IFSR International Series on Systems Science and Engineering

A.H. Louie

The Reflection of Life

Functional Entailment and Imminence in Relational Biology

Springer

A.H. Louie

The Reflection of Life

Functional Entailment and Imminence in Relational Biology

Springer

A.H. Louie
Ottawa, Ontario
Canada

ISSN 1574-0463
ISBN 978-1-4614-6927-8 ISBN 978-1-4614-6928-5 (eBook)
DOI 10.1007/978-1-4614-6928-5
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013932649

Mathematics Subject Classification (2010): 92B99

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Exordium
An Introduction to Relational Biology

My 2009 book *ML* has garnered some attention and has engendered/ sustained/renewed interest on the subject of relational biology. The journal *Axiomathes* (the theme of which is 'Where Science Meets Philosophy') dedicated a recent issue (volume 21 number 3, September 2011; [Poli 2011]) to discussing the nuances of *ML*. Entitled 'Essays on *More Than Life Itself*', the special topical issue comprises four essays commenting on *ML* and my responses [Louie 2011] to these comments. The growing interest also led to my being invited to conferences to speak on the subject. This Exordium is a representation of one of these lectures. It is included herein as a review, or a 'refresher of the whys and wherefores', as it were, of concepts considered in detail in *ML*.

E.1 The Interrogative Science is an activity based on the interrogative: one poses questions about nature and attempts to gain knowledge by answering these questions.

Aristotle contended that one did not really know a 'thing' (which to Aristotle meant a natural system) until one had answered its '*why?*' with its *αἴτιον* (primary or original 'cause'). In other words, Aristotle's *science* is precisely the subjects for which one seeks the *αἰτία* to the interrogative '*?*'.

Aristotle's original Greek term *αἴτιον* (*aition*) was translated into the Latin *causa*, a word which might have been appropriate initially, but which had unfortunately diverged into our contemporary notion of 'cause', as 'that which produces an effect' (more on this shortly). The possible semantic equivocation may be avoided if one understands that Aristotle's original idea had more to do with 'grounds or forms of explanation', so a more appropriate Latin rendering, in retrospect, would probably have been *explanatio*.

E.2 What Is Life? Biology is the study of life. The ultimate biological question is, then, "What is life?"

This was the question Erwin Schrödinger posed in 1943 and attempted to answer in a series of lectures delivered in Dublin; the corresponding book was published in 1944 [Schrödinger 1944]. With decades of hindsight and further advances in biology, parts of the book may now appear dated. But the originality expressed in this book is not diminished, and the fact that it is still in print is a testimony to its continuing significance.

The Schrödinger question "What is life?" is an abbreviation. A more explicitly posed expansion is

"What distinguishes a living system from a non-living one?"

alternatively,

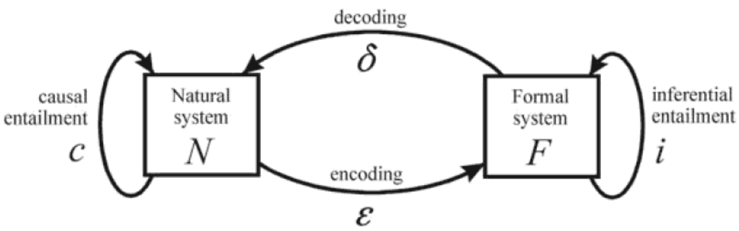
"What are the defining characteristics of a natural system for us to perceive it as being alive?"

These are epistemological forms of the question.

E.3 The Modelling Relation *Causality* in the modern sense, the principle that every effect has a cause, is a reflection of the belief that successions of events in the world are governed by definite relations. *Natural Law* posits the existence of these *entailment* relations *and* that this causal order can be *imaged* by implicative order.

A *modelling relation* is a commutative functorial encoding and decoding between two systems. Between a natural system (an object partitioned from the physical universe) N and a formal system (an object in the universe of mathematics) F, the situation may be represented in the following canonical diagram:

(1)



The encoding ε maps the natural system N and its causal entailment c therein to the formal system F and its internal inferential entailment i; that is,

(2) $$\varepsilon : N \to F \quad \text{and} \quad \varepsilon : c \to i \, .$$

The decoding δ does the reverse. The entailments satisfy the commutativity condition

(3) $$c = \varepsilon \rhd i \rhd \delta \, .$$

(Stated graphically, equality (3) says that, in diagram (1), tracing through arrow c is the same as tracing through the three arrows ε , i , and δ in succession.) Thence related, F is a *model* of N , and N is a *realization* of F . In terms of the modelling relation, then, Natural Law is a statement on the existence of causal entailment c and the encodings $\varepsilon : N \to F$ and $\varepsilon : c \to i$.

A formal system may simply be considered as a *set* with additional mathematical structures. So the mathematical statement $\varepsilon : N \to F$, that is, the posited existence for every natural system N a model formal system F , may be stated as the axiom

Everything is a set.

A *mapping* is an inference that assigns to each element of one set a unique element of another set. In elementary mathematics, when the two sets involved are sets of numbers, the inference process is often called a *function*. So 'mapping' may be considered a generalization of the term, when the sets are not necessarily of numbers. (The use of 'mapping' here avoids semantic equivocation and leaves 'function' to its biological meaning.)

Causal entailment in a natural system is a network of interacting processes. The mathematical statement $\varepsilon : c \to i$, that is, the functorial correspondence [*ML*: A.10] between causality c in the natural domain and inference i in the formal domain, may thus be stated as an epistemological principle, the axiom

Every process is a mapping.

Together, the two axioms are the mathematical formulation of Natural Law. These self-evident truths serve to explain "the unreasonable effectiveness of mathematics in the natural sciences".

E.4 Biology Extends Physics A living system is a material system, so its study shares the material cause with physics and chemistry. Reductionists claim this, therefore, makes biology reducible to 'physics'. *Physics*, in its original meaning of the Greek word *φύσις*, is simply (the study of) *nature*. So in this sense, it is tautological that everything is reducible to physics. But the hardcore reductionists, unfortunately, take the term 'physics' to pretentiously mean '(the toolbox of) *contemporary* physics'.

Contemporary physics that is the physics of mechanisms reduces biology to an exercise in molecular dynamics. This reductionistic exercise, for example, practised in biochemistry and molecular biology, is useful and has enjoyed popular success and increased our understanding life by parts. But it has become evident that there are incomparably more aspects of natural systems that the physics of mechanisms is *not* equipped to explain.

Biology is a subject concerned with organization of relations. Physicochemical theories are only surrogates of biological theories, because the manners in which the shared matter is organized are fundamentally different.

Hence, the behaviours of the realizations of these mechanistic surrogates are different from those of living systems. This in-kind difference is the impermeable dichotomy between *predicativity* and *impredicativity*. (I shall explicate these two antonyms presently.)

In his 1944 book, Schrödinger wrote:

> "... living matter, while not eluding the 'laws of physics' as established up to date, is likely to involve 'other laws of physics' hitherto unknown, which however, once they have been revealed, will form just as integral a part of science as the former."

There have, of course, been many interpretations of what these 'other laws of physics' might have been. Schrödinger himself likely thought of extensions in thermodynamical terms. It is, however, nothing new in the history of physics that 'other laws of physics' have been added to the repertoire from time to time when 'the toolbox of contemporary physics' became inadequate. The mathematical toolbox of calculus was sufficient for Newtonian mechanics. Tensor geometry had to be recruited for relativity. Operator theory was the appropriate mathematical language of quantum physics. I contend that biology extends physics, and to accordingly expand the toolbox, one needs to enlist *category theory*.

Any question becomes unanswerable if one does not permit oneself a large enough universe to deal with the question. The failure of presumptuous reductionism is that of the inability of a small surrogate universe to exhaust the real one. Equivocations create artefacts. The limits of mechanistic dogma are very examples of the restrictiveness of self-imposed methodologies that fabricate non-existent artificial 'limitations' on science and knowledge. The limitations are due to the nongenericity of the methods and their associated bounded microcosms. One learns something new and fundamental about the universe when it refuses to be exhausted by a posited method.

E.5 Relational Biology The study of biology from the standpoint of 'organization of relations' is a subject called *relational biology*. It was founded by Nicolas Rashevsky (1899–1972) in the 1950s, thence continued and flourished under his student Robert Rosen (1934–1998), my PhD supervisor.

The essence of reductionism in biology is to keep the matter of which an organism is made, and throw away the organization, with the belief that, since physicochemical *structure implies function*, the organization can be effectively reconstituted from the analytic material parts.

Relational biology, on the other hand, keeps the organization and throws away the matter; *function dictates structure*, whence material aspects are entailed.

In terms of the modelling relation, reductionistic biology is physicochemical process seeking models, while relational biology is organization seeking realizations. Stated otherwise, reductionistic biology begins with the material system and relational biology begins with the mathematics. Thus, the principles of relational biology may be considered the operational inverse of (and complementary to) reductionistic ideas. It must be emphasized that both approaches are valuable, each answering questions that the other is not equipped to answer. 'Structure implies function' has beneficial epistemological implications, while 'function dictates structure' better addresses ontological issues. What renders hardcore reductionism a falsehood is their practitioners' overreaching claim of genericity, their indignant exclusion of other approaches (which they presumptuously consider to be illegitimate), and their self-declared exclusive ownership of objectivity besides. One world is not enough.

In the relational-biological approach, the answer to our "What is life?" question will define an organism as a material system that realizes a certain kind of relational pattern, whatever the particular material basis of that realization may be. For the remainder of this exposition, I shall proceed to answer this question and use the process of reaching this goal to illustrate the methods of relational biology.

E.6 Mapping and Its Relational Diagram In relational biology, we begin with a formal system, with biology entailed as its realization. So let me begin with a mathematical object, a *mapping* f from set A to set B. It is commonly denoted thus:

(4) $$f : A \to B .$$

The mapping (4) may alternatively be represented in its category-theoretic notation

(5) $$f \in H(A,B),$$

where $H(A,B)$ denotes a set of mappings from set A to set B and is called a *hom-set*. Essentially, (5) says that $H(A,B)$ is a collection of mappings from set A to set B, and f, being a member of this collection, is one such mapping.

Another way to represent the mapping (4) is its 'element-chasing' version: if $a \in A$, $b \in B$, and the variables are related as $b = f(a)$, then one may use the 'maps to' arrow (note the short vertical line segment at the tail of the arrow) and write

(6) $$f : a \mapsto b .$$

Let me introduce a final representation of the mapping f, its *relational diagram in graph-theoretic form*. It may be drawn as a network with three *nodes* and two *directed edges*, that is, a directed graph (or *digraph* for short):

(7)

This graph-theoretic representation allows a ready identification of components of a mapping with the four Aristotelian causes that respond to the interrogative "Why mapping?".

The input $a \in A$ is the *material cause*. The output $b \in B$ is the *final cause*. The *hollow-headed arrow* denotes the *flow* from input $a \in A$ to output $b \in B$, whence the final cause of the mapping may be identified also as the hollow-headed arrow that terminates on the output:

(8)

The *efficient cause* is the *function* of the mapping f as a *processor*; thus, it may be identified as f itself. The *solid-headed arrow* denotes the induction of or constraint upon the flow by the processor f, whence the efficient cause of the mapping may be identified also as the solid-headed arrow that originates from the processor:

(9)

The *formal cause* of the mapping is the ordered pair of arrows:

(10)

that is, the ordered pair of $\langle$ processor, flow $\rangle$.

E.7 Efficient Cause Since the efficient cause will turn out to be the crucial *aition* in relational biology, I shall explicate it further. Aristotle's *κινητικός* (kinetikos) is rendered into *efficare* in Latin: the efficient cause is "one who puts in motion, that which brings the thing into being, the source of change, that which makes what is made, the 'production rule'". Note that efficient cause in the Aristotelian sense is simply 'the processor', and the adjective 'efficient' has nothing to do with its common-usage sense that is 'productive with minimum waste or effort'.

The Natural Law axiom "Every process is a mapping." encodes natural processes into mappings; in particular, the encoding identifies an efficient cause of a natural process with the efficient cause of the corresponding mapping. The isomorphic correspondence between the *solid-headed arrow* (9) and the efficient cause of a mapping then completes the linkage in our formalism. Each statement on entailment thus has three analogous formulations, concerning:

i. Causal entailment patterns among efficient causes of natural processes

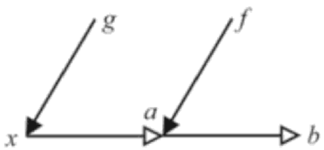
ii. Inferential entailment paths among efficient causes of mappings

iii. Graphical entailment networks among solid-headed arrows

E.8 Compositions The relational diagrams of mappings may *interact*: two mappings, with the appropriate domains and codomains, may be connected at different common nodes.

As a first example, consider $g : x \mapsto a$ and $f : a \mapsto b$; thus, *the output of* g *is the input of* f (the common 'middle' element a). In terms of hom-sets, one has $g \in H(X,A)$ and $f \in H(A,B)$ (where, naturally, $x \in X$, $a \in A$, and $b \in B$); thus, *the codomain of* g *is the domain of* f (the common 'middle' set A). The relational diagrams of these two mappings connect at the common node a as

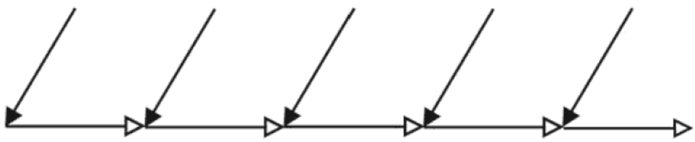
(11)



This *sequential composition* of relational diagrams represents the composite mapping $f \circ g \in H(X,B)$ with $f \circ g : x \mapsto b$.

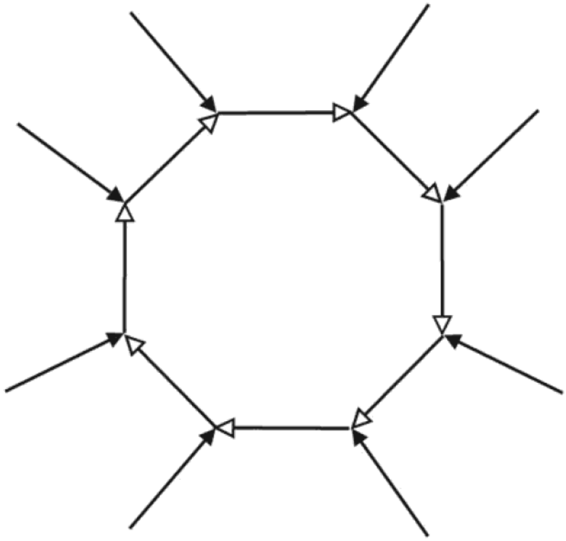
When several mappings are linked by sequential compositions, one has a *sequential chain*:

(12)



When the first and last mappings in a sequential chain are themselves linked by sequential composition, the chain folds up into a *sequential cycle*:

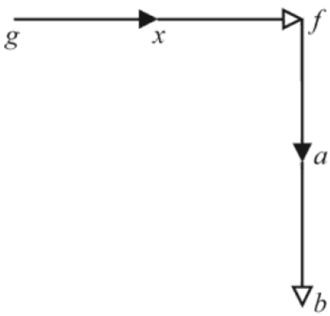
(13)



Note that *within* a sequential cycle, the arrows involved have a consistent direction and are *all hollow-headed* (with solid-headed arrows *peripheral* to the cycle). That is, the compositions involved in the closed path are all sequential, and each final cause has the additional role of being the material cause of the subsequent mapping. A sequential cycle may, therefore, be called a *closed path of material causation*.

Next, consider two mappings g and f with $g : x \mapsto f$ and $f : a \mapsto b$—now *the output of* g *is itself the mapping* f. The hom-sets involved are $g \in H\left(X, H(A,B)\right)$ and $f \in H\left(A,B\right)$: thus, *the codomain of* g *contains* f. Because of this 'containment', the mapping g may be considered to occupy a higher 'hierarchical level' than the mapping f (and that the hom-set $H\left(X, H\left(A,B\right)\right)$ is at a higher hierarchical level than $H\left(A,B\right)$). For these two mappings, one has the *hierarchical composition* of relational diagrams:

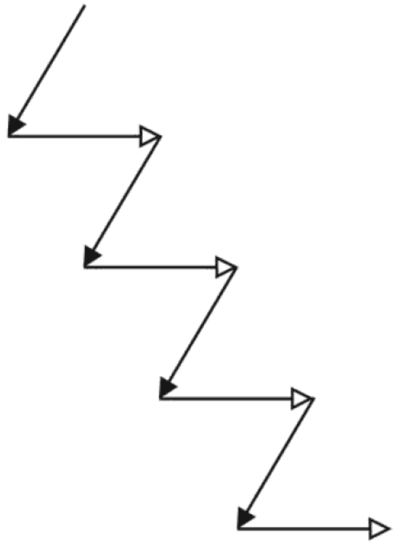
(14)



Since the final cause (i.e. output) of g is the efficient cause of f, the mapping g may be considered an 'efficient cause of efficient cause'. An iteration of efficient causes is inherently hierarchical, in the sense that a lower-level efficient cause is contained within a higher-level efficient cause. In sequential composition, the first mapping g produces something to be operated on, but in hierarchical composition, the first mapping g produces instead an operator itself. Hierarchical composition thus concerns a 'different' mode of entailment, which is given the name of *functional entailment*.

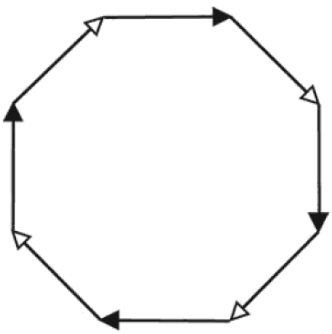
Similar to sequential compositions, hierarchical compositions may form a *hierarchical chain*:

(15)



and a *hierarchical cycle*:

(16)



Note that, in contrast to a sequential cycle (13), *solid-headed arrows* (along with hollow-headed arrows) are definitive components of a hierarchical cycle. Efficient causes are relayed; thus, a hierarchical cycle is a *closed path of efficient causation*.

E.9 Impredicativity In logic, the *predicate* is what is said or asserted about an object. It can take the role as either a property or a relation between entities. Thus, *predicate calculus* is the type of symbolic logic that takes into account the contents (i.e. predicate) of a statement. The defining property $p(x)$ of a subset P in the universe U, as in

(17) $$P = \{x \in U : p(x)\},$$

is an example of a predicate, since it *asserts unambiguously* the property that x must have in order to belong to the set P.

Contrariwise, a definition of an object is said to be *impredicative* if it invokes (mentions or quantifies over) the object itself being defined, or perhaps another set which contains the object being defined. In other words, *impredicativity* is the property of a *self-referencing definition* and may *entail ambiguities*. An impredicative definition often appears circular, as what is defined participates in its own definition.

Impredicative definitions usually cannot be bypassed and are mostly harmless. But there are some that lead to paradoxes. The most famous of a problematic impredicative construction is Russell's paradox, which involves the set of all sets that do not contain themselves:

(18) $$\{x : x \notin x\}.$$

(This foundational difficulty is only avoided by the restriction to a naive set-theoretic universe that explicitly prohibits self-referencing constructions.)

It is evident that a hierarchical cycle, with its cyclic collection of mutually entailing efficient causes, is impredicative. In other words, a hierarchical cycle is an *impredicative cycle of inferential entailment*. A closed path of efficient causation must form a hierarchical cycle of containment: both the hierarchy of containment and the cycle are essential attributes of this closure.

Through the encoding that identifies an efficient cause of a natural process with the efficient cause of the corresponding mapping, one may conclude that

A natural system has a model containing a hierarchical cycle
if and only if it has a closed path of efficient causation.

Stated otherwise, a hierarchical cycle is the relational diagram in graph-theoretic form of a closed path of efficient causation.

E.10 Nonsimulability An *algorithm* is a computation procedure that requires in its application a *rigid stepwise mechanical execution of explicitly stated rules*. It is presented as a prescription, consisting of a finite number of instructions. It halts after a finite number of steps. It has no room for ambiguity.

Predicates are algorithmic. Impredicativity is everything that an algorithm is *not*.

A mapping is *simulable* if it is definable by an algorithm. A formal system, an object in the universe of mathematics, may be considered a collection of mappings connected by the system's entailment pattern (i.e. its graph, which may itself be considered a mapping). So by extension, a formal system is *simulable* if its entailment pattern and all of its mappings are simulable. Simulability entails finiteness: that the corresponding Turing machine halts after a *finite* number of steps, that the corresponding algorithmic process is of *finite* length, and that the corresponding program is of *finite* length.

Impredicativity has many consequences. In view of its being the antithesis of things algorithmic, one of these consequences is, therefore, nonsimulability.

Among the entailment networks (12), (13), (15), and (16) that we have considered, the first three, namely, sequential chain, sequential cycle, and hierarchical chain, are simulable, but the last one, hierarchical cycle, is not. The nonsimulability of a hierarchical cycle has been proven using lattice theory. I state this theorem formally as

A formal system that contains a hierarchical cycle is not simulable.

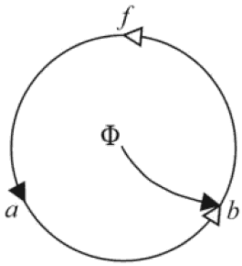
For natural systems, a *deadlock* is a situation wherein competing actions are waiting for one another to finish, and thus none ever does. A set of processes is in a deadlock state when every process in the set is waiting for an event that can be *caused* only by another process in the set. This is a realization, a relational analogue, of impredicativity. In computer science, deadlock refers to a specific condition when two or more processes are each waiting for another to release a resource, or more than two processes are waiting for resources in a circular chain. Implementation of hierarchical cycles (or attempts to execute ambiguous codes in general) will lead a program to either a deadlock or an endless loop. In either case, the program does not terminate. This is practical verification that a hierarchical cycle is *not* simulable.

E.11 Biological Realization: Metabolism-Repair System Every process is a mapping. The crucial biological process of *metabolism* may, therefore, be represented as a mapping $f : a \mapsto b$ (equivalently, $f \in H(A,B)$); an *enzyme* may be the realization of the efficient cause f, with material input and output metabolites realizations of a and b. Networks of mappings in sequential composition are, then, models of metabolic pathways.

Some biochemical processes produce enzymes as outputs. Such a process may naturally be modelled as a mapping of the form $\Phi : x \mapsto f$ (equivalently, $\Phi \in H(X, H(A,B))$). The morphism Φ may be considered *repair*: its codomain is $H(A,B)$, so it is a mapping that creates new copies of enzymes f, hence a *gene* that 'repairs' (or replenishes) the metabolism process. The repair map Φ and the metabolism map f are thus in hierarchical composition.

A typical eukaryotic cell is compartmentalized into two observably different regions, the cytoplasm and the nucleus. Metabolic activities mainly occur in the cytoplasm, while repair processors (i.e. genes) are contained in the nucleus. Repair in cells generally takes the form of a continual synthesis of basic units of metabolic processor (i.e. enzymes), using as inputs materials provided by the metabolic activities themselves. In particular, the simplest domain of the repair map Φ may be the codomain of metabolism f, the latter's 'output set' B (i.e. $\Phi: b \mapsto f$, $\Phi \in H(B, H(A,B))$), whence metabolism and repair combine into the relational diagram

(19)



This geometry gives a graphic representation of the metabolism component as the abstract equivalent of 'cytoplasm' and the repair component as the abstract counterpart of 'nucleus'.

What if the repair components themselves need repairing? New mappings representing *replication* (serving to replenish the repair components) may be defined. A replication map must have as its codomain the hom-set $H(X, H(A,B))$ to which repair mappings Φ belong, so it must be of the form

$$\beta : Y \to H\big(X, H(A,B)\big) \tag{20}$$

for some set Y (where Y contains ingredients already present in the cell). In the simplest case, when $X = B$, one may choose $Y = H(A,B)$; so (20) becomes

$$\beta : H(A,B) \to H\big(B, H(A,B)\big). \tag{21}$$

It turns out that under stringent but not prohibitively strong conditions, the replication mapping β may already be entailed within the components present. There are many ways in which this happens; one natural way is that an isomorphic correspondence may be defined between b and β, whence the mapping (21) may be equivalently represented as

(22) $$b : f \mapsto \Phi .$$

The relational diagram of the entailment among the metabolism-repair-replication mappings

(23) $$\{ f : a \mapsto b,\ \Phi : b \mapsto f,\ b : f \mapsto \Phi \}$$

is then

(24)

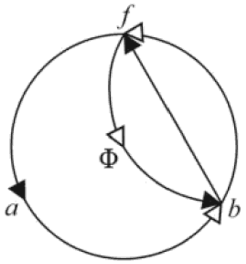
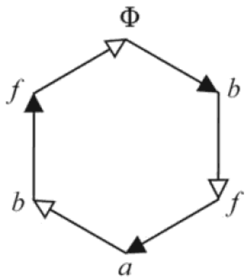


Diagram (24) is the relational diagram in graph-theoretic form of the simplest *metabolism-repair system* (or (M,R)-*system* for short), introduced by Robert Rosen in the late 1950s.

Note that (24) is a hierarchical cycle. The entailment pattern is more evident when the relational diagram is unfolded thus:

(25)

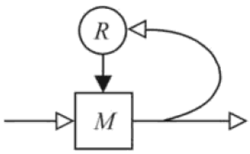


One may also note that there is no 'privileged' position of any of the three mappings involved. They are in cyclic entailment and may be assigned the labels of metabolism, repair, and replication in any cyclic permutation. The all-important feature is that the mappings form a hierarchical cycle; stated otherwise, the simplest (M,R)-system is a hierarchical-cycle model of a cell.

In the specialization of the replication map β from (20) to (21), many simplifying assumptions have been made to create the three-mapping $\{f, \Phi, b\}$ hierarchical cycle. A more sophisticated (M,R)-system model of a cell would contain a large number of metabolism and repair components connected in a complex entailment network, since in a cell there are obviously many more than three interacting processes. (Diagram (24) actually already captures the essence of all (M,R)-systems, and indeed it is possible in principle to reduce every abstract (M,R)-system to this simple form by making the three mappings involved sufficiently complex. One must, nevertheless, not lose sight of the network aspect of (M,R)-systems.)

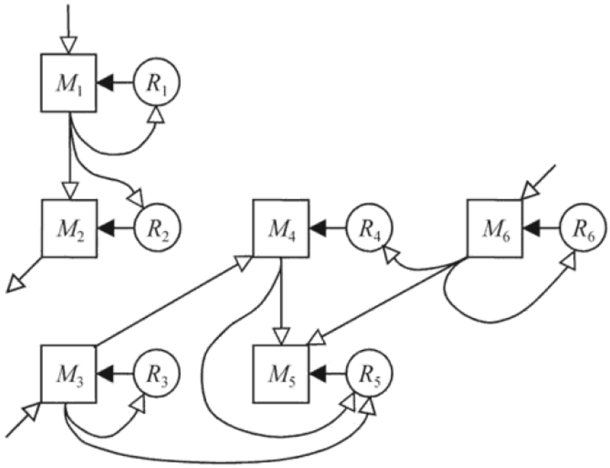
Metabolism may alternatively be considered an input-output system, with the mapping f representing the transfer function of the 'block', the domain A as the set of inputs, and the codomain B as the set of outputs. Similarly, *repair* may be considered an input-output system, with the mapping Φ representing the transfer function of the block, the domain B as the set of inputs, and the codomain $H(A,B)$ as the set of outputs. With the addition of entailment arrows for environmental inputs and outputs, and the abbreviated representation by the symbols M and R of the components, the relational diagram (19) may be represented as this simple network of one metabolism component and one repair component:

(26)



In general, a metabolism-repair network consists of many metabolism and repair components, with the requisite connections that the outputs of a repair component are observables in the hom-set of its corresponding metabolism component; the metabolism components may be connected among themselves by their inputs and outputs; and repair components must receive at least one input from the outputs of the metabolism components of the network. The following is a sample (M,R)-network (still relatively simple) with six pairs of metabolism-repair components:

(27)



One may easily visualize larger (M,R)-networks with thousands of components.

E.12 Closure to Efficient Causation Suppose a natural system *contains* a closed path of efficient causation, then *some* of its efficient causes are in cyclic entailment of one another. Their corresponding mappings must then form a hierarchical cycle. If it so happens that *all* of a natural system's efficient causes entail one another, then it must have a model in which *all* solid-headed arrows are components of hierarchical cycles (e.g. diagram (24) of the simplest (M,R)-system). Having *all* efficient causes entailed within the system is a more stringent requirement than having just *some*, and members of this subset of natural systems are given a special description: *closed to efficient causation*.

> A natural system is *closed to efficient causation* if its every efficient cause is entailed within the system.

The correspondence between an efficient cause and a solid-headed arrow implies:

> A natural system is *closed to efficient causation* if and only if each connected component in its relational diagram has a closed path that contains all the solid-headed arrows.

I mention in passing that "a closed path that contains all the solid-headed arrows" is related to the concept of *traversability* (one continuous trace of the edges in a graph, passing along each edge exactly once) in network topology. Thus, the study of 'closed to efficient causation' can make use of the powerful results from the mathematical theory of topology (in addition to lattice theory and category theory that we have already encountered).

Not all metabolism-repair networks satisfy the stringent requirements for entailment closure. The defining characteristic of an *(M,R)-system* that makes it a model of cells is the self-sufficiency in the networks of metabolism and repair components, in the sense that every mapping is entailed within, in short, closure to efficient causation.

The answer to our “What is life?” question according to the Rashevsky-Rosen school of relational biology, in a nutshell, is that an *organism*—the term is used in the sense of an ‘autonomous life form’, that is, any living system (including, in particular, cells)—admits a certain kind of relational description, that it is ‘closed to efficient causation’. Explicitly:

A material system is an organism
if and only if it is closed to efficient causation.

This ‘self-sufficiency’ in efficient causation is what we implicitly recognize as the one feature that distinguishes a living system from a nonliving one.

In terms of (M,R)-systems, we may state the **Postulate of Life**:

A natural system is an organism
if and only if it realizes an (M,R)-system.

Thus, an (M,R)-system is the very model of life, and, conversely, life is the very realization of an (M,R)-system.

Prolegomenon
Cardinalis

> Not everything that counts can be counted, and not everything that can be counted counts.
>
> — attributed to Albert Einstein

As I did in *ML*, in this book, I assume that the reader is familiar with the basic facts of *naive set theory*, as presented, for example, in Halmos [1960]. In this prologue, however, I shall present some set-theoretic and logical preliminaries; this is more for the clarity of notations (especially for those non-standardized ones) than for the concepts themselves.

Sets

0.1 Definition If A and B are sets and if every element of A is an element of B, then A is a *subset* of B, denoted

(1) $$A \subset B\,.$$

Note that this symbolism of containment means *either* $A = B$ (which means the sets A and B have the same elements; *ML*: 0.2: Axiom of Extension) *or* A is a proper subset of B (which means that B contains at least one element that is not in A). Two sets A and B are equal if and only if $A \subset B$ and $B \subset A$ (*ML*: 0.4).

0.2 Definition If X is a set, the *power set* $\mathcal{P}X$ of X is the family of all subsets of X.

An alternate notation of the power set $\mathcal{P}X$ is 2^X (*cf.* *ML*: A.3 for the etymology).

0.3 Definition The *relative complement* of a set A in a set B is the set of elements in B but not in A:

(2) $$B \sim A = \{x \in B : x \notin A\}.$$

When B is the 'universal set' U (of some appropriate universe under study, e.g. the set **N** of all natural systems), the set $U \sim A$ is denoted A^c, that is,

(3) $$A^c = \{x \in U : x \notin A\},$$

and is called simply the *complement* of the set A.

0.4 Number Sets Various sets of numbers are denoted thus:

i. *Natural numbers* ('positive integers') $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$
ii. *Whole numbers* ('nonnegative integers') $\mathbb{N}_0 = \{0\} \cup \mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}$
iii. *Integers* $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
iv. *Rational numbers* ('fractions') $\mathbb{Q} = \left\{ \frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{N} \right\}$
v. *Real numbers* $\mathbb{R}$
vi. *Complex numbers* $\mathbb{C}$

The six number sets are related by

(4) $$\mathbb{N} \subset \mathbb{N}_0 \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

in which all containments are proper.

Equipotence

0.5 Definition Two sets are *equipotent* (to each other) if there exists a bijective mapping, that is, a one-to-one correspondence, between them (*cf. ML*: 1.8).

Stated otherwise, two sets are equipotent if they are isomorphic in the category **Set** (*cf. ML*: A.6). It is evident that equipotence is an equivalence relation (*ML*: 1.11). The symmetry of the relation also allows the usage 'set A is equipotent to set B', since it implies 'set B is equipotent to set A', whence A and B are equipotent to each other. One also occasionally sees the usage of 'equipollent', or even 'equinumerous', for the same concept.

0.6 Schröder-Bernstein Theorem *If each of two sets is equipotent to a subset of the other, then the two sets are equipotent.*

Since every set itself is its own subset, the converse of the Schröder-Bernstein Theorem, that if two sets are equipotent then each is equipotent to a subset of the other, is trivially true.

0.7 Law of Trichotomy of Equipotence *Two sets are either equipotent to each other, or one is equipotent to a subset of the other.*

If two sets are equipotent, then it is easy to see that their power sets are equipotent. But a set is never equipotent to its own power set; this is

0.8 Cantor's Theorem *Every set is equipotent to a proper subset of its power set, but is not equipotent to the power set itself.*

Cardinality

0.9 Definition A set is *finite* if it is either empty or equipotent to the set $\{0,1,2,...,n-1\}$ for a natural number n; otherwise it is *infinite*. An infinite set that is equipotent to the set $\mathbb{N}$ of all natural numbers is called *countably infinite*; otherwise the infinite set is *uncountable*. The term *countable* means either finite or countably infinite.

With the formal definition $0=\varnothing$ and $n=\{0,1,2,...,n-1\}$ for $n\in\mathbb{N}$, a finite set is equivalently 'equipotent to a whole number'. Each finite set X is equipotent to a *unique* whole number $|X|=n\in\mathbb{N}_0$, the 'number of elements of X'. In short, a finite set is a set consisting of a finite number of elements.

The property that each finite set is equipotent to a unique whole number may be extended to infinite sets. The generalized 'number of elements' of a set is called its *cardinality*, and formally one has the

0.10 Property Every set is equipotent to a unique *cardinal number*.

I will not go into the formal definition of cardinal number (and its related concept ordinal number) here. The interested reader may read Halmos [1960]. The usual partial order $\leq$ of whole numbers may be extended to all cardinal numbers. One uses the same notation $|X|=n$ for the cardinality of the set X, where n may be an 'infinite cardinal' in addition to a whole number. Infinite cardinal numbers are usually denoted by the first letter $\aleph$ (*aleph*) of the Hebrew alphabet. When $|X|=n$, one may simply say 'X has cardinal number n' or 'X has cardinality n'.

When $|X|=n$, a bijective mapping from the cardinal number n (as a set) to the set X is called an *enumeration*, a 'listing of the elements' of X. While 'to enumerate' literally means 'to count out' (i.e. 'to have a number as output'), the domain of an enumeration may be any cardinal number, countable or uncountable. The enumeration map is not uniquely defined by the correspondence $n \leftrightarrow X$, since any permutation of the assignment also serves as a bijection (each different permutation—there being $n!$ of them for finite n—defining its own distinct listing of elements of the set X).

0.11 Theorem

i. *Every set has a cardinal number.*
ii. *Two sets A and B are equipotent if and only if they have the same cardinal number*, that is, *iff* $|A| = |B|$.
iii. $|A| \le |B|$ *if and only if A is equipotent to a subset of B.*
iv. $|A| < |B|$ *if and only if A is equipotent to a subset of B but B is not equipotent to a subset of A.*

Some trivial properties of finite sets are:

0.12 Corollary

i. *Every finite set has a unique number of elements.*
ii. *Two finite sets are equipotent if and only if they have the same number of elements.*
iii. *If a set is finite, then every one of its subsets is finite.*
iv. *If a finite set X has n elements and a subset $A \subset X$ has k elements, then $k \le n$; further, $k = n$ iff $A = X$.*
v. *If a set is finite, then it is not equipotent to any of its proper subsets.*

Property v, that a finite set is not equipotent to any of its proper subsets, in fact characterizes finite sets. The inverse thus characterizes infinite sets; stated formally:

0.13 Theorem

i. *A set is infinite if and only if it is equipotent to a proper subset of itself.*
ii. *A set is finite if and only if it is not equipotent to any proper subset of itself.*

One also has the following concerning countability:

0.14 Lemma

i. *Every subset of a countable set is countable.*
ii. *Every infinite set has a countably infinite subset.*

0.15 Cardinality of the Power Set If $|X| = n$, then $|\mathrm{P}X| = 2^n$ (for all cardinal numbers n, finite and infinite). The proof is immediate from the fact that $\mathrm{P}X$ is equipotent to $\mathbf{Set}(X,2)=2^X$, the hom-set of all mappings from X to $2=\{0,1\}$ (*cf. ML*: A.3). One may succinctly write

(5) $$\left|2^X\right| = 2^{|X|}.$$

0.16 Cantor's Continuum Hypothesis The cardinality of the set $\mathbb{N}$ of all natural numbers (whence of all countably infinite sets) is denoted by $\aleph_0$. In view of Lemma 0.14.ii, $\aleph_0$ is, then, the least infinite cardinal number. Analogously, the least uncountable cardinal number is usually denoted by $\aleph_1$. In terms of the canonical order relation $\leq$ of cardinal numbers, $\aleph_1$ is the *least* cardinal number strictly following $\aleph_0$.

Cantor's Theorem (Theorem 0.8) dictates that the set $\mathbb{N}$ is equipotent to a subset of its power set $\mathcal{P}\mathbb{N}$, but is not equipotent to $\mathcal{P}\mathbb{N}$ itself. Whence, it follows from Theorem 0.11.iv and Section 0.15 that $\aleph_0 < \left|\mathcal{P}\mathbb{N}\right| = 2^{\aleph_0}$. Since $\aleph_1$ is the least cardinal number larger than $\aleph_0$, one must have

(6) $$\aleph_1 \leq 2^{\aleph_0}.$$

In his famous *continuum hypothesis*, Cantor conjectured that $\aleph_1 = 2^{\aleph_0}$. (The word 'continuum' is used because $2^{\aleph_0}$ is also the cardinal number of the set $\mathbb{R}$ of all real numbers, the 'cardinality of the continuum', usually denoted $\left|\mathbb{R}\right| = c$.)

The consistency of the continuum hypothesis with the usual axioms of set theory has been proven, that is, the equality $\aleph_1 = 2^{\aleph_0}$ is non-contradictory. It has likewise been proven that the continuum hypothesis is independent of the usual axioms of set theory; that is, the inequality $\aleph_1 < 2^{\aleph_0}$ is also non-contradictory.

Indexed Sets

0.17 Indexed Family Let I and X be sets. A *family of elements in* X *indexed by* I is a mapping $x : I \to X$. The domain I is called the *index set* (note the noun adjunct 'index'), an element $i \in I$ is called an *index*, the range $x(I) \subset X$ is called an *indexed set* (note the past participle 'indexed'), and the value $x(i)$ of the mapping x at an index i, written as x_i (whence the element-chasing form of the mapping x may be written as $x : i \mapsto x_i$), is a *term* (or more precisely 'the i th term') of the family. Such a mapping is often denoted

(7) $$\{x_i\}_{i\in I},$$

and the mapping is also called an *indexed family* (*in* X). Note that notation (7) represents the indexed family, which is a mapping, whence $\{x_i\}_{i\in I} \in X^I$), while the notation

$$\{x_i : i \in I\} \tag{8}$$

represents the indexed set (i.e. the range of the indexed family, whence $\{x_i : i \in I\} = x(I) \subset X$. The notion of 'the i th term' only makes sense with respect to the indexed family (7) but not the indexed set (8). Occasionally one may simply use $\{x_i\}$ for (7) if the index set I is implicitly understood, but this is not good notation (although it is commonly accepted) because of the possible equivocation between the two different entities (7) and (8), essentially the identification of a mapping with its range.

One may also note that the mapping $x : I \to x(I)$ is surjective (which is simply the statement that a mapping maps *onto* its range), but an indexed family is not required to be injective. Explicitly, it may happen for $i, j \in I$ that $i \neq j$ but $x_i = x_j$; that is, there may be 'duplicated terms'.

0.18 Indexed Family of Sets An *indexed family of sets* is an indexed family $A : I \to \mathrm{P}X$ (of elements in $\mathrm{P}X$), denoted

$$\{A_i\}_{i\in I}, \tag{9}$$

where each $A_i \subset X$.

0.19 Indexed Partition An *indexed partition* of a set X is an indexed family of *nonempty* sets $A : I \to \mathrm{P}X$ for which the collection of subsets $\{A_i : i \in I\}$ forms a *partition* of X (*cf. ML*: 1.16), that is, for each $i \in I$, $A_i \neq \varnothing$, and

$$X = \bigcup_{i\in I} A_i \tag{10}$$

with

$$A_i \cap A_j = \varnothing \quad \text{for} \quad i \neq j. \tag{11}$$

0.20 Axiom of Choice If $\{A_i\}_{i\in I}$ is an indexed family of nonempty sets indexed by a *nonempty* index set I, then there exists an indexed family $\{x_i\}_{i\in I}$ such that for each $i\in I$, $x_i\in A_i$.

Compare this with the equivalent statement from *ML*: 1.37: Given a nonempty family $\mathfrak{A}$ of nonempty sets, there is a mapping f with domain $\mathfrak{A}$ such that $f(A)\in A$ for all $A\in\mathfrak{A}$. The correspondence is $\mathfrak{A}=\{A_i : i\in I\}$. The mapping $f:\mathfrak{A}\to\bigcup_{A\in\mathfrak{A}} A$ (i.e. $f:\{A_i : i\in I\}\to\bigcup_{i\in I} A_i$) defined by

$$f(A_i)=x_i \tag{12}$$

is called a *choice mapping*. When the index set I is finite, choosing, for each $i\in I$, an x_i from a nonempty set A_i (i.e. defining a choice mapping f) is a simple procedure; not so when I is infinite. There is no prescription of how infinitely many choices are to be made, and that is why the existence of the choice mapping has to be postulated in an axiom. It is almost a convention in mathematics that one explicitly acknowledges when a consequence depends on the Axiom of Choice.

Sequences

0.21 Sequence An indexed family $\{x_i\}_{i\in I}$ in X with an index set $I=\{1,2,\ldots,n\}$ (for some natural number n) or $I=\mathbb{N}$ is called a *sequence* (*finite* or *infinite*, respectively) *in* X.

A finite sequence is often written as a list of its terms:

$$\{x_i\}_{i\in\{1,2,\ldots,n\}}=\{x_1,x_2,\ldots,x_n\}; \tag{13}$$

so also is an infinite sequence:

$$\{x_i\}_{i\in\mathbb{N}}=\{x_1,x_2,x_3,\ldots\}. \tag{14}$$

Note that in the listing of the elements on the right-hand side of each of (13) and (14), the distinction between indexed family and indexed set is already somewhat blurred (again, this is not good notation but is commonly accepted). A caveat of the blurred listing notation is that in an indexed family, duplicated terms are kept, while in an indexed set, duplicated terms are (almost) always eliminated. Consider, for example, a finite sequence of two vectors v_1 and v_2, with the two vectors identical and nonzero, that is, $v_1 = v_2 \neq 0$. The *sequence* (i.e. indexed family) $\{v_1, v_2\}$ is linearly dependent, but the *set* $\{v_1, v_2\} = \{v_1\}$ is linearly independent, since it consists actually of just one nonzero vector.

0.22 Preorder The terms of finite and infinite sequences are well-ordered (*ML*: 3.39) by the natural order of integers of their index sets ($\{1, 2, ..., n\}$ and $\mathbb{N}$ respectively). Thus, one may *truncate*, for example, an infinite sequence $\{x_1, x_2, x_3, ...\}$ *after m terms* (where $m \in \mathbb{N}$) to split off the infinite 'tail end' $\{x_{m+1}, x_{m+2}, x_{m+3}, ...\}$ and obtain the finite sequence $\{x_1, x_2, ..., x_m\}$. One may say, as another example, that a term x_i *precedes* another term x_j if $i < j$ (and that x_j *follows* x_i in the sequence).

It is important to note that the ordering of the terms in a sequence $\{x_i\}$ has to do with the *positions* of the terms, and not the ordering of the elements themselves in the indexed set. This is because the codomain X is not necessarily equipped with an order, and unless it is, a statement such as $x_i \leq x_j$ is meaningless.

There is, however, a way to *define* the binary relation of *precedence* on the *range* $x(I) = \{x_i : i \in I\}$ using the order inherent in $I \subset \mathbb{N}$, by

(15) $$x_i \preceq x_j \text{ in } x(I) \quad \text{iff} \quad i \leq j \text{ in } \mathbb{N}.$$

Note that the relation $\preceq$ is defined only on the range $x(I)$ and not on the rest of the codomain $X \sim x(I)$. It is easy to see that the relation of precedence on $x(I)$ is reflexive and transitive (*ML*: 1.10), but not necessarily either antisymmetric or symmetric. A relation that is reflexive and transitive is called a *preorder* (something that is 'not quite' a partial order, *ML*: 1.20, or an equivalence, *ML*: 1.11). A set equipped with a preorder is called a *preordered set* or *proset*. (I shall revisit binary relations, especially these with special properties, in Chapter 3 of this book.)

Each preordered set $\langle S, \preceq \rangle$ is itself a category (*cf. ML*: A.1). This category S has objects the elements of S, and for $a, b \in S$, the hom-set $S(a, b)$ either contains a single S-morphism or is empty, depending on whether $a \preceq b$ or not. Transitivity of $\preceq$ provides for the composition of morphisms, and reflexivity provides the identity morphisms in $S(a, a)$. (I have discussed *poset*, i.e. partially ordered set, as category in *ML*: 1.31, but indeed a *proset* suffices; the antisymmetry is not needed.)

0.23 Monotonic Sequence As a mapping $x : \langle I, \leq \rangle \to \langle x(I), \preceq \rangle$ of prosets, the sequence $\{x_i\}$ preserves the ordering relation by the very definition of $\preceq$ in (15), and is therefore a morphism in the category of prosets and order-preserving mappings. But the order-preserving property of a sequence $\{x_i\}$ may also exist, as a mapping $x : \langle I, \leq \rangle \to \langle x(I), \preceq \rangle$ of prosets, when the codomain X is already equipped with its own preorder $\preceq$ (even when $\preceq$ is not the precedence defined on $x(I) \subset X$; in particular, when $\langle X, \preceq \rangle$ is in fact a *poset*). As the mapping $x : \langle I, \leq \rangle \to \langle X, \preceq \rangle$ of prosets, the sequence $\{x_i\}$ is *isotone* (*cf. ML*: 1.23) if

(16) $$i \leq j \text{ in } I \quad \Rightarrow \quad x_i \preceq x_j \text{ in } X .$$

An isotone sequence $\{x_i\}$ is more commonly called *monotonically increasing*, and implication (16) is equivalent to

(17) $$x_i \preceq x_{i+1} \quad \text{for} \quad i \in \mathbb{N}$$

(or for $i \in \{1, 2, ..., n-1\}$ in the case of a finite sequence). (The sequence $\{x_i\}$ is, of course, monotonically increasing with respect to the relation of precedence on $x(I)$.) If the mapping $x : \langle I, \leq \rangle \to \langle X, \preceq \rangle$ is order reversing (i.e. 'antitone'), then

(18) $$i \leq j \text{ in } I \quad \Rightarrow \quad x_i \succeq x_j \text{ in } X ,$$

which is equivalent to

(19) $$x_i \succeq x_{i+1} \quad \text{for} \quad i \in \mathbb{N}$$

(or, again, for $i \in \{1, 2, ..., n-1\}$ in the case of a finite sequence); such is a *monotonically decreasing* sequence.

If the ordering in (17) is strict, that is, $x_i \prec x_{i+1}$ which means '$x_i \preceq x_{i+1}$ and $x_i \neq x_{i+1}$' (*ML*: 1.22), then the sequence $\{x_i\}$ is *strictly increasing*. Likewise, a strict inequality $x_i \succ x_{i+1}$ ($x_i \succeq x_{i+1}$ and $x_i \neq x_{i+1}$) in (19) defines a *strictly decreasing* sequence. The class of *monotonic sequences* consists of all the increasing and the decreasing sequences.

0.24 Subsequence There is an important way of obtaining new sequences from a given infinite sequence $\{x_i\}_{i\in\mathbb{N}} = \{x_1, x_2, x_3, \ldots\}$. Let $\{n_k\}_{k\in\mathbb{N}} = \{n_1, n_2, n_3, \ldots\}$ be an infinite sequence in $\mathbb{N}$ such that

(20) $$n_{k+1} > n_k \quad \text{for} \quad k \in \mathbb{N}$$

(i.e. the sequence $n : \langle \mathbb{N}, \leq \rangle \to \langle \mathbb{N}, \leq \rangle$ is strictly increasing). The sequence $k \mapsto x_{n_k}$ is called a *subsequence* of the sequence $\{x_i\}_{i\in\mathbb{N}} = \{x_1, x_2, x_3, \ldots\}$ and is denoted

(21) $$\{x_{n_k}\}_{k\in\mathbb{N}} = \{x_{n_1}, x_{n_2}, x_{n_3}, \ldots\}.$$

One may see that the subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ is simply the (sequential) composition (*ML*: 5.13) of the mapping $k \mapsto n_k$ (in $\mathbb{N}^{\mathbb{N}}$) followed by the mapping $n \mapsto x_n$ (in $X^{\mathbb{N}}$).

One may also readily verify that every sequence $\{x_i\}$ is a subsequence of itself, and if $\{z_i\}$ is a subsequence of $\{y_i\}$ and $\{y_i\}$ is a subsequence of $\{x_i\}$, then $\{z_i\}$ is a subsequence of $\{x_i\}$. Stated otherwise, the relation 'is a subsequence of' on the set $X^{\mathbb{N}}$ of all infinite sequences in X is reflexive and transitive; it is, therefore, a preorder (Section 0.22). Trivially, the relation 'is a subsequence of' is not symmetric, so it is not an equivalence relation; that it is not antisymmetric (whence not a partial order) may be seen in the following example. Let

(22) $$\{x_i\} = \{0,1,0,1,0,1,\ldots\} \quad \text{and} \quad \{y_i\} = \{1,0,1,0,1,0,\ldots\}.$$

The mapping $n : k \mapsto k+1$, that is, the sequence

(23) $$\{n_k\} = \{2,3,4,5,6,7\ldots\},$$

is such that

(24) $$y \circ n = x \quad \text{and} \quad x \circ n = y,$$

that is,

(25) $$\{y_{n_k}\} = \{x_i\} = \{0,1,0,1,0,1,\ldots\} \text{ and } \{x_{n_k}\} = \{y_i\} = \{1,0,1,0,1,0,\ldots\}.$$

So $\{x_i\}$ and $\{y_i\}$ are subsequences of each other, but $\{x_i\} \neq \{y_i\}$.

0.25 Enumerating Sequence Recall (Definition 0.9) that a nonempty set is finite if it is equipotent to the set $\{0,1,2,...,n-1\} \cong \{1,2,...,n\}$ for a natural number n, and a set is countably infinite if it is equipotent to the set $\mathbb{N}$ of all natural numbers. Equipotence implies that a nonempty finite set X with cardinality $|X| = n$ has a *bijective indexed family* $x:\{1,2,...,n\} \to X$ listing its elements in order and representing it as a finite sequence $\{x_i\}_{i\in\{1,2,...,n\}} = \{x_1, x_2, ..., x_n\}$. (This means, in particular, that for $i, j \in I$, if $i \neq j$, then $x_i \neq x_j$.) Similarly, a countably infinite set X (with cardinality $|X| = \aleph_0$) may be represented as an infinite sequence $\{x_i\}_{i\in\mathbb{N}} = \{x_1, x_2, x_3, ...\}$ with its corresponding *bijective* indexed family $x:\mathbb{N} \to X$. The bijective indexed family x, a mapping turning a countable set into a sequence, is called an *enumeration* of X (*cf.* Section 0.10).

For both finite and infinite sets, the choice of the enumeration is, as previously mentioned, not unique: any permutation of the assignment also serves as an enumeration (each different permutation defining its own distinct listing of elements and sequential representation of the set).

PART I
Pentateuchus
Becoming Mapping

He had brought a large map representing the sea,
    Without the least vestige of land:
And the crew were much pleased when they found it to be
    A map they could all understand.

“What’s the good of Mercator’s North Poles and Equators,
    Tropics, Zones, and Meridian Lines?”
So the Bellman would cry: and the crew would reply
    “They are merely conventional signs!

Other maps are such shapes, with their islands and capes!
    But we’ve got our brave Captain to thank”
(So the crew would protest) “that he’s bought *us* the best —
    A perfect and absolute blank!”

— Lewis Carroll (1876)
The Hunting of the Snark
Fit the Second (The Bellman’s Speech)

This introductory Part I is an exploration in five chapters of the algebraic theory of set-valued mappings.

My emphasis is on the topics that will be of use to us on our continuing journey in relational biology. Some theorems will only be stated in this introduction without proofs; their proofs may be found in Chapter 1 of Aubin and Frankowska [1990], Chapter 2 of Berge [1963], or Chapter 1 of Burachik and Iusem [2008]. These are among the very few books that contain the subject of set-valued mappings, and even therein, the algebraic theory is only a prelude that is quickly passed over to concentrate on the analytic and topological aspects. I should note that the 'forked arrow' notation $F : X \multimap Y$, to be introduced in Section 2.1, for a set-valued mapping F from set X to set Y, is my own.

1
Mapping Origins

> He drove out the man; and at the east of the garden of Eden he placed the cherubim, and a sword flaming and turning to guard the way to the tree of life.
>
> — *Genesis* 3:24

In Principio: Mappings

1.1 Definition Given two sets X and Y, one denotes by $X \times Y$ the set of all *ordered pairs* of the form (x, y) where $x \in X$ and $y \in Y$. The set $X \times Y$ is called the *product* (or *Cartesian product*) of the sets X and Y.

The definition of product may be extended to any finite sequence of sets (*cf.* Sections 0.18 and 0.21) $\{X_i\}_{i \in \{1,2,\ldots,n\}} = \{X_1, X_2, \ldots, X_n\}$, of which the product is the set of all *ordered n-tuples* of the form $(x_1, x_2, \ldots, x_n)$ where, for $i = 1, 2, \ldots, n$, $x_i \in X_i$, and may alternatively be denoted

(1) $$X_1 \times X_2 \times \cdots \times X_n = \prod_{i=1}^{n} X_i = \prod_{i \in (1,2,\ldots,n)} X_i$$

(the Cartesian product being the product in the category **Set**; *ML*: A.26).

If either X or Y is empty, then $X \times Y$ is empty. If $X \neq \varnothing$ and $Y \neq \varnothing$, then there is an element $x \in X$ and an element $y \in Y$, whence $(x, y) \in X \times Y$ and $X \times Y \neq \varnothing$. These remarks may trivially be extended to a finite sequence of sets $\{X_i\}_{i \in \{1,2,\ldots,n\}} = \{X_1, X_2, \ldots, X_n\}$ — $X_1 \times X_2 \times \cdots \times X_n = \varnothing$ if and only if at least one $X_i = \varnothing$. For an *infinite* indexed family of sets (i.e. $\{X_i\}_{i \in I}$ where the index set I is infinite), the sufficiency of the previous statement is still trivial: if at least

A.H. Louie, *The Reflection of Life*, IFSR International Series on Systems Science and Engineering 29, DOI 10.1007/978-1-4614-6928-5_1,

one $X_i = \varnothing$, then $\prod_{i \in I} X_i = \varnothing$. The necessity, however, is nontrivial, and the inverse statement is in fact an alternate statement of the

1.2 Axiom of Choice The product of a nonempty family of nonempty sets is nonempty.

(For a review of necessity versus sufficiency and the logic of conditional statements in general, see the Prolegomenon of *ML*.) Stated otherwise, if $\{X_i\}_{i \in I}$ is a family of nonempty sets indexed by a nonempty set I, then there exists an indexed family $\{x_i\}_{i \in I}$ such that for each $i \in I$, $x_i \in X_i$ (which is the Axiom of Choice stated in 0.20). The 'ordered I-tuple' $(x_i)_{i \in I}$ is an element in the product $\prod_{i \in I} X_i$, whence $\prod_{i \in I} X_i \neq \varnothing$.

1.3 Definition A *relation* is a set R of ordered pairs; that is, $R \subset X \times Y$ for some sets X and Y.

If $(x, y) \in R$, then one may say that x *is related to* y.

Equivalently, a relation R is an element of the power set $\mathrm{P}(X \times Y)$ (Definition 0.2), that is, $R \in \mathrm{P}(X \times Y)$. The collection of *all* relations between two sets X and Y is thus the power set $\mathrm{P}(X \times Y)$. The relation $U = X \times Y \in \mathrm{P}(X \times Y)$ is the *universal relation*, in which every $x \in X$ is related to every $y \in Y$. The relation $\varnothing \in \mathrm{P}(X \times Y)$ is the *empty relation*, in which no $x \in X$ is related to any $y \in Y$. In the partially ordered set $\langle \mathrm{P}(X \times Y), \subset \rangle$, U is the greatest element and $\varnothing$ is the least element (*cf. ML*: 1.28). For all relations $R \in \mathrm{P}(X \times Y)$, $\varnothing \subset R \subset U$.

1.4 Definition A *mapping* is a set f of ordered pairs with the property that, if $(x, y) \in f$ and $(x, z) \in f$, then $y = z$.

Note that the requirement for a subset of $X \times Y$ to qualify as a mapping is in fact quite a stringent one: an element $x \in X$ cannot be related to more than one element of $y \in Y$. Most relations, that is, *generic* members of $\mathrm{P}(X \times Y)$, do not have this 'single-valued' property.

1.5 Definition Let f be a mapping. One defines two sets, the *domain* of f and the *range* of f, respectively, by

(2) $$\operatorname{dom}(f) = \{x \in X : (x,y) \in f \text{ for some } y \in Y\}$$

and

(3) $$\operatorname{ran}(f) = \{y \in Y : (x,y) \in f \text{ for some } x \in X\}.$$

Thus $\operatorname{dom}(f) \subset X$ and $\operatorname{ran}(f) \subset Y$, and f is a subset of the product $\operatorname{dom}(f) \times \operatorname{ran}(f)$. If $\operatorname{ran}(f)$ contains exactly one element, then f is called a *constant mapping*.

Various words, such as ‘function’, ‘transformation’, and ‘operator’, are used as synonyms for ‘mapping’. The mathematical convention is that these different synonyms are used to denote mappings having special types of sets as domains or ranges. Because these alternate names also have interpretations in biological terms, to avoid semantic equivocation, in this book I shall—unless convention dictates otherwise—use *mapping* (and often *map*) for the mathematical entity.

1.6 Notations The traditional concept of a mapping is that which assigns to each element of a given set a definite element of another given set; that is, a ‘point-to-point’ map. I shall now reconcile this with the formal definition given above. Let X and Y be sets, and let $f \subset X \times Y$ be a mapping. This implies $f \subset \operatorname{dom}(f) \times \operatorname{ran}(f) \subset X \times Y$. If one further requires that $\operatorname{dom}(f) = X$ (I shall have more to say about this restriction in Section 1.24 below.), then one says that f is a *mapping of* X *into* Y, denoted by

(4) $$f : X \to Y$$

and occasionally (mostly for typographical reasons) by

(5) $$X \xrightarrow{f} Y.$$

The collection of all mappings of X into Y is a proper subset of the power set $\mathrm{P}(X \times Y)$; this subset is denoted Y^X. Suggestively, one has $Y^X \subset 2^{X \times Y}$.

To each element $x \in X$, by Definition 1.4, there corresponds a unique element $y \in Y$ such that $(x,y) \in f$. Traditionally, y is called the *value of the mapping* f *at the element* x, and the relation between x and y is denoted by $y = f(x)$ instead of $(x,y) \in f$. Note that the $y = f(x)$ notation is only logically consistent when f is a mapping (i.e. single-valued). For a general relation f, it is possible that $y \neq z$ yet both $(x,y) \in f$ and $(x,z) \in f$; if one were to write $y = f(x)$ and $z = f(x)$ in such a situation, then one would be led to the conclusion that $y = z$: a direct contradiction to $y \neq z$.

With the $y = f(x)$ notation, one has

(6) $$\operatorname{ran}(f) = \{ y \in Y : y = f(x) \text{ for some } x \in X \},$$

which may be further abbreviated to

(7) $$\operatorname{ran}(f) = \{ f(x) : x \in \operatorname{dom}(f) \}.$$

One then also has

(8) $$f = \{ (x, f(x)) : x \in X \}.$$

From this last representation, we observe that when $X \subset \mathbb{R}$ and $Y \subset \mathbb{R}$ (where $\mathbb{R}$ is the set of real numbers), my formal definition of a mapping coincides with that of the 'graph of f ' in elementary mathematics.

1.7 Element Chase Sometimes it is useful to trace the path of an element as it is mapped. If $a \in X$, $b \in Y$, and $b = f(a)$, one uses the 'maps to' arrow (note the short vertical line segment at the tail of the arrow) and writes

(9) $$f : a \mapsto b .$$

One occasionally also uses the 'maps to' arrow to define the mapping f itself:

(10) $$x \mapsto f(x).$$

Mappings of Sets

1.8 Definition Let f be a mapping of X into Y . If $E \subset X$, the *image* of E under f is defined to be the set $f(E)$ of all elements $f(x) \in Y$ for $x \in E$; that is,

(11) $$f(E) = \{ f(x) : x \in E \} \subset Y .$$

In this notation, $f(X)$ is the range of f . One may also note that, for all $x \in X$,

(12) $$f(\{x\}) = \{f(x)\}.$$

1.9 Definition If f is a mapping of X into Y, the set Y is called the *codomain* of f, denoted by $\mathrm{cod}(f)$.

While the domain and range of $f: X \to Y$ are specified by f in $\mathrm{dom}(f) = X$ and $\mathrm{ran}(f) = f(X)$, the codomain is not yet uniquely determined. All that is required so far is that the codomain contains the range as a subset, $Y \supset \mathrm{ran}(f)$. One needs to invoke a category theory axiom that assigns to each mapping f a unique set $Y = \mathrm{cod}(f)$ as its codomain. The axiom is on the mutual exclusiveness of hom-sets in a category **C**:

(13) $$\mathbf{C}(A,B) \cap \mathbf{C}(C,D) = \varnothing \text{ unless } A = C \text{ and } B = D.$$

Thus each **C**-morphism f determines a unique pair of **C**-objects, its domain $A = \mathrm{dom}(f)$ and codomain $B = \mathrm{cod}(f)$, such that $f \in \mathbf{C}(A,B)$. One may consider that associated with a category **C** there is a pair of 'mappings' (hence with unique images), dom and cod, that takes **C**-morphisms to **C**-objects. (I shall elaborate on this pair of mappings in Sections 6.8 and 6.9.) Alternatively, one may consider a **C**-morphism as a *triple* (A, B, f) consisting of two **C**-objects A, B and a **C**-morphism $f \in \mathbf{C}(A,B)$; equality between triples occurs when they are component-wise equal.

1.10 The Category Set The category in which the collection of objects is the collection of all sets (in a suitably naive universe of small sets) and the morphisms are mappings is denoted **Set**. Given two sets X and Y, the hom-set $\mathbf{Set}(X,Y)$ is the collection Y^X of all mappings from X to Y.

I often employ non-full subcategories of **Set**, and I use $H(X,Y)$ for appropriate subsets of $Y^X = \mathbf{Set}(X,Y)$ under consideration (e.g. when mappings $f: X \to Y$ represent metabolic functions). These collections of hom-sets $H(X,Y)$ in the subcategory, of course, still have to satisfy the category axioms.

Axiom (13) is interpreted in the category **Set** to yield unique codomains. If a given mapping f from A to B in fact maps A into a proper subset B' of B, then (A, B, f) and (A, B', f) count as different **Set**-morphisms, although as 'mappings' they are the same. For an illustration, consider the mapping $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ versus the mapping $g: \mathbb{R} \to \{y \in \mathbb{R} : y \geq 0\}$

defined by $g(x)=x^2$. While f and g are the same 'squaring mapping', they are different as **Set**-morphisms, $(\mathbb{R},\mathbb{R},f) \neq (\mathbb{R},\{y\in\mathbb{R}: y\geq 0\},g)$.

1.11 Surjection The range $f(X)=\operatorname{ran}(f)$ is a subset of the codomain $Y=\operatorname{cod}(f)$, but they need not be equal. When they are, that is, when $f(X)=Y$, one says that f is a *mapping of* X *onto* Y and that $f:X\to Y$ is *surjective* (or is a *surjection*). Note that every mapping maps onto its range.

1.12 Definition If $E\subset Y$, $f^{-1}(E)$ denotes the set of all $x\in X$ that f maps into E:

$$f^{-1}(E)=\{x: f(x)\in E\}\subset X \tag{14}$$

and is called the *inverse image* of E under f.

Note that $f^{-1}(Y)=X$, even though $\operatorname{ran}(f)=f(X)$ may be a proper subset of Y. If $y\in Y$, $f^{-1}(\{y\})$ is abbreviated to $f^{-1}(y)$, whence

$$f^{-1}(y)=\{x\in X: f(x)=y\}. \tag{15}$$

1.13 Injection Note that $f^{-1}(y)$ may be the empty set or may contain more than one element. If, for each $y\in Y$, $f^{-1}(y)$ consists of at most one element of X, then f is said to be an *injective* (also *one-to-one* or *1-1*) *mapping* of X into Y. Other commonly used names are '$f:X\to Y$ is an *injection*' and '$f:X\to Y$ is an *embedding*'. This may also be expressed as follows: f is a one-to-one mapping of X into Y provided $f(x_1)\neq f(x_2)$ whenever $x_1,x_2\in X$ and $x_1\neq x_2$.

For $A\subset X$, the embedding $i:A\to X$ defined by $i(x)=x$ for all $x\in A$ is called the *inclusion map* (of A in X). The inclusion map of X in X is called the *identity map* on X, denoted 1_X.

1.14 Lemma

i. *$f:X\to Y$ is injective iff for every $y\in\operatorname{ran}(f)$, $f^{-1}(y)$ is a singleton set in X.*
ii. *$f:X\to Y$ is surjective iff for every nonempty subset $E\subset Y$, $f^{-1}(E)$ is a nonempty subset of X.*

1.15 Inverse Mapping In view of the equivalence in Lemma 1.14.i, when $f : X \to Y$ is injective, it defines an *inverse mapping* $f^{-1} : \mathrm{ran}(f) \to X$ (with the mild notational equivocation of each singleton set $f^{-1}(y)$ with the element it contains). Indeed, as a mapping, f^{-1} is necessarily a *one-to-one* mapping of $\mathrm{ran}(f)$ *onto* $X = \mathrm{dom}(f)$.

A mapping and its inverse (when it exists) compose to identity mappings; thus,

(16) $$f^{-1} \circ f = 1_X \quad \text{but} \quad f \circ f^{-1} = 1_{f(X)}$$
(and not necessarily $f \circ f^{-1} = 1_Y$).

One also has the following simple

1.16 Lemma *Let $f : X \to Y$ and $g : Y \to X$ be mappings. If $g \circ f = 1_X$ then f is injective and g is surjective.*

1.17 Bijection If a mapping $f : X \to Y$ is both one-to-one and onto, that is, both injective and surjective, then f is called *bijective* (or is a *bijection*) and that the mapping f establishes a *one-to-one correspondence* between the sets X and Y.

1.18 The Power Set Functor The *power set functor* $\mathrm{P} : \mathbf{Set} \to \mathbf{Set}$ assigns to each set X its power set $\mathrm{P}X$ (Definition 0.2) and assigns to each mapping $f : X \to Y$ the mapping

(17) $$\mathrm{P}f : \mathrm{P}X \to \mathrm{P}Y$$

that sends each $A \subset X$ to its image $f(A) \subset Y$. One readily verifies that this definition satisfies the functorial requirements $\mathrm{P}(g \circ f) = \mathrm{P}(g) \circ \mathrm{P}(f)$ (the mapping that sends a subset A of the domain of f to the subset $g(f(A))$ of the codomain of g) and $\mathrm{P}1_X = 1_{\mathrm{P}X}$ (the identity morphism gets sent to the identity morphism), so P is a covariant functor from **Set** to **Set**.

Dually, the *contravariant power set functor* $\overline{\mathrm{P}} : \mathbf{Set} \to \mathbf{Set}$ assigns to each set X its power set $\mathrm{P}X$ and to each mapping $f : X \to Y$ the mapping

(18) $$\overline{\mathrm{P}}f : \mathrm{P}Y \to \mathrm{P}X$$

that sends each $B \subset Y$ to its inverse image $f^{-1}(B) \subset X$.

Thus a 'point-to-point' mapping $f : X \to Y$ naturally defines two 'point-to-point' mappings, $\mathrm{P}f : \mathrm{P}X \to \mathrm{P}Y$ and $\overline{\mathrm{P}}f : \mathrm{P}Y \to \mathrm{P}X$, for which a 'point' is an element of a power set, hence a set. Alternatively (with mild notational equivocation), the 'image map' $\mathrm{P}f$ may be considered a 'set-to-set' mapping f from X to Y, sending subsets of X to subsets of Y, while the 'inverse image map' $\overline{\mathrm{P}}f$ may be considered a 'set-to-set' mapping f^{-1} from Y to X, sending subsets of Y to subsets of X.

Note that the traffic $f \mapsto \mathrm{P}f$ (or $f \mapsto \overline{\mathrm{P}}f$) from a 'point-to-point' mapping to a 'set-to-set' mapping only goes one way. For a given mapping $g : \mathrm{P}X \to \mathrm{P}Y$ of power sets, there is in general no mapping $f : X \to Y$ for which $g = \mathrm{P}f$ (or $f : Y \to X$ for which $g = \overline{\mathrm{P}}f$). This is because, in the covariant case (the argument for the contravariant case being similar), for $x \in X$, one would have to have $f(x) = g(\{x\})$. But there is no guarantee that $g(\{x\}) \subset Y$ is a singleton set, which is what is required for f to be a mapping.

Some properties of the 'set-to-set' mapping $\mathrm{P}f : \mathrm{P}X \to \mathrm{P}Y$ are listed in the following theorem:

1.19 Theorem *Let $f : X \to Y$ and $A, B \subset X$. Then:*

i. $f(A) \neq \varnothing$ *iff* $A \neq \varnothing$.
ii. $A \subset B \Rightarrow f(A) \subset f(B)$.
iii. $f(A \cup B) = f(A) \cup f(B)$.
iv. $f(A \cap B) \subset f(A) \cap f(B)$.
v. $f(B \sim A) \supset f(B) \sim f(A)$.

The simple example of $f(1) = f(2) = 0$, $A = \{1\}$, and $B = \{2\}$ (whence $f(A) = f(B) = \{0\}$) shows why the converse of property ii is false and why properties iv and v are not equalities.

One sees from property iv that the mapping $\mathrm{P}f : \mathrm{P}X \to \mathrm{P}Y$ does not preserve the Boolean algebraic structure of power sets: one does not have $f(A \cap B) = f(A) \cap f(B)$. Note property v implies that

(19) $$f(X \sim A) \supset f(X) \sim f(A) = \mathrm{ran}(f) \sim f(A),$$

but in general there is no inclusion relation between the sets $f(X \sim A)$ and $Y \sim f(A)$ (since $Y \sim f(X)$ may be nonempty; i.e. f may not be surjective).

On the other hand, the dual 'set-to-set' mapping $\overline{\mathrm{P}}f : \mathrm{P}Y \to \mathrm{P}X$ has the following properties:

1.20 Theorem *Let* $f : X \to Y$ *and* $A, B \subset Y$. *Then:*

i. $A = \varnothing \;\Rightarrow\; f^{-1}(A) = \varnothing$.
ii. $A \subset B \;\Rightarrow\; f^{-1}(A) \subset f^{-1}(B)$.
iii. $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.
iv. $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.
v. $f^{-1}(B \sim A) = f^{-1}(B) \sim f^{-1}(A)$.

Going from Theorem 1.19 to Theorem 1.20, one sees that now in properties iv and v, set inclusion has been replaced by equality. Recall (Section 1.15) that the inverse mapping f^{-1} is necessarily a *bijective* mapping of $\mathrm{ran}(f)$ onto $X = \mathrm{dom}(f)$. These 'improvements' result as a consequence. Theorem 1.20 says that $\overline{\mathrm{P}}f : \mathrm{P}Y \to \mathrm{P}X$ is in fact a Boolean algebra homomorphism. Note also that $f^{-1}(\varnothing) = \varnothing$ and $f^{-1}(Y) = X$; thus the least and greatest elements are preserved by the inverse mapping f^{-1}.

The converse of property 1.20.i is not true (in contrast to property 1.19.i). This is again because f itself may not be surjective: if $A \subset Y \sim f(X)$ is a nonempty subset, then one still has $f^{-1}(A) = \varnothing$. But $f^{-1} : \mathrm{ran}(f) \to X$ is surjective, so $f^{-1}(Y) = X$; property 1.20.v then also says that

$$f^{-1}(Y \sim A) = X \sim f^{-1}(A) \tag{20}$$

(contrast this with (19) and its subsequent discussion).

The following properties of the composites of f and f^{-1} may also be readily verified:

1.21 Theorem *Let* $f : X \to Y$, $A \subset X$, *and* $B \subset Y$. *Then:*

i. $A \subset f^{-1}(f(A))$.
ii. $B \supset f(f^{-1}(B))$.
iii. $f(A \cap f^{-1}(B)) = f(A) \cap B$.

The inclusion relations in properties 1.19.iv and v and properties 1.21.i and ii become equality under special conditions:

1.22 Theorem *Let $f : X \to Y$. The following are equivalent:*

i. *f is injective.*

ii. *For all $A \subset X$, $A = f^{-1}\big(f(A)\big)$.*

iii. *For all $A, B \subset X$, $f(A \cap B) = f(A) \cap f(B)$.*

iv. *For all $A, B \subset X$, $f(B \sim A) = f(B) \sim f(A)$.*

v. *For all $A \subset X$, $f(X \sim A) \subset Y \sim f(A)$.*

1.23 Theorem *Let $f : X \to Y$. The following are equivalent:*

i. *f is surjective.*

ii. *For all $B \subset Y$, $B = f\big(f^{-1}(B)\big)$.*

iii. *For all $A \subset X$, $Y \sim f(A) \subset f(X \sim A)$.*

What Is a Mapping?

1.24 Hardy's Idea of a Mapping A mapping (i.e. 'function' in the mathematical sense) $y = f(x)$ often possesses three characteristics:

i. y is determined for *every value* of x.

ii. y is determined *uniquely* for each value of x; that is, to each value of x corresponds *one and only one* value of y.

iii. The relation f between x and y is expressed by means of an *analytic formula*, from which the value of y corresponding to a given x may be calculated by direct substitution of the latter.

G. H. Hardy, in his seminal textbook *A Course of Pure Mathematics* [10th edition, 1952], dismissed each of these characteristics as "by no means essential to a function". That characteristic iii is not essential is evident: not all functional correspondences are given by neat formulae such as $y = 3x^2 + x - 2$ and $y = a\sin(bx + c)$. Indeed, the existence of an "analytic formula" depends on the set of "elementary functions" one has in one's toolbox, which is expanded by necessity with the augmentation, when circumstances warrant, of "special functions": consider, for example, $y = \log(x)$, $y = \cosh(x)$, $y = \Gamma(x)$, and $y = Li(x)$. But toolbox collections are finite. From another viewpoint, the negation of iii alludes to the fact that not all mappings are computable.

In all the editions of his book (from first edition (1908) to the final tenth edition (1952), which is still being reprinted and available), Hardy maintained:

> "All that is essential is that there should be some relation between x and y such that to some values of x at any rate correspond values of y."

Note the quantifier in "*some* values of x" and the plurality of "correspond *values* of y".

That characteristic i is not essential is inherent in the definition of a relation R as any subset of the Cartesian product $X \times Y$. There is no requirement that for each $x \in X$, there has to be a $y \in Y$ such that $(x, y) \in R$. This condition may be passed on to mappings, hence the negation of i: $y = f(x)$ (i.e. $(x, y) \in f$) may only hold for "*some* values". For example, one may consider $f(x) = \sqrt{x}$ as a mapping from ℝ to ℝ, although $f(x)$ is not determined for $x < 0$, whence $\mathrm{dom}(f) = \{x \in \mathbb{R} : x \geq 0\}$. Since $\mathrm{dom}(f) \subset X$, the issue may be bypassed by restricting f to $X' = \mathrm{dom}(f)$, and considering $f \subset X' \times Y$ instead of $f \subset X \times Y$, then y is determined for *every value* of $x \in X'$. This is commonly practised; with the $f(x) = \sqrt{x}$ example, the mapping is more properly considered as from $X' = \{x \in \mathbb{R} : x \geq 0\}$ to ℝ. Indeed, in the notation $f : X \to Y$ (Section 1.6), the convention is that one implicitly takes $\mathrm{dom}(f) = X$ (unless otherwise stated).

The single-valued requirement of characteristic ii is now standard, universally accepted as an integral part of the definition of a mapping (*cf.* Definition 1.4 above). As I remarked in Section 1.6 above, the notation $y = f(x)$ (attributed to Leonhard Euler) only makes logical sense when $f(x) \in Y$ is uniquely determined. In this context of a mapping being single-valued by definition, the term 'multi-valued mapping' is therefore a misnomer; a mapping has to be single valued to be called 'well defined'. But as Hardy declared as late as 1952, it is at times useful to relax characteristic ii to include "*values* of y".

1.25 Well-Posed Problem Jacques Hadamard stated that mathematical models of physical phenomena should have the properties that:

i. A solution exists.
ii. The solution is unique.
iii. The solution depends continuously on the data, in some reasonable topology.

The formulation of such a model is termed a *well-posed problem* (and an *ill-posed problem* otherwise). Hadamard's well-posed problem is often used as an explanation of why mappings are defined thus, especially their unique-value requirement: compare Hadamard's three properties with the three characteristics in the previous section.

1.26 Examples of Multi-valued Mappings There are, however, many situations in which existence fails, when no output is associated with an input, and in which uniqueness fails, when more than one output are associated with an input. When a mapping is not surjective, its inverse is not defined on its codomain: for $f: X \to Y$ and $y \in Y \sim \operatorname{ran}(f)$, $f^{-1}(y)$ is not defined (at least not by the role of f^{-1} as an inversion of f). When a mapping is not injective, its inverse is not single-valued: for $f: X \to Y$ and $y \in \operatorname{ran}(f)$, $f^{-1}(y)$ may contain more than one element. Thus the 'inverse' f^{-1} of a mapping f is not necessarily itself a mapping. Stated otherwise, the 'inverse' of a **Set**-morphism is not necessarily a **Set**-morphism.

As a simple example, consider the inverse of the square mapping $y = x^2$ from $\mathbb{R}$ to $\mathbb{R}$, that is, real solutions to the equation $y^2 = x$. If x is a negative real number, there are no real solutions for y. (This is, of course, famously the genesis of complex numbers.) If x is a positive real number, this equation defines two values of y corresponding to x, namely, $y = \pm\sqrt{x}$. Indeed, the 'square-root mapping' is not a mapping, unless one follows the convention that the symbol $\sqrt{x}$ is defined to mean the *positive square root* of a positive real number x (whence $-\sqrt{x}$ is the negative square root). The proper 'square-root mapping' is thus the 'double-valued'

$$x \mapsto \left\{ \sqrt{x}, -\sqrt{x} \right\}. \tag{21}$$

In general, for a complex number $z = re^{i\theta}$ (where $-\pi < \theta \le \pi$), there are n (distinct when $z \neq 0$) n th roots of z, given by

$$r^{\frac{1}{n}} e^{i\frac{1}{n}(\theta + 2k\pi)}, \qquad k = 0, 1, \ldots, n-1. \tag{22}$$

The proper 'n th-root mapping' from $\mathbb{C}$ to $\mathbb{C}$ is thus the multi-valued

$$z \mapsto \left\{ r^{\frac{1}{n}} e^{i\frac{1}{n}\theta}, r^{\frac{1}{n}} e^{i\frac{1}{n}(\theta + 2\pi)}, \ldots, r^{\frac{1}{n}} e^{i\frac{1}{n}(\theta + 2(n-1)\pi)} \right\}. \tag{23}$$

In complex analysis, there are many situations in which 'multi-valued mappings' arise, and most of them stem from the *ambiguity* of the argument $\theta = \arg z$ of a complex number

$$z = re^{i\theta}, \tag{24}$$

since θ plus any multiple of 2π may be substituted for θ in (24). The 'mapping' arg is thus not well defined:

(25) $$\arg : z \mapsto \{\theta + 2k\pi : k \in \mathbb{Z}\}.$$

In order to have a (single-valued) mapping, one restricts $\arg : \mathbb{C} \to \mathbb{R}$ to one branch (its *principal branch*), and one defines

(26) $$-\pi < \arg z \leq \pi ,$$

that is, as the mapping $\arg : \mathbb{C} \to (-\pi, \pi]$.

It is often useful to have domains and codomains of complex mappings as open sets; thus one may even restrict further. For example, the complex logarithm of a complex number $z = re^{i\theta}$ is the multi-valued mapping

(27) $$\log : z \mapsto \{ \log r + i(\theta + 2k\pi) : k \in \mathbb{Z} \}.$$

Its principal branch is restricted to the domain $\mathbb{C} \sim \{s : s \leq 0\}$ (i.e. the complex plane with a 'slit' along the negative real axis) and defined as

(28) $$\log z = \log|z| + i \arg z$$

(where $-\pi < \arg z < \pi$), whence $\log : \mathbb{C} \sim \{s : s \leq 0\} \to \{w \in \mathbb{C} : \mathrm{Re}\, w > 0\}$.

A more sophisticated treatment replaces complex 'multi-valued mappings' with mappings with Riemann surfaces as domains, but I shall not digress thence.

2
From Points to Sets

> He made loops of blue on the edge of the outermost curtain of the first set; likewise he made them on the edge of the outermost curtain of the second set; he made fifty loops on the one curtain, and he made fifty loops on the edge of the curtain that was in the second set; the loops were opposite one another.
>
> — *Exodus* 36:11–12

Congregatio: Set-Valued Analysis

2.1 Set-Valued Mapping From the forms of the 'point-to-set mappings' $F : \bullet \mapsto \{\cdots\}$ in Section 1.26 (*cf.* (21), (23), (25), and (27) therein), one may naturally proceed to define a set-valued mapping thus:

Definition A A *set-valued mapping* from set X to set Y is a relation $F \subset X \times Y$ (Definition 1.3). It may be denoted

(1) $$F : X \multimap Y,$$

such that for each $x \in X$,

(2) $$F(x) = \{ y \in Y : (x, y) \in F \} \subset Y.$$

Note the *point-to-set* nature of a set-valued mapping (as opposed to 'point-to-point' for a standard mapping; *cf.* Section 1.6). This relaxation of characteristic 1.24.ii thus includes, when $F(x)$ contains more than one element, Hardy's allowance of mappings in which to a point may plurally "correspond *values* of y". Note, also, the possibility that for some $x \in X$, it may happen that $F(x) = \varnothing$. This relaxation of characteristic 1.24.i thus includes Hardy's allowance of mappings in which values may correspond to only "*some* values of x".

A.H. Louie, *The Reflection of Life*, IFSR International Series on Systems Science and Engineering 29, DOI 10.1007/978-1-4614-6928-5_2,

Note the special ‘forked arrow’ $-\!\sqsubset$ that I have chosen to denote set-valued mappings, in distinction from $\rightarrow$ for a standard (single-valued) mapping. In this chapter when I introduce the concept of set-valued mapping and its properties, I shall also use capital letters to denote set-valued mappings, e.g., $F : X -\!\sqsubset Y$, while use lowercase letters to denote standard mappings, e.g., $f : X \rightarrow Y$. This F–versus–f distinction may not, however, necessarily continue in later chapters, but the two different arrows will remain as the characterizing form.

In a set-valued mapping’s element-chasing form, one may write

(3) $$F : x \mapsto F(x).$$

The ‘source’ of F is still a *point* $x \in X$, but now the *value* of the mapping F at the element x is a *set* $F(x) \subset Y$. The source (material cause) and the value (final cause) of a set-valued mapping are thus different in kind from each other, they belonging to different hierarchical levels (‘point’ versus ‘set’). (For a review of the identification of Aristotle’s four causes with components of a mapping, see *ML*: Chapter 5.)

A standard (single-valued) mapping (as defined in 1.4) $f : X \rightarrow Y$ may be considered a very specialized set-valued mapping $F : X -\!\sqsubset Y$ such that, for each $x \in X$, the value

(4) $$F(x) = \{f(x)\}$$

is a singleton set. Indeed, one can make the formal definition: a set-valued mapping $F : X -\!\sqsubset Y$ is called *single-valued* if for each $x \in X$, $F(x)$ is a singleton set. A ‘single-valued set-valued mapping’ $F : X -\!\sqsubset Y$ therefore defines a ‘standard’ mapping $f : X \rightarrow Y$ by $f : x \mapsto$ the single element in $F(x)$. Thus ‘single-valued set-valued mapping’ and ‘mapping’ are equivalent terms.

Since a set-valued mapping $F : X -\!\sqsubset Y$ takes its values in the family of subsets of Y (i.e., the power set $\mathrm{P}Y$ of Y), one may alternatively consider

Definition B A *set-valued mapping* from set X to set Y is a (single-valued) mapping $F : X \rightarrow \mathrm{P}Y$.

In algebraic terms, the two definitions are equivalent. In topological terms (*cf.* Hadamard’s property iii in 1.25), however, because of the complicated power-set topology of $\mathrm{P}Y$ induced by the topology of Y, it is often advantageous to use Definition A.

2.2 Definition Let $F : X -\!\sqsubset Y$ be a set-valued mapping. The *graph* of F is defined as F in its relational form; i.e.,

(5) $$F = \{(x,y) \in X \times Y : y \in F(x)\} = \{(x,y) \in X \times Y : (x,y) \in F\} \subset X \times Y .$$

(Compare this with the 'graph of f ' in Section 1.6.)

2.3 Domain The *domain* of the set-valued mapping $F : X \multimap Y$ is the set X, denoted by $\operatorname{dom}(F)$.

The word 'domain' is from the Latin *domus*, 'house, home'. Thus the domain of a mapping is the set of values for which the mapping 'feels at home' (in the idyllic and idealistic sense of the set of values that 'do not cause the mapping any trouble'). In addition, the related Latin word *dominus* means 'lord, master' literally 'one who rules the home', or 'one who owns the domain'. Thus the domain of a mapping is the set of values that the mapping 'owns' or 'has control of'.

There is a subtle difference in the definitions of 'domain' of a set-valued mapping and a (single-valued) mapping, as respectively given in 2.3 and 1.5. When a mapping is considered as a relation $f \subset X \times Y$, one has $\operatorname{dom}(f) \subset X$. But, as I mentioned in Section 1.24, in the notation $f : X \to Y$ for a standard mapping, the convention is that one implicitly takes $\operatorname{dom}(f) = X$ (whence for every $x \in X$, $f(x)$ is defined and it is a single element in Y). Contrariwise, for a set-valued mapping $F : X \multimap Y$, F is still defined at those $x \in X$ for which $F(x) = \varnothing$. One has $\operatorname{dom}(F) = X$ in both interpretations of F, as the relation $F \subset X \times Y$ and as the point-to-set mapping $F : x \mapsto F(x)$ from X to $\mathcal{P}Y$.

2.4 Definition The projections of the graph of F onto its first and second components are, respectively, the *corange* and the *range* of F,

(6) $$\operatorname{cor}(F) = \{x \in X : F(x) \neq \varnothing\},$$

(7) $$\operatorname{ran}(F) = \{y \in Y : y \in F(x) \text{ for some } x \in X\}.$$

Thus $\operatorname{cor}(F) \subset X$ and $\operatorname{ran}(F) \subset Y$, and both inclusions may be proper.

$X \sim \operatorname{cor}(F) = \operatorname{dom}(F) \sim \operatorname{cor}(F)$ is the subset of X that contains all those $x \in X$ at which $F(x) = \varnothing$. Note that some authors, however, define the domain of F as $\{x \in X : F(x) \neq \varnothing\}$ instead of X itself. But there are category-theoretic advantages in allowing $F(x) = \varnothing$ for $x \in \operatorname{dom}(F)$. (I shall return to this point when I presently introduce the category **Rel** of sets and relations.) The range of F may also be expressed as

(8) $$\operatorname{ran}(F) = \bigcup_{x \in X} F(x) \subset Y .$$

F (as a relation in $X \times Y$) is thus a subset of the product $\operatorname{cor}(F) \times \operatorname{ran}(F)$. $x \in \operatorname{cor}(F)$ means there exists $y \in \operatorname{ran}(F)$ such that $(x, y) \in F$; dually, $y \in \operatorname{ran}(F)$ means there exists $x \in \operatorname{cor}(F)$ such that $(x, y) \in F$.

If there exists a subset C of Y such that $F(x) = C$ for all $x \in X$, then F is called a *constant set-valued mapping*. As a relation in $X \times Y$, F is the subset $X \times C$. The constant mapping $f : x \mapsto c$ (where $c \in Y$) thus defines the constant set-valued mapping $F : x \mapsto \{c\}$. The universal relation $U = X \times Y$ from X to Y (*cf.* Section 1.3) is the constant set-valued mapping $U : X \multimap Y$ that sends everything to the set Y, i.e., such that $F(x) = Y$ for all $x \in X$.

2.5 The Constant Empty-Set-Valued Mapping The constant set-valued mapping $F : X \multimap Y$ that sends everything to the empty set, i.e., such that

(9) $$F(x) = \varnothing \quad \text{for all } x \in X ,$$

has

(10) $$\operatorname{cor}(F) = \{x \in X : F(x) \neq \varnothing\} = \varnothing ,$$

(11) $$\operatorname{ran}(F) = \varnothing ,$$

and

(12) $$\{x \in X : F(x) = \varnothing\} = X \sim \operatorname{cor}(F) = X .$$

As a relation in $X \times Y$, F is thus the 'empty relation' $\varnothing$ (*cf.* Section 1.3).

Note that the 'empty relation' $\varnothing$ is a legitimate set-valued mapping from set X to set Y, for all sets X and Y. This is in contrast to standard mappings, when the 'empty mapping' $\varnothing : X \to Y$ is only a mapping when $X = \varnothing$. Recall (*ML*: A.4) that by convention $Y^{\varnothing} = \{\varnothing\}$; thus the 'empty mapping' $\varnothing$ is the only mapping from the empty set to any set Y. If $X \neq \varnothing$, however, then $f(X) \neq \varnothing$ for any mapping f with $\operatorname{dom}(f) = X$, whence $\operatorname{ran}(f) \neq \varnothing$; so one has $\varnothing^X = \varnothing$ whence $\varnothing \notin \varnothing^X$.

It is interesting to note that for any two sets X and Y, whatever their nature, the constant empty-set-valued mapping $\varnothing : X \multimap Y$ is the same one. There is only one constant empty-set-valued mapping because there is only one empty set. Suppose $\varnothing_1$ and $\varnothing_2$ are two empty sets. Then $x \in \varnothing_1 \Rightarrow x \in \varnothing_2$, since there is no $x \in \varnothing_1$ to contradict this statement; thus $\varnothing_1 \subset \varnothing_2$. Likewise, $\varnothing_2 \subset \varnothing_1$. Therefore, $\varnothing_1 = \varnothing_2$.

The map that is a 'perfect and absolute blank' of Lewis Carroll's Bellman is an example of a constant empty-set-valued mapping (indeed, a manifestation of *the* empty set) $\varnothing$. As a material system, a blank sheet of paper is, of course, *structurally* nonempty, but, as a map, it *functions* as the empty set.

2.6 Definition For a set-valued mapping $F : X \multimap Y$, the set Y is called the *codomain* of F, denoted by $\operatorname{cod}(F)$.

Thus one has the dual relations

(13) $$\operatorname{ran}(F) \subset \operatorname{cod}(F) = Y, \qquad \operatorname{cor}(F) \subset \operatorname{dom}(F) = X.$$

2.7 Definition A set-valued mapping $F : X \multimap Y$ is:

i. *Surjective* if

(14) $$\operatorname{ran}(F) = \operatorname{cod}(F) = Y$$

ii. *Semi-single-valued* if

(15) $$F(x_1) \cap F(x_2) \neq \varnothing \quad \Rightarrow \quad F(x_1) = F(x_2)$$

iii. *Injective* if

(16) $$x_1 \neq x_2 \quad \Rightarrow \quad F(x_1) \cap F(x_2) = \varnothing$$

(which is contrapositively equivalent to

(17) $$F(x_1) \cap F(x_2) \neq \varnothing \quad \Rightarrow \quad x_1 = x_2\text{)}$$

A semi-single-valued mapping $F : X \multimap Y$ defines a *partition* of its range $\operatorname{ran}(F)$; its distinct values are pairwise disjoint subsets of Y, forming the *blocks* of the partition. It also defines a partition of its domain X: one block is $X \sim \operatorname{cor}(F)$ (which contains all those $x \in X$ for which $F(x) = \varnothing$), and then

2.12 Theorem *Let $F : X \multimap Y$ be a set-valued mapping. The following are equivalent:*

i. *F is surjective.*
ii. *For all $A \subset X$, $Y \sim F(A) \subset F(X \sim A)$.*

Inverse Mapping

2.13 Definition Given a set-valued mapping $F : X \multimap Y$, its *inverse* is the set-valued mapping $F^{-1} : Y \multimap X$ (equivalently, the relation $F^{-1} \subset Y \times X$) defined by interchanging the ordered components in the graph (5) of F:

$$F^{-1} = \{(y,x) \in Y \times X : y \in F(x)\} = \{(y,x) \in Y \times X : (x,y) \in F\} \subset Y \times X . \quad (20)$$

A (single-valued) mapping is not necessarily injective, and so its inverse is not necessarily single-valued and hence not (well defined as) a mapping. But the inverse of a set-valued mapping is always a set-valued mapping. Note, however, that F^{-1} is itself a point-to-set mapping (not a 'set-to-point mapping', as a direct reversal-of-roles 'inverse' of a point-to-set mapping would have been), with its value at the point $y \in Y$ defined as the set

$$F^{-1}(y) = \{x \in X : (x,y) \in F\} \subset X . \quad (21)$$

Indeed, since both $F(x)$ and $F^{-1}(y)$ are defined by the membership $(x,y) \in F$ (*cf.* (2) and (21)), one trivially has

2.14 Lemma *Let $F : X \multimap Y$, $x \in X$, and $y \in Y$. Then*

$$y \in F(x) \ \textit{iff} \ x \in F^{-1}(y). \quad (22)$$

While F maps points in X to subsets of Y, the inverse F^{-1} maps points in Y to subsets of X; so the involvements of the sets X and Y in F and F^{-1} are asymmetric. The situation is more evident if one considers the maps in terms of Definition 2.1B:

$$F : X \to \mathrm{P}Y , \qquad F^{-1} : Y \to \mathrm{P}X . \quad (23)$$

There is, however, symmetry in corange and range:

$$\mathrm{cor}(F) = \mathrm{ran}(F^{-1}) = F^{-1}(Y), \qquad F(X) = \mathrm{ran}(F) = \mathrm{cor}(F^{-1}) . \quad (24)$$

Note also that

(25) $$\operatorname{dom}(F) = \operatorname{cod}(F^{-1}) = X, \qquad Y = \operatorname{cod}(F) = \operatorname{dom}(F^{-1}).$$

And that

(26) $$(F^{-1})^{-1} = F.$$

For $F : X \mathrel{-\!\sqsubset} Y$, all the $x \in X$ for which $F(x) = \varnothing$ are not members of $\operatorname{cor}(F)$ and, therefore, not members of $\operatorname{ran}(F^{-1})$. In other words, when $X \sim \operatorname{cor}(F) \neq \varnothing$, F^{-1} is not surjective. If $y \in Y \sim \operatorname{ran}(F)$, then $F^{-1}(y) = \varnothing$. Consider the simple example of $F : \{1,2\} \mathrel{-\!\sqsubset} \{p,q\}$ with $F(1) = \{p,q\}$ and $F(2) = \{q\}$; then $F^{-1}(p) = \{1\}$ and $F^{-1}(q) = \{1,2\}$. This F^{-1} is not semi-single-valued and (hence) not injective. Thus, in contrast to an inverse mapping f^{-1} (which is only defined from $\operatorname{ran}(f)$ to X but is both injective and surjective thence, *cf.* Section 1.15), an inverse set-valued mapping F^{-1} is defined from Y to X, but is not necessarily either injective or surjective.

2.15 Theorem *Let $F : X \mathrel{-\!\sqsubset} Y$ and $F^{-1} : Y \mathrel{-\!\sqsubset} X$ be its inverse. Then:*

i. *If F is single-valued, F^{-1} is injective.*
ii. *If F is injective, F^{-1} is single-valued.*
iii. *If F is semi-single-valued, F^{-1} is semi-single-valued.*

Inverse Images

If $f : X \to Y$ is a mapping and $E \subset Y$, the inverse image of E under f, the set $f^{-1}(E) = \{x \in X : f(x) \in E\}$, may be considered in two equivalent ways:

i. As the set $\{x \in X : \{f(x)\} \cap E \neq \varnothing\}$
ii. As the set $\{x \in X : \{f(x)\} \subset E\}$

When these two sets are interpreted in set-valued mapping terms (recalling that f defines the special singleton-set-valued mapping $x \mapsto \{f(x)\}$), they give two different notions of the inverse image of a set $E \subset Y$:

2.16 Definition For a set-valued mapping $F : X \multimap\sqsubset Y$ and $E \subset Y$,

i. The *inverse image* of E by F is the set

(27) $$F^{-1}(E) = \begin{cases} \{x \in X : F(x) \cap E \neq \varnothing\} & \text{if } E \neq \varnothing \\ \varnothing & \text{if } E = \varnothing \end{cases}.$$

ii. The *core* of E by F is the set

(28) $$F^{+1}(E) = \{x \in X : F(x) \subset E\}.$$

The two notions i and ii coincide (and are identical to the inverse image in Definition 1.12) when the mapping is single-valued, since $F(x) \cap E \neq \varnothing$ iff $F(x) \subset E$ when $F(x)$ is a singleton set.

Note that when $F^{-1} : Y \multimap\sqsubset X$ is considered a set-valued mapping itself (as opposed to its role as the inverse of another set-valued mapping), for $E \subset Y$ the set $F^{-1}(E)$, the image of E under F^{-1}, has already been defined in 2.9. It is the set

(29) $$F^{-1}(E) = \bigcup_{y \in E} F^{-1}(y) \subset X .$$

One may verify that this defines the same set as in (27), so the notation is consistent. In particular, for $y \in Y$, $F(x) \cap \{y\} \neq \varnothing$ iff $y \in F(x)$ iff $(x, y) \in F$, thus $F^{-1}(\{y\})$ as defined by (27) when $E = \{y\}$ is identical to $F^{-1}(y)$ as defined in (21).

The similarity of the word 'core' to the symbol ' cor ' for corange may lead to confusion, so it is perhaps opportune to clarify here at the outset. For a set-valued mapping $F : X \multimap\sqsubset Y$ and $E \subset Y$, both the corange of F and the core of E by F are subsets of the domain X of F:

(30) $$\operatorname{cor}(F) \subset X \quad \text{and} \quad F^{+1}(E) \subset X .$$

But there are no general inclusion relations between $\operatorname{cor}(F)$ and $F^{+1}(E)$. Other than having the first three letters of their names in common, corange and core are very different entities: $\operatorname{cor}(\bullet)$, the corange of $\bullet$, accepts one argument F that is a set-valued mapping, whereas $\bullet^{+1}(\bullet)$, the core of $\bullet$ by $\bullet$, accepts two arguments, the first being a set-valued mapping F and the second being a subset E of the mapping's codomain.

The definition of $F^{+1}(E)$ implies that

(31) $$F^{+1}(\varnothing) = \{x \in X : F(x) = \varnothing\} = X \sim \operatorname{cor}(F) = \operatorname{dom}(F) \sim \operatorname{cor}(F);$$

i.e., $F^{+1}(\varnothing)$ is the subset of X that contains all those $x \in X$ at which $F(x) = \varnothing$, and it is not necessarily the empty set. Equivalently, (31) says

(32) $$\operatorname{cor}(F) = \{x \in X : F(x) \neq \varnothing\} = X \sim F^{+1}(\varnothing) = \operatorname{dom}(F) \sim F^{+1}(\varnothing).$$

Note that for every $E \subset Y$,

(33) $$F^{+1}(\varnothing) \subset F^{+1}(E),$$

and

(34) $$F^{-1}(E) \subset X \sim F^{+1}(\varnothing).$$

This last inclusion says that $F^{-1}(E) \cap F^{+1}(\varnothing) = \varnothing$, which means if $x \in F^{-1}(E)$, then $F(x) \neq \varnothing$.

Consider the simple example of $F : \{1,2\} \multimap \{p,q\}$ with $F(1) = \{p,q\}$ and $F(2) = \varnothing$; then $\operatorname{cor}(F) = \{1\}$, $F^{-1}(\{p\}) = \{1\}$, and $F^{+1}(\{p\}) = \{2\}$. This shows that in general there are no inclusion relations between $\operatorname{cor}(F)$ and $F^{+1}(E)$ and between $F^{-1}(E)$ and $F^{+1}(E)$.

The same authors who define the domain of F as $\{x \in X : F(x) \neq \varnothing\}$ (i.e., my $\operatorname{cor}(F)$) also define their alternate inverse (sometimes called *upper inverse*) accordingly, for $E \subset Y$, as

(35) $$F^{\wedge 1}(E) = \begin{cases} \{x \in X : F(x) \neq \varnothing \text{ and } F(x) \subset E\} & \text{if } E \neq \varnothing \\ \varnothing & \text{if } E = \varnothing \end{cases}.$$

This puts, for all $E \subset Y$,

(36) $$F^{\wedge 1}(E) \subset X \sim F^{+1}(\varnothing) = \operatorname{cor}(F).$$

One sees that

(37) $$F^{+1}(E) = F^{\wedge 1}(E) \cup F^{+1}(\varnothing) \text{ and } F^{\wedge 1}(E) \cap F^{+1}(\varnothing) = \varnothing$$

(i.e., $F^{+1}(E)$ is the union of the disjoint sets $F^{\wedge 1}(E)$ and $F^{+1}(\varnothing)$), and

(38) $$F^{\wedge 1}(E) \subset F^{-1}(E).$$

Also

(39) $$F^{-1}(E) \cap F^{+1}(\varnothing) = \varnothing .$$

In particular,

(40) $$F^{-1}(Y) = F^{\wedge 1}(Y) = X \sim F^{+1}(\varnothing) = \operatorname{cor}(F)$$

and

(41) $$F^{+1}(Y) = X .$$

2.17 Lemma *For a set-valued mapping* $F : X \multimap Y$ *and* $E \subset Y$ *,*

(42) $$F^{-1}(Y \sim E) = X \sim F^{+1}(E);$$

(43) $$F^{+1}(Y \sim E) = X \sim F^{-1}(E).$$

With the identities (32) and (37), one has

2.18 Corollary *For a set-valued mapping* $F : X \multimap Y$ *and* $E \subset Y$ *,*

(44) $$F^{-1}(Y \sim E) = \operatorname{cor}(F) \sim F^{\wedge 1}(E);$$

(45) $$F^{\wedge 1}(Y \sim E) = \operatorname{cor}(F) \sim F^{-1}(E).$$

Note that among the three varieties of 'inverse images' that I have defined, inverse image $F^{-1}(E)$, core $F^{+1}(E)$, and upper inverse $F^{\wedge 1}(E)$, only the first is associated with an 'inverse mapping', viz., $F^{-1} : Y \multimap X$, with

(46) $$F^{-1}(y) = F^{-1}(\{y\}) = \{x \in X : F(x) \cap \{y\} \neq \varnothing\} = \{x \in X : y \in F(x)\}.$$

While one may similarly define $F^{+1}(y)$ and $F^{\wedge 1}(y)$,

(47) $$F^{+1}(y) = F^{+1}(\{y\}) = \{x \in X : F(x) \subset \{y\}\},$$

Let $A \subset \operatorname{cor}(F)$, whence $A \subset F^{-1}(F(A))$ by Lemma 2.21. A nonempty $F^{-1}(F(A)) \sim A$ means the existence of an element $x \in F^{-1}(F(A)) \sim A$. This element x must be in $\operatorname{cor}(F) \sim A$, which means $F(x) \cap F(\operatorname{cor}(F) \sim A) \neq \varnothing$. At the same time, $x \in F^{-1}(F(A)) \sim A$, so *a fortiori* $x \in F^{-1}(F(A))$, which means $F(x) \cap F(A) \neq \varnothing$. In other words, $A = F^{-1}(F(A))$ iff there is no $x \in \operatorname{cor}(F)$ such that $F(x) \cap F(A) \neq \varnothing$ and $F(x) \cap F(\operatorname{cor}(F) \sim A) = \varnothing$. But this equivalent condition is the same when A is replaced by $\operatorname{cor}(F) \sim A$, since $\operatorname{cor}(F) \sim (\operatorname{cor}(F) \sim A) = A$, whence it also defines the conditions under which $\operatorname{cor}(F) \sim A = F^{-1}(F(\operatorname{cor}(F) \sim A))$. Thus $A = F^{-1}(F(A))$ iff $\operatorname{cor}(F) \sim A = F^{-1}(F(\operatorname{cor}(F) \sim A))$, and this says $A \in \mathfrak{S}$ iff $\operatorname{cor}(F) \sim A \in \mathfrak{S}$. $\mathfrak{S}$ is therefore complemented.

Let $A, B \in \mathfrak{S}$. Then, using Theorem 2.10.iii and Theorem 2.19.iii,

$$F^{-1}(F(A \cup B)) = F^{-1}(F(A)) \cup F^{-1}(F(B)) = A \cup B, \tag{49}$$

so $A \cup B \in \mathfrak{S}$. Since $\operatorname{cor}(F) \sim (A \cap B) = (\operatorname{cor}(F) \sim A) \cup (\operatorname{cor}(F) \sim B) \in \mathfrak{S}$, one also has $A \cap B \in \mathfrak{S}$. □

Theorem 1.22 says that a mapping $f : X \to Y$ is injective iff $A = f^{-1}(f(A))$ for all $A \subset X$. Correspondingly, one has

2.24 Lemma *A set-valued mapping* $F : X \multimap Y$ *is injective iff* $A = F^{-1}(F(A))$ *for all* $A \subset \operatorname{cor}(F)$.

When a set-valued mapping $F : X \multimap Y$ is injective, every subset of $\operatorname{cor}(F)$ is stable; the complemented lattice $\mathfrak{S}$ of stable subsets is thus all of $\mathcal{P}(\operatorname{cor}(F))$.

An injective mapping $f : X \to Y$ means $f^{-1} \circ f = 1_X$, the identity mapping $x \mapsto x$ on the domain X. But an injective set-valued mapping $F : X \multimap Y$ means

$$F^{-1}(F(x)) = \begin{cases} \{x\} & \text{if } x \in \operatorname{cor}(F) \\ \varnothing & \text{if } F(x) = \varnothing \end{cases}, \tag{50}$$

i.e., $x \mapsto F^{-1}\left(F(x)\right)$ is a disjoint union, the concatenation of the inclusion map i of $\operatorname{cor}(F)$ in X and the constant empty-set-valued mapping $\varnothing$ on $F^{+1}(\varnothing) = X \sim \operatorname{cor}(F)$. So even for an injective F, the combination $G(x) = F^{-1}\left(F(x)\right)$ is still not quite the identity mapping on X (unless $F^{+1}(\varnothing) = \varnothing$). When i and G are considered as subsets of $X \times X$, $G = i = \left\{(x,x) : x \in A\right\}$ (but not necessarily $G = 1_X = \left\{(x,x) : x \in X\right\}$).

2.25 Second Combination Given a mapping $f : X \to Y$ and $B \subset Y$, one has $B \supset f\left(f^{-1}(B)\right)$ (Theorem 1.21.ii). But there is no containment relation between B and $F\left(F^{-1}(B)\right)$ for a set-valued mapping $F : X \multimap Y$. Consider the example $F : \{1,2\} \multimap \{p,q,r\}$ with $F(1) = \{p,q\}$ and $F(2) = \varnothing$; then $F\left(F^{-1}\left(\{q,r\}\right)\right)$ $= F\left(\{1\}\right) = \{p,q\}$. This time, even a restriction to $B \subset \operatorname{ran}(F)$ (dual to $A \subset$ $\operatorname{cor}(F)$ in Lemma 2.21) does not help: in the example, $\{p\} \not\supset F\left(F^{-1}\left(\{p\}\right)\right) =$ $F\left(\{1\}\right) = \{p,q\}$. Neither does the specialization to surjections: Theorem 1.23 says that a mapping $f : X \to Y$ is surjective iff $B = f\left(f^{-1}(B)\right)$ for all $B \subset Y$. But the same F in my example is a surjective set-valued mapping from $\{1,2\}$ onto $\{p,q\}$, and still $\{p\} \neq F\left(F^{-1}\left(\{p\}\right)\right) = F\left(\{1\}\right) = \{p,q\}$.

Since neither $x \mapsto F^{-1}\left(F(x)\right)$ nor $y \mapsto F\left(F^{-1}(y)\right)$ is necessarily the identity mapping on its respective domain, one must understand the usage of the term '*inverse* set-valued mapping' with this in mind: it is not the usual algebraic definition in connection with a 'reversal entity for the recovery of the identity'. For this reason, some authors call $F^{-1} : Y \multimap X$ the 'converse' of $F : X \multimap Y$ instead of the 'inverse'.

2.26 Third Combination Given a mapping $f : X \to Y$, $A \subset X$, and $B \subset Y$, one has $f\left(A \cap f^{-1}(B)\right) = f(A) \cap B$ (Theorem 1.21.iii).

Consider the example $F : \{1,2\} \multimap \{p,q\}$ with $F(1) = \{p,q\}$ and $F(2)$ $= \{q\}$; then $F\left(\{1\} \cap F^{-1}\left(\{p\}\right)\right) = F\left(\{1\} \cap \{1\}\right) = \{p,q\}$, but $F\left(\{1\}\right) \cap \{p\}$ $= \{p,q\} \cap \{p\} = \{p\}$. So they are not equal. But one does have inclusion:

2.27 Theorem *Let $F : X \multimap Y$, $A \subset X$, and $B \subset Y$. Then $F\left(A \cap F^{-1}(B)\right)$ $\supset F(A) \cap B$.*

Operations on Set-Valued Mappings

2.28 Definition If $F: X \multimap Y$ and $G: X \multimap Y$ are two set-valued mappings, then:

i. Their *union* is the mapping $F \cup G: X \multimap Y$ defined by $(F \cup G)(x) = F(x) \cup G(x)$.

ii. Their *intersection* is the mapping $F \cap G: X \multimap Y$ defined by $(F \cap G)(x) = F(x) \cap G(x)$.

iii. Their *Cartesian product* is the mapping $F \times G: X \multimap Y \times Y$ defined by $(F \times G)(x) = F(x) \times G(x)$.

2.29 Theorem *Let $F: X \multimap Y$ and $G: X \multimap Y$. Then, for $A \subset X$:*

i. $(F \cup G)(A) = F(A) \cup G(A)$.

ii. $(F \cap G)(A) \subset F(A) \cap G(A)$.

iii. $(F \times G)(A) \subset F(A) \times G(A)$.

Recall (Section 2.4) that $F: X \multimap Y$ is a constant (set-valued) mapping if there exists a subset C of Y such that $F(x) = C$ for all $x \in X$. This implies $F(A) = C$ for all $A \subset X$.

2.30 Corollary *Let $F: X \multimap Y$ be a constant mapping. Let $G: X \multimap Y$ and $A \subset X$. Then $(F \cap G)(A) = F(A) \cap G(A)$.*

2.31 Theorem *If both $F: X \multimap Y$ and $G: X \multimap Y$ are semi-single-valued, then the set-valued mappings $F \cap G: X \multimap Y$ and $F \times G: X \multimap Y \times Y$ are semi-single-valued.*

2.32 Theorem *If one of $F: X \multimap Y$ and $G: X \multimap Y$ is injective, then the set-valued mappings $F \cap G: X \multimap Y$ and $F \times G: X \multimap Y \times Y$ are injective.*