B)$
- (d)  $\delta_{\text{get0}} : (\text{get } A: (A:B)) \mapsto B$
- (e)  $\delta_{\text{atoms}} : (\text{atoms} : b B) \mapsto (:) (\text{atoms} : B)$
- (f)  $\delta_{\text{bin}} : (\text{bin } A:B) \mapsto (\text{bin } A:B), \text{atom}$

### 2. Control

- (a)  $\delta_{\text{if}} : (\text{if } ():B) \mapsto B$
- (b)  $\delta_{\text{if}} : (\text{if } a A:B) \mapsto ()$
- (c)  $\delta_{\text{while}} : (\text{while } A:B) \mapsto B \text{ if } (A:B)=B$
- (d)  $\delta_{\text{while}} : (\text{while } A:B) \mapsto (\text{while } A: A : B)$

### 3. Semiring

- (a) prod
- (b) sum

- 4. Definition
- 5. Combinatorics
  - (a) map
  - (b) various aps
  - (c) aq
  - (d) ar
- 6. Sequence
  - (a)  $(\text{nat}:n) \rightarrow n$  ( $\text{nat}:n+1$ ), example of infinite data. MOVE UP TO TOP?
  - (b)  $\delta_{\text{first}} : (\text{first} : b B) \mapsto b$
  - (c)  $\delta_{\text{first}} : (\text{first} : ()) \mapsto ()$
  - (d)  $\delta_{\text{last}} : (\text{last} : B b) \mapsto b$
  - (e)  $\delta_{\text{last}} : (\text{last} : ()) \mapsto ()$
  - (f) has
  - (g) hasnt
  - (h) is
  - (i) isnt
  - (j) once
  - (k)  $\delta_{\text{rev}} : (\text{rev} : B b) \mapsto b (\text{rev}:B)$
  - (l)  $\delta_{\text{rev}} : (\text{rev} : ()) \mapsto ()$
  - (m) rem
  - (n) sort

## References

- [1] Saul Youssef. Pure Data Foundation of Mathematics and Computing. *arXiv preprint arXiv:2305.05507*, 2023. Available at: <https://arxiv.org/abs/2305.05507>.
- [2] T. Jech. *Set Theory*. Springer, 3rd Millennium ed., 2002.
- [3] Proof of rem...
- [4] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 6th edition, 2008.
- [5] Saul Youssef. Coda: Pure Data Spaces software repository. Available at: <https://github.com/Saul-Youssef/coda>, 2025.
- [6] Robert Geroch. *Mathematical Physics*. University of Chicago Press, 1985. ISBN: 978-0-226-28862-3.
- [7] S. Mac Lane and G. Birkhoff. *Algebra*. Chelsea Publishing Company, 3rd edition, 1999. ISBN: 978-0-8218-1646-6.
- [8] J. C. Baez. The octonions. *Bulletin of the American Mathematical Society*, 39(2):145–205, 2002. Available at: <https://arxiv.org/abs/math/0105155>.