Mathematical Foundations of Consciousness

Willard L. Miranker, Gregg J. Zuckerman
Departments of Computer Science, Mathematics
Yale University
TR1383

Mathematical Foundations of Consciousness

Willard L. Miranker[1], Gregg J. Zuckerman[2]
Departments of [1]Computer Science, [2]Mathematics
Yale University

This work is dedicated to the memory of our mentors.

"Mathematics as an expression of the human mind reflects the active will, the contemplative reason, and the desire for aesthetic perfection. Its basic elements are logic and intuition, analysis and construction, generality and individuality. Though different traditions may emphasize different aspects, it is only the interplay of these antithetic forces and the struggle for their synthesis that constitute the life, usefulness, and supreme value of mathematical science." [1]*Richard Courant* (1941).

"One expects that logic, as a branch of applied mathematics, will not only use existing tools from mathematics, but also that it will lead to the creation of new mathematical tools, tools that arise out of the need to model some real world phenomena not adequately modeled by previously known mathematical structures." [2]*Jon Barwise* (1988).

Abstract: We employ the Zermelo-Fränkel Axioms that characterize sets as mathematical primitives. The Anti-foundation Axiom plays a significant role in our development, since among other of its features, its replacement for the Axiom of Foundation in the Zermelo-Fränkel Axioms motivates Platonic interpretations. These interpretations also depend on such allied notions for sets as pictures, graphs, decorations, labeling and various mappings that we use. A syntax and semantics of operators acting on sets is developed. Such features enable construction of a theory of non well-founded sets that we use to frame mathematical foundations of consciousness. To do this we introduce a supplementary axiomatic system that characterizes experience and consciousness as primitives. The new axioms proceed through characterization of so-called consciousness operators. The Russell operator plays a central role and is shown to be one example of a consciousness operator. Neural networks supply striking examples of non well-founded graphs the decorations of which generate associated sets, each with a Platonic aspect. Employing our foundations, we show how the supervening of consciousness on its neural correlates in the brain enables the framing of a theory of consciousness by applying appropriate consciousness operators to the generated sets in question.

Key words: foundations of consciousness, neural networks, non well founded sets, Russell operator, semantics of operators

1. Introduction

Analytic writing on consciousness dates to Aristotle's De Anima (Ross, ed., 1961). Yet to this day the phenomena of consciousness continue to elude illuminating scientific characterization. We should not be surprised at this since,

"A physical scientist does not introduce awareness (sensation or perception) into his theories, and having thus removed the mind from nature, he cannot expect to find it there" (Schrödinger, 1958).

The self-referential qualities of consciousness place it outside conventional logic(s) upon which scientific models and frameworks have heretofore been constructed. However more contemporary mathematical development has begun to deal with features of self- reference. We shall address Schrödinger's critique by assembling and extending such development thereby putting self-reference as a form of awareness into theory. In this way we shall frame mathematical foundations for a theory of consciousness. Then as an application to a neural network model of brain circuitry, we shall exhibit a theory of consciousness using these foundations.

1.1 Mathematical thought and its limits

Platonism, that is, the interplay of ideal and physical worlds characterizes a central feature of mathematical thought. The briefest summary of the evolution of this Platonic dualism in mathematical thought and modeling might be made by citing the contributions of Euclid (the axiomatic method), Aristotle (the law of the excluded middle), Cantor, 1895 (set theory), Russell, 1910 (the paradox of set theory), Zermelo, 1908 and Fränkel, 1922 (the axioms of set theory that serve to accommodate the paradox) and Gödel (incompleteness, a self-referential development). Indeed we shall extend this line and employ the axiomatic theory of sets to further characterize self-referential features.

The work of Zermelo-Fränkel and others transformed sets from so-called naïve objects into mathematical primitives (i.e., ideal Platonic objects). The Russell paradox and its accommodation demonstrate limitations on mathematical thought (about sets and related constructs). Today we are not surprised at the existence of such a limitation, since we have the well-known example of Heisenberg. The Platonic character of the latter is characterized by the Heisenberg inequality, its ideal form (Dym, McKean, 1972) and its real world character by the limitation on the accuracies with which certain concurrent measurements can be made. The quality of self-reference (set self-membership) underlying the Russell paradox informs development of the ideal Platonic structures (i.e., of placing awareness into theory) required for constructing the mathematical foundations we seek.

1.2 Consciousness and its limits

As within naïve set theory, the self-referential character of consciousness appears paradoxical. It seems to be an illusion. The incompleteness of mathematical thought shown by Gödel, suggests that all thought, and so, consciousness in particular, is not explainable via a conventional approach such as by a Turing machine computation for instance (Penrose, 1989). Incompleteness, while precluding establishment of certain knowledge within a system, allows for its establishment by looking onto the system from the outside. This knowledge from the outside (a kind of observing) is reminiscent of consciousness that provides as it does a viewing or experiencing of what's going on in thought processing. Note the correspondence of these observations to Freud's meta-psychology where he recognizes a disconnect between mental and physical states,

> "*...mental and physical states represent two different aspects of reality, each irreducible to the other.*" (Solms, 1994).

However we may say that Freud's psychoanalytic method is a tool devised for penetrating the mental from the outside via the physical. Compare these dual aspects of reality with the *res cogitans* (Platonic) and the *res extensa* (physical) of Descartes, 1637.

To frame a set theoretic correspondent to these features note that a set has an inside (its elements) and an outside (the latter is not a set, as we shall see), and this allows a set to be studied from the outside. We liken this to interplay between the ideal (Platonic) and physical (computable) worlds, the latter characterizing a model for study from the outside of the former. So we expect consciousness to be accessible to study through extensions of the self-reference quality characterized by axiomatic set theory, in particular, by a special capacity to study a set from the outside. We pursue this approach as an effectual way to introduce awareness into theory.

1.3 Summary

Sect. 2 begins with a description of the crises in mathematical thought precipitated by Cantor's set theory and characterized by the Russell paradox. We describe how Gödel's discoveries inform the crises and furnish motivation for our development. We introduce a mathematical framework that includes sets, graphs, decorations, and the notion of non well-founded sets and which enables annunciation of the anti-foundation axiom of set theory. This axiom allows replacement of the Russell paradox by a logically coherent dichotomy and is key to framing our approach characterized by observation of sets from the outside.

In Sect. 3 we introduce the Russell operator $\mathcal{R}$ a distinguisher between so-called normal and abnormal sets. A number of properties of $\mathcal{R}$ is collected, these to play a central role in the foundations to be developed. Then we introduce a number of other operator mappings along with interrelations, these to supplement $\mathcal{R}$ in the analysis of sets to follow. This operator syntactic framework is followed by a semantic development in which experience and consciousness are introduced as primitives. A Semantic Thesis for consciousness is then proposed, and a list of axioms for associated operators, along with a descriptive semantics for each axiom is given (compare Aleksander, Dunmall, 2003). The axioms along with their semantics are used for characterizing both the primitives and the Consciousness Thesis. $\mathcal{R}$ is shown to satisfy the axioms, giving it thereby the role of a so-called consciousness operator. This existence of a consciousness operator establishes consistency of the new axioms. Examples both of sets and operators illustrating the syntax and semantics are given.

In Sect. 4 we give a description of tools for building a theory of consciousness upon the foundations developed. This begins with a formal process for labeling and then decorating a graph. The process establishes a way to induce existence of a virtual set associated intrinsically with a graph (a two-level or self-referential feature). A mapping construct called a histogram is then introduced, a tool for applying this set with graph association process to a special class of graphs arising in brain circuitry. The M-Z equation is then developed, this equation characterizing a method for specifying the intrinsic set in question, including those that arise in brain circuitry. Finally the theory of consciousness is formulated as an application in which we employ neural network theory (Hebb's rule for synaptic weight change and the McCulloch-Pitts equation for neuronal input-output dynamics, (see Haykin, 1999)) to specify the special class of labeled graphs in question. This two-level procedure is interpreted as a Platonic process (that is, the association of a virtual set with a graph) by means of what we call a Neural Network Semantic Thesis. To complete the description of information processing from sensory perception through to consciousness, a third, purely physical, so-called Neuro-physiological Thesis is introduced. Sect. 4 concludes with a critical description both of these three theses and the analytic formalisms developed earlier. This critique serves to illuminate the mathematical foundations of consciousness developed.

In Sect. 5 we ascribe syntactic and semantic nomenclature to a collection of basic operators, also offering speculative interpretations of the role each plays in our theory. The flow of information from sensory input to conscious experience is described along with a speculation on the role of the sets we have constructed in the experience. Finally speculations on a class of operators that produce qualia are offered.

The axioms of set theory that we employ are given in an appendix. This is followed by a glossary.

2. Preliminaries

In this section we describe the crises in mathematical thought engendered by the notions developed by Cantor, Russell and others. Then we describe the evolution of the crises according to the development of Zermelo-Fränkel, Gödel and others. We continue with the introduction of terminology and properties that provide the setting for our work.

2.1 Crises in mathematical thought

We begin with Cantor's definition of a set, often regarded as the *naïve notion of set.*

> "*A set is a collection into a whole of definite, distinct objects of our intuition or <u>thought</u>.*"

When specificity is required, we shall hereafter use the term *collection* for a set in the sense of Cantor's definition.

Cantor's use of the word "thought" shows that set theory is entwined with consciousness from the start. In fact, Cantor's definition is circular, replacing one mystery by another. It replaces the unanswered questions: what is a definite object? what is thought? by others, namely: who does the collecting? the thinking? The latter have a correspondence to the questions often raised in consciousness studies, "Who is doing the looking? the experiencing?" Suppose the words "intuition or thought" in Cantor's definition are replaced by the word "consciousness". This would make it an exception to Schrödinger's critique, relating it to what is perhaps the only other known exception, namely to Von Neumann's (mysterious) appeal to the observer's consciousness (of the outcome of a measurement) to specify the moment of collapse of the wave function during a quantum mechanical measuring process.

Cantor's definition of a set supports a logical inconsistency, resulting in several paradoxes. The most accessible of these is the Russell paradox that goes to the essence of that inconsistency. This paradox is expressed in terms of the *Russell set N*, which is the collection of all sets x such that x is not a member of x. The logical inconsistency of N is revealed by the following observations:

1. Since N is a set, either $N \in N$ or $N \notin N$.
2. If $N \in N$, then $N \notin N$. If $N \notin N$, then $N \in N$. (2.1)

The annunciation of this paradox by Russell (Zermelo is thought to have known earlier of the paradox) precipitated a major crisis in mathematical and philosophical thought. Frege had just completed development of an axiomatic treatment of sets when a letter to him from Russell informing him of the paradox overturned his central thesis. Various mathematicians (Bernays, Gödel, Hilbert, Russell, Von Neumann, Whitehead...) attempted to rework the foundations of mathematics so as to resolve the(se) paradox(es). It is the axiomatic approach to set theory that provides for us the most fruitful resolution, motivating our own development. (See the appendix for these axioms.) The key feature of the axiomatic approach is to regard the concept "set" as a primitive (an undefined notion), and the concept "is an element of" as a primitive relation. The axioms are chosen to ensure that there does not exist a set y such that $x \in y$ if and only if $x \notin x$; in other words, *within axiomatic set theory, there is no Russell set.* Even so, this axiomatic approach allows for a coherent elaboration of the quality of self-reference in set theory, and so, it supports the connection of the study of sets to the development of the mathematical foundations we are after.

For our development we use Z-F, the Zermelo-Fränkel axioms of set theory, however replacing FA: the Foundation Axiom (a latter day addition by Von Neumann to the original Z-F list) by AFA: the Anti-foundation Axiom. When it is necessary to distinguish a set in the sense of these axioms from a collection of Cantor, we shall use the terminology, *bona fide set.*

Although successfully accommodating the paradox, the axiomatic development of set theory brought with it a deeper problem: is the axiomatic system itself consistent? That is, can we derive a logical inconsistency from the axioms? Gödel produced a two level approach to this issue. At a mathematical level is a set theoretic formula, and at a meta-mathematical level is the proposition asserting the consistency of set theory. We interpret this as an instance of self-reference, a viewing of a mathematical object meta-mathematically, that is from the outside. Gödel showed that if axiomatic set theory is consistent then it is incomplete. This incompleteness is widely celebrated (see Gödel-Escher-Bach of Hofstadter, 1979, Emperor's New Mind of Penrose, 1989, Scientific American, 1968...).

One might say that Gödel replaced one crisis in mathematical thought by another. Subsequently, mathematicians (Aczel, 1988...) did show that if Z-F with FA deleted is consistent, then Z-F with AFA replacing FA is also consistent. These results of Gödel and his successors provide for us the framework to develop our self-referential two level approach that consists, in particular, of *a syntactic level and a semantic level.*

2.2 Sets, graphs, decorations, the axiom of anti-foundation

The special nature of set theory can be traced in part to the use of two different notions of belonging associated with sets. One is denoted by $\in$ (the primitive concept '*is an element of*') and the other by $\subset$ (for the concept '*is a subset of*').

For clarity we adopt the following notational conventions.

a) Sets will be denoted by Latin characters, *a, A, b*...
Braces will also denote a set, the contents of which and/or conditions specifying the set placed within the braces: $\{\textit{list of set elements and / or conditions for being a set element}\}$.
b) Mappings between sets will be denoted by lower case Greek characters, α, β...
c) Relations and operators as well as certain special objects to be introduced called classes will be denoted with upper case script Latin letters, $\mathcal{A}, \mathcal{B}...\mathcal{R}...$ A generic operator will be denoted by an upper case script $\mathcal{O}$.
d) The empty set $\{x \mid x \neq x\}$ will, as usual, be denoted by $\varnothing$. The existence of $\varnothing$ follows from the Z-F Axioms of Existence and Comprehension (see the appendix).

We shall restrict our attention to pure sets:

Definition: A set is a *pure set* if its elements are sets.

Note that any finite collection (naïve set) of objects that are not bona fide sets furnishes an example of a not pure set.

We also note the distinction between normal sets and abnormal sets:

Definition: A set x is *normal* if $x \notin x$. It is *abnormal* if $x \in x$.

Our presentation involves normal and abnormal sets. The *Quine atom* (denoted by Ω), the set defined by the condition $\Omega = \{\Omega\}$, supplies an example of an abnormal set.

We shall make use of a collection of notions specified in the following paragraph. (See Aczel, 1988, Chap. 1.)

A *graph* will consist of a collection N of *nodes* and a collection E of *edges,* each edge being a pair (n, n') of nodes. We have no knowledge of the nature of the elements of N. If (n, n') is an edge, we shall write $n \to n'$ and say that n' is a *child* of its *parent* n. A *path* is a sequence (finite or infinite)

$$n_0 \to n_1 \to n_2 \to \cdots$$

of nodes $n_0, n_1, n_2 \ldots$ linked by edges $(n_0, n_1), (n_1, n_2) \ldots$ A *pointed graph* is a graph together with a distinguished node called its *point*. A pointed graph is *accessible* if for every node n there is a path $n_0 \to n_1 \to \cdots \to n$ from the point n_0 to the node n. If this path is always unique then the pointed graph is a *tree*, and the point is the *root* of the tree. A *decoration* of a graph is an assignment of a set to each node of the graph so that the elements of the set assigned to a node are the sets assigned to the children of that node. Alternatively a decoration is a set valued function d on N such that

$$\forall a \in N, \quad da = \{db \mid a \to b\}. \tag{2.2}$$

A *picture* of a set is an accessible pointed graph (apg) that has a decoration in which the set is assigned to the point.

Being well founded is a key quality of graphs and sets:

Definition: A graph is *well founded* if it has no infinite path. It is *non well founded* otherwise. An alternate name for a non well-founded set is a hyper-set, but we prefer never to use the latter.

With this terminology, we collect the known results stated in the following proposition.

Proposition: *i*) every well-founded graph has a unique decoration.
ii) Every well-founded apg is a picture of a unique set.
iii) Every set has a picture.

Continuing, we define well foundedness for sets.

Definition: A set is *well founded* if its picture is well founded. It is *non well founded* otherwise.

We now state the *anti-foundation axiom* that is central to the development. Note it is stated for general graphs that are not necessarily accessible.

AFA: *Every graph has a unique decoration.*

Some consequences of this axiom are given in the following proposition.

Proposition: 1. Every pointed graph is the picture of a unique set.
2. Non well founded sets exist.
3. Every non well founded graph will have to picture a non well-founded set.

The relationship between these concepts is summarized in terms of two mappings, the tree mapping τ and the decoration of the point P mapping δ, schematized as follows.

$$\text{Pointed graphs} \underset{\delta:\,\text{decoration of the point}}{\overset{\tau:\,\text{tree map}}{\rightleftarrows}} \text{Sets}$$

There are many graphs Γ_i, the decoration of whose point is a given set A. That is, for the map δ, we have

$$\delta\Gamma_1 = \delta\Gamma_2 = \cdots = A.$$

However there is a unique pointed graph, $\Gamma_* = \Gamma_*(A)$ called the tree of A, such that $\delta\Gamma_* = A$ and

$$\tau A = \Gamma_*(A).$$

We shall use $\delta(\Gamma, p)$ to denote the set associated with the node p of the graph Γ in the decoration of the latter. So $\delta\Gamma = \delta(\Gamma, P)$ is the set in the decoration of the pointed graph Γ that corresponds to the point P of Γ. Then a sufficient condition for normality if a set is given in the following proposition.

Proposition: If for every child c of P, $\delta(\Gamma, c) \neq \delta(\Gamma, P)$, then $\delta(\Gamma, P)$ is normal.

2.3 Classes and mappings

Classes are primitives introduced by Gödel. A collection of sets with a common property is called a *class*. A set is also a class; a class that is not a set is called a *proper class*. The elements of a class are sets, the sets being the primitives defined by the Z-F axioms with the AFA replacing the AF. Conversely, any set is a member of a class.

We now formalize the notions of several types of mappings to be used. These are: relations, function, and operators. They are illustrated by the nest of concepts the outermost member of which is comprised of the Classes as shown in Fig. 2.1.

Inside of classes is the collection of relations. A *relation* is a class consisting of ordered pairs of sets.

Inside of relations is the collection of functions. A *function* is relation with the graph property. Namely, if (x,y) and (x,z), both being ordered pairs of sets in a relation $\mathcal{F}$, implies that $y = z$, then $\mathcal{F}$ is said to have the *graph property*.

Inside of functions is the collection of operators. An *operator* is a function whose domain is the class of all sets. To see that an operator O is a relation, note that Ox equals the unique y such that $(x,y) \in O$.

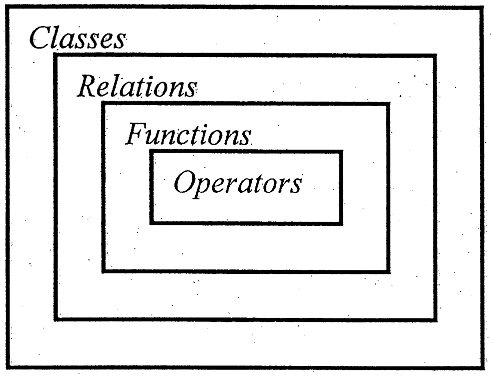

Figure 2.1: Nesting within the model

3. The Russell Operator, Operator Syntax and Semantics, Semantic Thesis, Axioms

In this section we supply syntax and semantics for some operators of relevance for our axiomatic treatment of consciousness. We start in Sect. 3.1 with the characterization of the Russell operator, since it plays a central role. Then in Sect. 3.2, we introduce a relevant collection of operators and develop mathematical properties (syntax) for them. In Sect. 3.3, we state the Semantic Thesis that characterizes consciousness as the action of operators on experience. *Consciousness and experience are introduced as primitives*, and an open axiom system for them is elaborated. The axioms are accompanied by semantic characterizations of the associated operators.

3.1 The Russell operator

The Russell operator $\mathcal{R}$ plays a special role in the syntax and semantics of the development of the Semantic Thesis. $\mathcal{R}$ is defined by its action on a set A as follows.

Definition 3.0: $\mathcal{R}A = \{x \in A \mid x \notin x\}$.

So we see that $\mathcal{R}$ may be viewed as a selector of the normal elements of A and a rejecter of the abnormal. $\mathcal{R}$ is a special case of a generic operator O_P specified in terms of a predicate $P(y)$ as

$$O_P A = \{y \in A \mid P(y)\}.$$

We recognize this as the Axiom of Comprehension. So $O_P A$ is a bona fide set. Then it follows that

$$x \subseteq y \Rightarrow O_P x = x \cap O_P y. \qquad (3.1)$$

This relation holds in particular for O_P set equal to $\mathcal{R}$.

The Russell paradox is no longer relevant as a paradox. It is replaced by the operator $\mathcal{R}$ as examination of the proof of the following theorem reveals (compare (2.1).

Theorem 3.1: $\forall A$, $\mathcal{R}A \notin A$.

Proof: Assume there exists a set z such that $\mathcal{R}z \in z$. Then by the definition of $\mathcal{R}$ there are two options, both of which lead to contradictions. Namely,

1. $\mathcal{R}z \in \mathcal{R}z$, in which case $\mathcal{R}z \notin \mathcal{R}z$,
2. $\mathcal{R}z \notin \mathcal{R}z$, in which case $\mathcal{R}z \in \mathcal{R}z$.

A corresponding result is

Proposition: $\forall A$, $A \notin \mathcal{R}A$.

Proof:

By definition, if $x \in \mathcal{R}$A, then it is both true that $x \notin x$ and $x \in A$. Then $A \in \mathcal{R}$A implies both $A \notin A$ and $A \in A$, a contradiction. $\square$

We make the following observations associated with Theorem 3.1.

a) There is no set of all sets.
b) Every set has an inside and an outside,
where the inside of a set consists of its elements.
c) The compliment of a set (the balance of a class) is not a set.
d) If $\forall y \in B, y \notin y$, then $B \notin B$.
e) If $\forall y \in C, y \in y$, we cannot conclude that $C \in C$.

Since $\mathcal{R}$ takes a part of A outside itself, note the relevance of b) to the ability to observe a set from the outside described in Sect. 1.

e) is illustrated by the following two examples.

Example 1: Since $\Omega \in \Omega$, taking $C = \Omega$ satisfies the hypotheses, and we have $C \in C$.

Example 2: Take x and y to be unequal normal sets, and let $C = \{x^*, y^*\}$ where $x^* = \{x^*, x\}$ and $y^* = \{y^*, y\}$. We claim that $C \notin C$.

Let $\mathcal{U} = \{x \mid x = x\}$ be the *class of sets*. ($\mathcal{U}$ may be referred to as the *universe of sets*.) Let $\mathcal{N} = \{x \mid x \notin x\}$, the collection of normal sets, and let $\mathcal{A} = \{x \mid x \in x\}$, the class of abnormal sets. Then we have the following proposition concerning the classes $\mathcal{A}$, $\mathcal{N}$ and $\mathcal{U}$ and the Russell operator $\mathcal{R}$.

Proposition 3.2: a) $\mathcal{N}$ is a proper class
b) $\mathcal{U}$ is a proper class
c) $\mathcal{A}$ is a proper class
d) $\mathcal{U} = \mathcal{N} \cup \mathcal{A}$.
e) $\mathcal{R}A = \mathcal{N} \cap A, \forall A$.

Proof: We shall prove a), b) and c).

a) Assume not. Then $\mathcal{N} = A$ for some set A. $\mathcal{R}A \notin A$ by Theorem 3.1. But then $\mathcal{R}A \in \mathcal{N}$, a contradiction.

b) Assume not. Then $\mathcal{U}=B$ for some set B. Now $\{x \in B | x \notin x\}$ is a set by the Axiom of Comprehension. However $\{x \in B | x \notin x\} = \mathcal{N}$, by definition. This is a contradiction since $\mathcal{N}$ is a proper class,

c) Suppose to the contrary that $\mathcal{A}$ is a set. Then $\exists!$ a set A such that $x \in A \Leftrightarrow x \in x$. Then using AFA, $\forall y \in \mathcal{N}$, $y^* \in A$. Let $C = \{x \in A | \exists y \in \mathcal{N} \ni x = y^*\}$. By the Axiom of Comprehension, C is itself a set that we may alternatively write as

$$C = \{y^* | y \in \mathcal{N}\} = \{\{y^*, y\} | y \in \mathcal{N}\}.$$

Then using the Axiom of Union, can write

$$\bigcup C = \bigcup_{y \notin y} \{y^*, y\},$$

where $\bigcup$ is the monadic union operator ($\bigcup A = \{x | x \in a \text{ for some } a \in A\}$). Then

$$\mathcal{R}(\bigcup C) = \bigcup_{y \notin y} \{y\} = \mathcal{N}.$$

This is a contradiction, since $\mathcal{R}(\bigcup C)$ is a set and $\mathcal{N}$ is a proper class. $\square$

3.2 Syntax

3.2.1 Fundamental operators

We shall employ the following four basic dyadic set operations $\circ, \cup, \cap, -$, defined as follows.

$$\begin{aligned}
\circ:&\quad O_1 \circ O_2 x = O_1 O_2 x \\
\cup:&\quad (O_1 \cup O_2)\, x = (O_1 x) \cup (O_2 x) \\
\cap:&\quad (O_1 \cap O_2)\, x = (O_1 x) \cap (O_2 x) \\
-:&\quad (O_1 - O_2)\, x = (O_1 x) - (O_2 x)
\end{aligned}$$

The last, the difference of operators, is defined in terms of set subtraction, which is specified by the following Boolean rule.

$$x - y = x - (x \cap y).$$

The associative law $(O_1 O_2) O_3 = O_1 (O_2 O_3)$ follows from the definition of $\circ$.

To supplement $\mathcal{R}$ we introduce four additional basic operators I, $\mathcal{E}$, $\mathcal{B}$ and $\mathcal{D}$. Let

a) I be the *identity operator*, $Ix = x$,
b) $\mathcal{E}$ be the *elimination operator*, $\mathcal{E}x = \varnothing$,
c) $\mathcal{B}$ be the *singleton operator*, $\mathcal{B}x = \{x\}$, and
d) $\mathcal{D}$ be the *duality operator*, $x^* = \mathcal{D}x = \{x^*, x\}$.

a) employs the Axiom of Extensionality, and c) the Axiom of Pairing.

The following are four observations about these basic operators.

i) $IO = OI$, for any operator O.
ii) $\mathcal{E}$ is only a left-zero operator, since for example, $\mathcal{B}\mathcal{E} \neq \mathcal{E}$. Note that $\mathcal{E}$ is *idempotent* ($\mathcal{E}^2 = \mathcal{E}$). Note also that $(\mathcal{B}\mathcal{E})x = \mathcal{B}\varnothing$, so that in particular, $(\mathcal{B}^n\mathcal{E})x = \mathcal{B}^n\varnothing$ for any non-negative integer n.
iii) Since $\{x, y\} = \{x\} \cup \{y\}$, showing that $\{x, y\}$ is a derivative notion (see a) in Sect. 2.2), we can write

$$x^* = (\mathcal{B}x^*) \cup \mathcal{B}x.$$

iv) The set x^* is called the *dual* of x. The existence and uniqueness of the dual of a set follows from the AFA (see Sect. 2.2). Note that

$$\mathcal{D} = (\mathcal{B}\mathcal{D}) \cup \mathcal{B}.$$

3.2.2 Properties of $\mathcal{R}$

The quadruple of basic operators $\mathcal{E}$, I, $\mathcal{R}$ and $\mathcal{B}$ form a non-closed system illustrated in the following operator multiplication table.

	$\mathcal{E}$	I	$\mathcal{R}$	$\mathcal{B}$
$\mathcal{E}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{E}$	$\mathcal{E}$
I	$\mathcal{E}$	I	$\mathcal{R}$	$\mathcal{B}$
$\mathcal{R}$	$\mathcal{E}$	$\mathcal{R}$	$\mathcal{R}$	$\mathcal{R}\mathcal{B}$
$\mathcal{B}$	$\mathcal{B}\mathcal{E}$	$\mathcal{B}$	$\mathcal{B}\mathcal{R}$	$\mathcal{B}\mathcal{B}$

Next we introduce the *counter-Russell* operator, $\mathcal{T} = I - \mathcal{R}$. Note that

$$\mathcal{T}A = A - \mathcal{R}A = A \cap \mathcal{A}.$$

Consider the following proposition relating $\mathcal{R}$ and $\mathcal{T}$ to normal and abnormal sets.

Proposition: Let B be a normal set and C an abnormal set. Then $\mathcal{R}B = B$ but $\mathcal{R}C = \varnothing$, and reversely, $\mathcal{T}C = C$ but $\mathcal{T}B = \varnothing$.

We also have the following proposition exhibiting properties of $\mathcal{R}$ and $\mathcal{B}$.

Proposition: *i*) $x \in A \Leftrightarrow \mathcal{B}x \subset A$.
ii) $\mathcal{R}A \notin A \Leftrightarrow \mathcal{B}\mathcal{R}A \not\subset A$.
iii) $\mathcal{R}A \notin \mathcal{R}A$.

iii) implies that $\mathcal{R}A$ is normal.

Additional syntactical relations (conceptual operator statements) are given in the following theorem, proposition and corollary.

Theorem 3.3: a) $I \cap \mathcal{R} = \mathcal{R}$
b) $\mathcal{B} \cap \mathcal{R} = \mathcal{E}$
c) $I \cap (\mathcal{B}\mathcal{R}) = \mathcal{E}$
d) $I \cap (\mathcal{R}\mathcal{B}) = \mathcal{E}$
e) $\mathcal{R}\mathcal{B} = \mathcal{B} - I$
f) $(\mathcal{R}\mathcal{B}) - (\mathcal{B}\mathcal{R}) \neq \mathcal{E}$
$(\mathcal{B}\mathcal{R}) - (\mathcal{R}\mathcal{B}) \neq \mathcal{E}$

The two statements in f) are not the same since there is no monadic minus for sets.

Proposition 3.4: $\mathcal{R}\mathcal{B}\mathcal{R} = \mathcal{B}\mathcal{R}$.

Proof:
$$\begin{aligned}
\mathcal{R}\mathcal{B}\mathcal{R} &= (\mathcal{R}\mathcal{B})\mathcal{R} \\
&= (\mathcal{B} - I)\mathcal{R}, \text{ using e)} \\
&= \mathcal{B}\mathcal{R} - \mathcal{R}, \text{ by def.} \\
&= \mathcal{B}\mathcal{R} - (\mathcal{B}\mathcal{R} \cap \mathcal{R}), \text{ using set subtraction} \\
&= \mathcal{B}\mathcal{R} - ((\mathcal{B}\mathcal{R} \cap I) \cap \mathcal{R}), \text{ using a)} \\
&= \mathcal{B}\mathcal{R} - (\mathcal{E} \cap \mathcal{R}), \text{ using c)} \\
&= \mathcal{B}\mathcal{R} - \mathcal{E} \\
&= \mathcal{B}\mathcal{R}. \qquad \square
\end{aligned}$$

Corollary: $((\mathcal{R}\mathcal{B}) - \mathcal{B}\mathcal{R})\mathcal{R} = \mathcal{E}$, and $((\mathcal{B}\mathcal{R}) - \mathcal{R}\mathcal{B})\mathcal{R} = \mathcal{E}$.

This corollary gives a connection between the Prop. 3.4 and relation f) in Thm. 3.3.

3.2.3 Characterization of $\mathcal{R}$

The following proposition and corollary gives a complete characterization of $\mathcal{R}$.

Proposition 3.5: Let O satisfy the hypothesis

$$x \subseteq y \Rightarrow Ox = x \cap Oy. \tag{3.2}$$

Then $O\mathcal{B}$ uniquely determines O.

The conclusion of the proposition may be restated alternatively as

$$\forall x,\, Ox = \{y \in x \,|\, O\mathcal{B}y = \mathcal{B}y\}. \tag{3.3}$$

Proof of Prop. 3.5: We make the following preliminary observations. (i) The hypothesis implies that $\forall x,\, Ox \subseteq x$, and hence (ii) $O^2x = Ox \cap Ox = Ox$.

We shall now address the question: when is $y \in Ox$? However $y \in Ox \Leftrightarrow \mathcal{B}y \subseteq Ox$, by definition. In the hypothesis we may replace x with $\mathcal{B}y$ and y with Ox to conclude that

$$O\mathcal{B}y = \mathcal{B}y \cap O(Ox) = \mathcal{B}y \cap Ox, \tag{3.4}$$

the last employing (ii). Now $\mathcal{B}y \subseteq Ox \Leftrightarrow \mathcal{B}y \cap Ox = \mathcal{B}y$, by definition. This and (3.4) implies that $y \in Ox$ if and only if

$$O\mathcal{B}y = \mathcal{B}y.$$

From this and (i) we conclude that $y \in Ox \Leftrightarrow y \in x$ and $O\mathcal{B}y = \mathcal{B}y$. □

Corollary: Let the operators O_1 and O_2 satisfy the hypothesis of the proposition, and let $O_1\mathcal{B} = O_2\mathcal{B}$. Then $O_1 = O_2$.

$\mathcal{R}$ is characterized by the following two properties.

1. $x \subseteq y \Rightarrow \mathcal{R}x = x \cap \mathcal{R}y$.
2. $\mathcal{R}\mathcal{B} = \mathcal{B} - \mathcal{I}$.

The first follows from (3.1). The second is the result e) of Thm. 3.3.

Selectors: Another class of operators of interest are those that satisfy the hypothesis $x \subseteq y \Rightarrow Ox = x \cap Oy$ of Prop. 3.5. We shall call such operators, *selectors* (see Def. 3.0). They form a commutative system. In particular, consciousness operators $\mathcal{K}$, being selectors, commute. Theorem 3.3(f) provides an example of a non-commuting pair of operators.

3.2.4 Schematic illustrating syntax of sets and operators

The Venn type diagram in Fig. 3.1 illustrates some of the notions being discussed. The diagram is intended to be composed in a homeomorphic representation of the Euclidean plane. In the diagram sets are represented by open topological sets. That is, they do not contain their boundaries. For example in terms of the rectangular coordinates α and β in the plane, the empty set is given by $\varnothing = \{\alpha, \beta \mid \alpha^2 + \beta^2 < 0^2\}$.

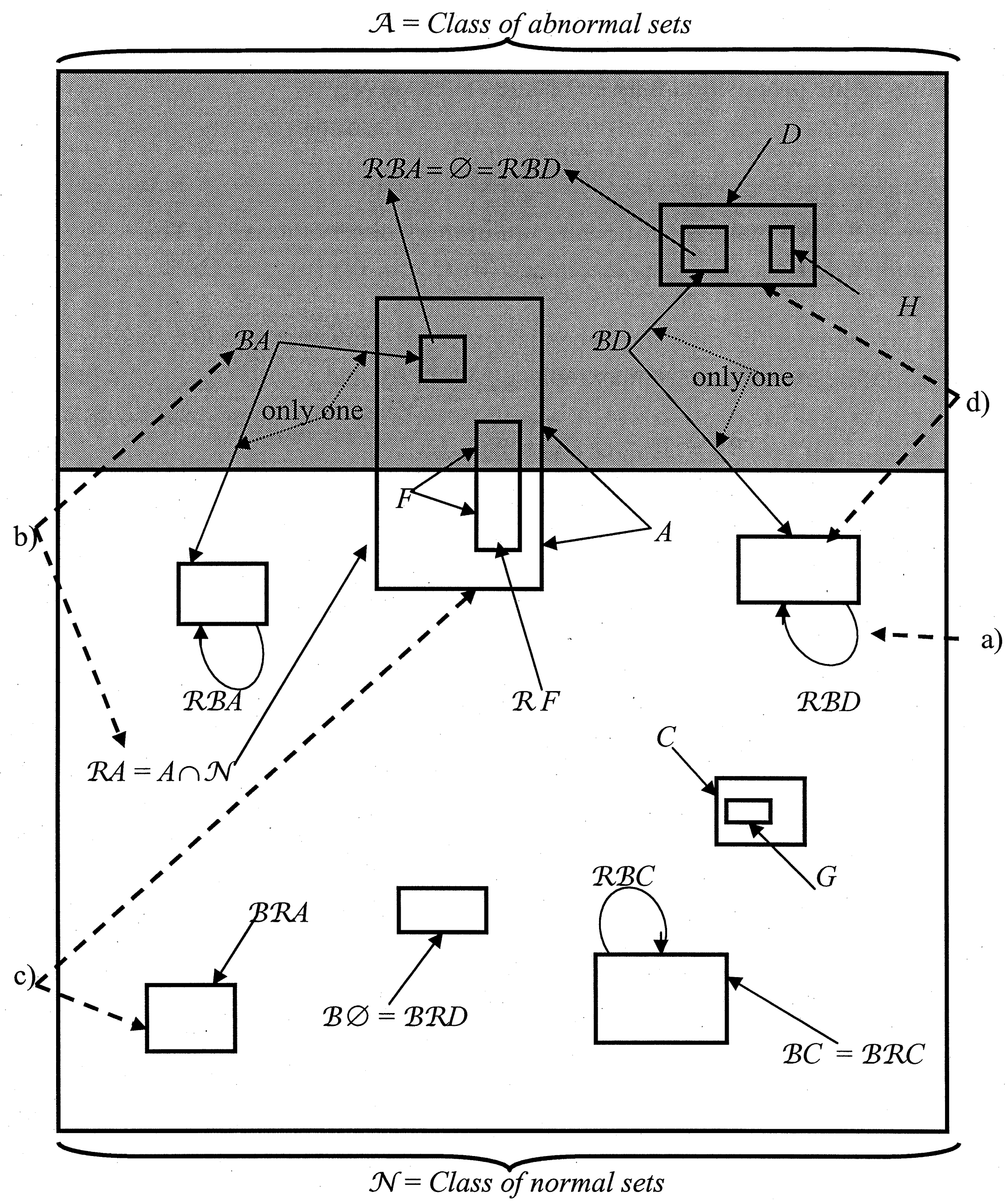

Figure 3.1: A schematic illustrating the properties assembled in Theorem 3.3.

In Fig. 3.1 the class of abnormal sets is shaded to distinguish it from the class of normal sets. Also illustrated are 6 possibilities for sets and 7 for fundamental operators:

2 for set A, depending on whether $\mathcal{B}A \subseteq A$ or not.

1 for set C, namely, $\mathcal{B}C \not\subset C$.

2 for set D, depending on whether $\mathcal{B}D \subseteq D$ or not.

1 for $\varnothing$, a technical possibility, since $\varnothing$ can not be illustrated.

The 7 illustrated fundamental operators are $\mathcal{E}$, I, $\mathcal{B}$, $\mathcal{R}$, $\mathcal{T}$, $\mathcal{BR}$, and $\mathcal{RB}$, although $\mathcal{E}$ and I and $\mathcal{T}$ are illustrated implicitly.

The conclusions a) - d) of Thm. 3.3 are illustrated in the figure by the sets and/or labels of sets that are pointed to by dashed arrows with the corresponding labels. These labels are placed in the margins of the figure. For example, the c) in the left hand margin labels both a dashed arrow pointing to the set $\mathcal{BR}A$ and a dashed arrow pointing to the label of the set $\mathcal{R}A$. These two sets are shown as disjoint in the figure, illustrating conclusion c) of the theorem. One can see that conclusions e) and f) are also illustrated.

The result

$$x \subseteq y \Rightarrow \mathcal{R}x = x \cap \mathcal{R}y,$$

which follows from (3.1) is illustrated in its three different cases.

1. $\mathcal{R}F$, the part of F in $\mathcal{N}$ equals $F \cap \mathcal{R}A$.
2. $H \subseteq D \in \mathcal{A}$ then $\mathcal{R}H = \varnothing$.
3. $G \subseteq C \in \mathcal{N}$ then $\mathcal{R}G = G \cap C$.

3.3 Semantics

We now develop a model in which *experience* and *consciousness* are taken as primitives. These primitives may be composed of layers. If so, our primitives model the corresponding basic layers, namely what we have knowledge and understanding about through our sensations and perceptions (this last being a Cantor-like statement). When necessary for clarity, the basic layers shall be called *primary experience* and *primary consciousness*, respectively. While we perceive these basic layers, they are essentially ineffable. The higher layers, should they exist, might very well be beyond ineffability. We focus on the basic layers, and we take our primitives to be models of them. Our goal is to specify an illuminating axiomatic system for these primitives. So we may say that as with set theory, we commence with a Cantor-like (naïve) manner and then refine it by means of an axiomatic approach.

We shall characterize a collection of operators called *consciousness operators*, the generic element of which is denoted by $\mathcal{K}$. We take a set x to model a primary experience. Such a set, being a primitive may be viewed as a Platonic object. Then our *Semantic Thesis* is stated as follows.

Consciousness is a result of operators being applied to experience.

We now give the first four axioms of an open (and developing) system that serves to characterize the experience and consciousness primitives. The axioms and their semantic interpretations justify the Semantic Thesis. We begin with the following definition.
Definition: Let x model a primary experience. Then $\mathcal{K}x$ models the *awareness*, an *induced experience*. *Consciousness* is a specific operator $\mathcal{K}$ acting on experience.

The first three axioms along with their semantic interpretations and a name for each are displayed in the following table.

	Axiom	Semantic interpretation of the axiom	Name of Axiom
a)	$\forall x,\ \mathcal{K}x \subseteq x$	Experience generates its own awareness	Generation
b)	$\forall x,\ x \not\subseteq \mathcal{K}x$	Awareness does not generate the primary experience	Irreversibility
c)	$\forall x,\ \mathcal{K}x \not\subseteq x$	Awareness is removed from experience	Removal

Axioms a) and b) are motivated by the properties of the Russell operator a) and b), respectively given in Theorem 3.3.

Note the following analytic statement of axiom c), asserting the normality of awareness.

$$(\mathcal{B}\mathcal{K}x) \cap x = \varnothing.$$

The following table displays algebraic statements of these axioms along with examples of operators that violate each statement. Ω shows that $\mathcal{B}$ and $\mathcal{I}$ violate c).

	Algebraic statement	Violating examples
a)	$\mathcal{K} \cap \mathcal{I} = \mathcal{K}$	$\mathcal{K} = \mathcal{I}$
b)	$\mathcal{B} \cap \mathcal{K} = \mathcal{E}$	$\mathcal{K} = \mathcal{B},\ \mathcal{T}$
c)	$(\mathcal{B}\mathcal{K}) \cap \mathcal{I} = \mathcal{E}$	$\mathcal{K} = \mathcal{E},\ \mathcal{B},\ \mathcal{I}$

We now append a fourth axiom (that in fact is stronger than axiom a)). Note the connection of this axiom to the notion of Selectors in Sect.3.2.3.

d)	If $x \subseteq y$, then $\mathcal{K}x = x \cap \mathcal{K}y$	Awareness of a sub-experience is determined by the sub-experience and awareness of the primary experience	Selection

Axiom d) is motivated by the condition (3.2) of Thm. 3.1.

The consistency of the axioms a) - d) is demonstrated by producing an operator that satisfies all of them. Indeed, $\mathcal{R}$ is such an operator as the following theorem shows.

Theorem: The Russell operator $\mathcal{R}$ satisfies the axioms a), b), c) and d).

Axiom c) implies the following proposition.

Proposition: $\mathcal{KBE} = \mathcal{BE}$.

There are other operators besides $\mathcal{R}$ that satisfy axioms a) - d), as the following example C of a consciousness operator shows.

$$Cx = \{y \in x \,|\, y \notin y;\ \forall z \in y,\ z \neq \Omega\}. \tag{3.5}$$

So since all elements of Cx are normal, Cx is a normal set. Moreover $Cx \subseteq \mathcal{R}x$, so that C is a sub-operator of $\mathcal{R}$. To show that $C \neq \mathcal{R}$ note that for the set $A = \{\{\varnothing, \Omega\}\}$, we have $\mathcal{R}A = A$, but $CA = \varnothing$. To show that C satisfies the axioms, we proceed as follows.

a) By definition $Cx \subseteq x$, so axiom a) is satisfied.
b) Since $Cx \subseteq \mathcal{R}x$ and $x \notin \mathcal{R}x$, then $x \notin Cx$. So axiom b) is satisfied.
c) To prove that C satisfies axiom c), we show the algebraic equivalent to the axiom already noted. Namely that $\mathcal{B}C \cap \mathcal{I} = \mathcal{E}$. Then suppose $\exists z$ such that $Cz \in z$. There are two options.
   1. $Cz \in Cz$. This implies that $Cz \notin Cz$, a contradiction, since by definition every element of Cx is normal.
   2. $Cz \notin Cz$. This implies that either $Cz \in Cz$ or $\Omega \in Cz$. Hence $Cz \notin Cz$ implies $\Omega \in Cz$. However $\Omega \in \Omega$, contradicting the normality of Cz.
d) (3.1) shows that C satisfies axiom d).

Note that the Anti-Russell operator $\mathcal{T}$ introduced in Sect. 3.2 is not a consciousness operator since it violates axiom b). The idempotency, $\mathcal{K}^2 = \mathcal{K}$ follows from axiom d). Generalizing (3.5) yields a collection $\mathcal{K}_k$ of consciousness operators parameterized by a set A.

$$\mathcal{K}_A x = \{y \in x \,|\, y \notin y,\ \mathcal{T}(y \cap A) = \varnothing\}. \tag{3.6}$$

In Sect. 5 we speculate on the connection of qualia to a diagonalization of this operator.

4. Labeling of Graphs, Histogram Construction, M-Z Equation, Neural Networks

We begin with a prescription for labeling a collection (Sect. 3.1). This is extended into a technique for labeling a decorated graph. Given a graph, this procedure forms the basis for inducing existence of a set intrinsically associated with the graph. A histogram construction is then made. The latter is a tool used in proposing the M-Z equation, which expresses the labeling of a decorated graph in terms of sets. Application of these constructs is then made to graphs arising in neural networks. An interpretation is made that portrays the sets decorating a graph as virtual sets, in particular as Platonic constructs, namely *consciousness*. This application and interpretation constitutes a theory of consciousness constructed on the foundations that we have developed.

4.1 Labeling of graphs

Let N be the collection of nodes of a graph Γ and E the edges. A *labeling* λ of Γ is a set valued function of N.

$$a \mapsto \lambda a, \forall a \in N.$$

A *labeled decoration* of Γ is a set valued function $a \mapsto d_\lambda a$, where (compare (2.1))

$$d_\lambda a = \{d_\lambda b \mid a \to b\} \cup \lambda a, \;\; \forall a \in N.$$

This system of equations along with the following theorem shows how labeled decorations are a basis for inducing existence of a set intrinsically associated with the graph. (Compare the notion of the picture of a graph in Sect. 2.2.) Existence and uniqueness of d_λ is the subject of the following theorem.

Theorem 4.1: Given (N, E, λ), a corresponding labeled decoration $d_\lambda a$ exists and is unique. (Aczel, Thm. 1.10.) (Compare with AFA in Sect.2.2.)

Example of a labeled decoration: Take the set Ω with node a. With λa being any set, we have

$$d_\lambda a = \{d_\lambda a\} \cup \lambda a. \tag{4.0}$$

If $\lambda a = \{b\}$, a singleton, then $d_\lambda a = \{d_\lambda a, b\}$. Then $d_\lambda a = b^* = \mathcal{D}b$ is the dual of b.

4.2 The histogram construction

We now give a histogram construct that replaces a set valued function on a collection by a set valued function on a pure set. This construct is used to apply Theorem 4.1 to a special collection of graphs abstracted from brain circuitry to be introduced in Sect. 4.4.

Let A be a collection (of indistinguishable elements), and let B be a set. Consider the mapping, $f : A \to B$, where

$$f^{-1}(b) = \{a \in A \mid f(a) = b\}, \forall b \in B.$$

We suppose that the number of elements in this set $|f^{-1}(b)|, \forall b$ is finite.

The *histogram* H_f of f is the set of pairs specified as follows.

$$H_f = \{(b, |f^{-1}(b)|) b \in B, f^{-1}(b) \neq \varnothing\}.$$

Note that H_f is a bona fide set (see Sect 2.1), and in particular, that $H_f \subseteq B \times \mathbf{N}_+$.

4.3 The M-Z equation, the weight function, the voltage function

We call the set valued function $w : E \to \mathbf{Q}$, a *weight function*. The Rationals $\mathbf{Q}$ comprise a set, since each rational q corresponds to the triple $(m,n,\pm)$, where $\pm m/n$ is the value of q. The choice of the Rationals for the range of w is made for simplicity.

Let E_a denote the set of edges of Γ that terminate in the node a, that is,

$$E_a = \{(p, a) \mid p \to a\}, \forall a \in N.$$

We make the local finiteness hypothesis: $\forall a$, E_a is finite. Then let $w_a = w|_{E_a}$, so that $w_a : E_a \to \mathbf{Q}$ is a function from a finite collection into the Rationals. (Recall that we have no way of distinguishing among the elements of E_a.)

Let H_{w_a} be the histogram of w_a. H_{w_a} is a finite set since E_a is. Note that

$$H_{w_a} \subseteq \mathbf{Q} \times \mathbf{N}_+.$$

Now given (N,E,w), label Γ with the labeling $\lambda : a \mapsto H_{w_a}$. Then the labeled decoration of Γ is specified by what we shall call the M-Z *equation*, namely

$$d_\lambda a = \{d_\lambda b \mid a \to b\} \cup H_{w_a}, \ \forall a \in N. \tag{4.1}$$

Comparing this to (2.1) where a decoration is defined and noting that the set H_{w_a} of labels is arbitrary, we may interpret the set H_{w_a} as a forcing term in the M-Z equation for the decoration $d_\lambda a$.

We shall be interested in an extension of this construction that involves what we call a *voltage function* $v : N \to \{0,1\}$. (The choice of $\{0,1\}$ is made for convenience.) Take

$$E_{a,v} = \{(p, a) \mid p \to a, v(p) = 1\}, \forall a \in N,$$

and let $w_{a,v} = w|_{E_{a,v}}$. Now label Γ with the labeling $\lambda : a \mapsto H_{w_{a,v}}$. Then the M-Z equation that specifies the labeled decoration of Γ (when both weights and voltages are prescribed) is

$$d_\lambda a = \{d_\lambda b \mid a \to b\} \cup H_{w_{a,v}}, \ \forall a \in N. \tag{4.2}$$

4.4 Application to a neural network model of brain circuitry, NN semantic thesis

The brain is commonly taken as the seat of consciousness, the latter supervening on the workings of the brain's neural networks. (While for some, it is the entire physical body that is taken as the seat of consciousness, there is no loss of meaning for our argument to take the more limited view.) We shall show how our model applies to a neural network to produce a labeled decorated graph, this in turn, supplying the mathematical foundation of consciousness we seek in the context of that neural network.

Take a neuron P and trace its inputs (afferents) backward and its outputs (efferents) forward to elaborate a neural network. Replacing the neurons by nodes and the synapses by edges, there results a graph Γ emanating from the node (also called P) corresponding to the chosen neuron. Typically this network has reentrant connections, and so, Γ is non well founded. An illustration of a simple possible Γ is given in Fig. 4.1. (Note the correspondence to the cords and knots of Kanger, 1957.)

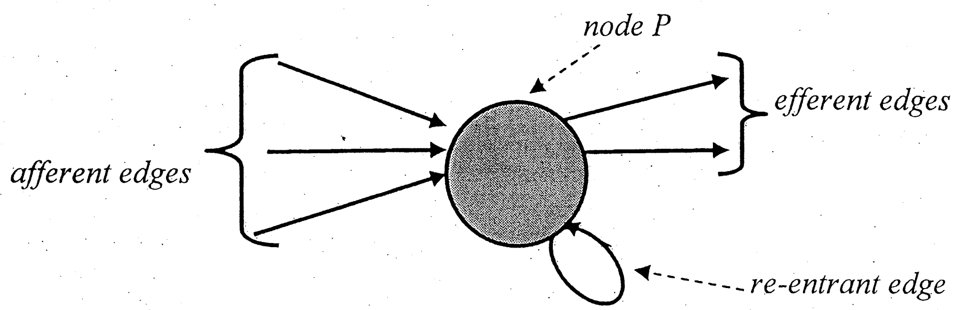

Figure 4.1: A neural net with a single node P interpreted as a graph picturing a set.

This network and so also Γ is associated with two families of parameters, namely, the synaptic weights w of its neurons and the output of its neurons' activities. The latter are expressed as voltages, denoted v. Hebb's rule is the customary model of synaptic weight change. The changes in voltage outputs are modeled by I-O threshold equations, the simplest version of which is the McCulloch-Pitts model (Haykin, 1999). For clarity, we use the simplest meaningful form of these two relations that specify updates of w and v, written as $w_{old}(a \to b) \xrightarrow{update} w_{new}(a \to b)$ and $v_{old}(a) \xrightarrow{update} v_{new}(a)$, respectively.

Hebb's rule:
$$w_{new}(a \to b) - w_{old}(a \to b) = \alpha v_{old}(a) v_{new}(b). \tag{4.3}$$

Here a is an afferent neuron and b a corresponding efferent. $w(a \to b)$ is the synaptic weight of the synapse connecting neuron a to neuron b. (For convenience we allow one such connection per pair of neurons.) $v(a)$ is the neuronal activity of a, that activity customarily modeled as a voltage. For consistency with Sect. 4.3, the scaling constant α is chosen to be a rational number.

McCulloch-Pitts eqn:
$$v_{new}(a) = h\left(\sum_{p:p \to a} w_{old}(p \to a) v_{old}(p) - \theta \right). \tag{4.4}$$

Here h is the Heaviside function, θ is a threshold, and the sum is over all neurons p that forward connect directly to neuron a.

At any instant of time, the dynamical systems (4.3) and (4.4) may be viewed as specifying all the current values in the parameter collections v and w. Referring to the weight and voltage functions of Sect.4.3, we use the v and w to specify a labeling,

$\lambda_{w,v} : a \mapsto H_{w_{a,v}}$ of the graph Γ as described in that section. Then we may use (4.2) to specify a labeled decoration, $d_{\lambda_{w,v}}$ of Γ. We shall also refer to $d_{\lambda_{w,v}}$ as the labeled decoration of the corresponding neural network.

We now state our *Neural Net Semantic Thesis* (see the Semantic Thesis of Sect. 3.3.)

> *Each value of $d_{\lambda_{w,v}}$, a Platonic set, encodes a dynamic preconscious experience associated with the corresponding neuron (equivalently, node of Γ).*

We take the set in $d_{\lambda_{w,v}}$ corresponding to the node P of Γ (namely $d_{\lambda_{w,v}}(P)$) to encode the preconscious experience of the entire graph Γ. The remaining sets in the decoration of Γ, being subsets of $d_{\lambda_{w,v}}$ can be viewed as encodings of subordinate or supporting preconscious experiences. Every neuron generates one or more graphs Γ. As the brain processes information, the weights and voltages change as characterized. When a neuron's efferent voltage, a binary valued variable changes, the graphs containing that neuron gain or lose an edge, as the case may be, and so, these graphs along with their corresponding labeled decorations (sets) undergo changes.

Platonism: The neural networks along with their weights and voltages are physical, that is, they may be observed and measured. The sets corresponding to the labeled decorations $d_{\lambda_{w,v}}$ are not physical and so are unobservable. *Since they are located in some virtual space, we regard the $d_{\lambda_{w,v}}$ as Platonic.* (Compare Schrödinger's quote in Sect. 1.)

If Γ is well founded, its labeled decoration $d_{\lambda_{w,v}}$ can be constructed in a straightforward recursive manner. However while the AFA supplies an existence statement for the decoration of a non well-founded graph, it does not give a method to construct that decoration. The universe of graphs is divisible into two parts, one in which labeled decorations are recursively computable and its compliment. Call the latter $d^{not}_{\lambda_{w,v}}(P)$. The computability of the former might be a reason for classifying these corresponding sets as physical and not Platonic. The non-computability of the $d^{not}_{\lambda_{w,v}}(P)$ reinforces their Platonic status.

4.5 Correspondence of the semantic theses

The relationship among the semantic theses of Sect. 3.3 and Sect. 4.4 is schematized in Figure 4.2. Each of the three arrows in Fig. 4.2 describes a flow of information. The lowest is a flow of physical information. The second is a flow from physical to *psychic* (i.e., to Platonic) information. The highest is a flow of psychic information. By the *Neuro-physiological Thesis* in Fig. 4.2, we shall mean the conventional movement of sensory information from a sense organ to the brain where it is processed to frame an internal physical representation of that information, and from where, according to our theory, a primitive called consciousness is called into existence in a virtual space. Note a parallel between the information flow in Fig. 4.2 with Plato's *line of knowledge* (Plato, 360 BCE).

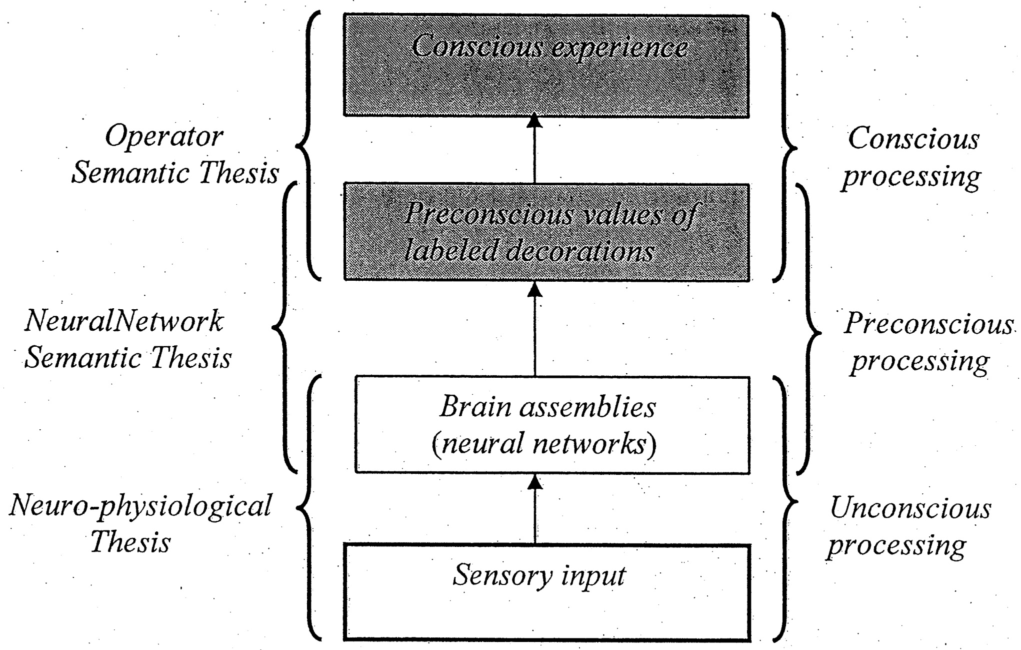

Figure 4.2: The theses of consciousness. The shading demarks the Platonic realm.

Section 5. Speculations: syntactic and semantic nomenclature, qualia, evolution

Further elaboration of the consciousness operators developed will provide additional applications of the theory presented here. The basic operators of set theory introduced will contribute to this elaboration. This is suggested by the syntactic and semantic nomenclature ascribed to these operators, summarized in the Table 5.1. Also shown in this table is an interpretation of each operator along with the Z-F axiom(s) that the operator codifies.

Op	Syntactic	Semantic	Interpretation	Axiom(s)
$\mathcal{E}$	*Elimination*	*Erasing/Forgetting*	*Set representing Platonic experience erased*	*Existence of* $\varnothing$
$\mathcal{I}$	*Identity*	*Accepting/Receiving*	*Leaves set unchanged*	*Extension*
$\mathcal{B}$	*Brace*	*Conceiving*	*Creates higher order set (a singleton) out of a set*	*Pair and singleton*
$\mathcal{R}$	*Russell*	*Perceiving*	*Bifurcates set contents & retains normal elements*	*Comprehension*
$\mathcal{T}$	*Anti-Russell*	*Rejecting/Denying*	*Counters* $\mathcal{R}$, *retaining the abnormal elements*	*Union*
$\mathcal{D}$	*Duality*	*Reinforcing/Elaborating*	*Extension of conceiving*	*AFA*

Table 5.1: Semantic interpretations of basic operators

In Fig. 5.2 we schematize the flow of information from sensory input to conscious experience. The upper boxes describe the syntactic level, the lower the semantic. The $d^{not}_{\lambda_w,v}(P)$ correspond to neural networks in the brain. They are schematized in the box labeled 'collection of virtual sets d_λ' in Fig. 5.2. Is it a time dependent one of these that emerges into consciousness? If so, how is this distinguished neural network selected?

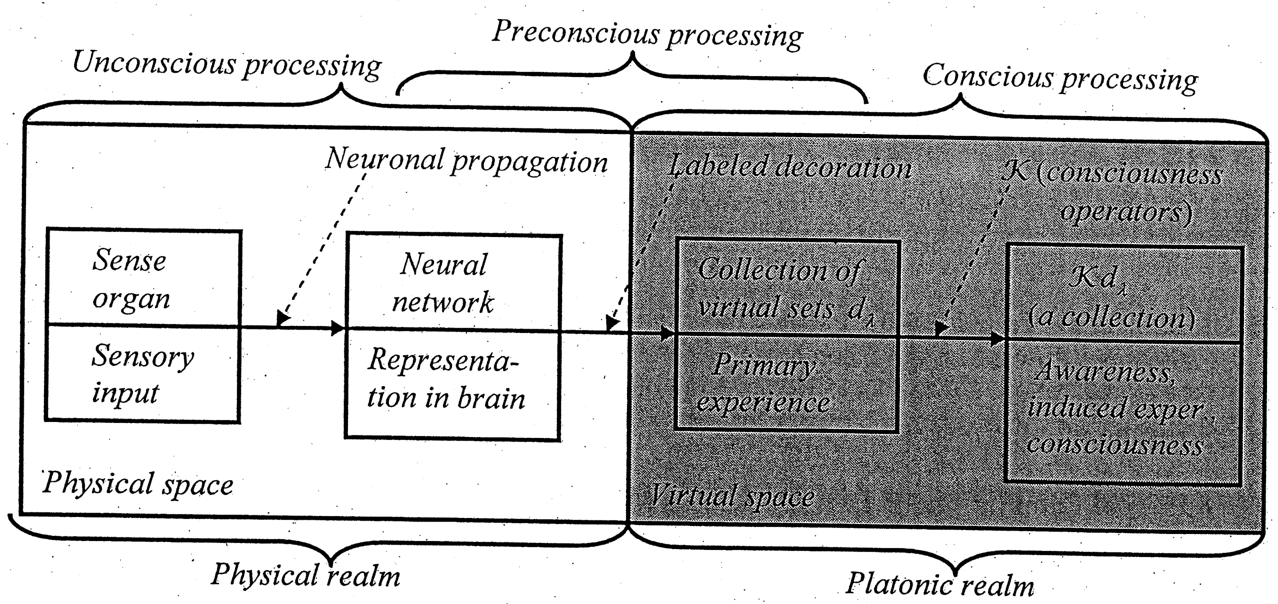

Figure 5.2: Consciousness: Syntactic and semantic views of the processing from the physical to the virtual. Shading distinguishes the Platonic realm from the physical.

Diagonalization of $\mathcal{K}_A$, qualia: Diagonalization of the operator $\mathcal{K}_A$ in (3.5) gives

$$\mathcal{K}_x^{diag}x = \{y \in x \mid y \notin y, \text{ and } \forall z \in x \cap y, z \notin z\}.$$

$\mathcal{K}_x^{diag}$ satisfies axioms a) – c) of Sect. 3.2. However taking $A = \{\{\varnothing, \Omega\}\}$ and $B = \{\varnothing, \Omega, \{\varnothing, \Omega\}\}$, it follows that $\mathcal{K}_A^{diag}A$ is not a subset of $\mathcal{K}_B^{diag}B$. So failing axiom d) precludes $\mathcal{K}_x^{diag}$ from being a consciousness operator. We expect this operator to be a member of an operator collection of interest. For instance, take the set $d_\lambda a$ specified in (4.1) and put a equal to p, the point of a graph corresponding to a neural network. If this neural network is the neural correlate of a quale, we ascrive the semantics of that quale to the Platonic set $\mathcal{K}_{d_\lambda p}^{diag}d_\lambda p$, itself located in a virtual space. This quale is positioned in the rightmost box in Fig. 5.2.

Evolution: We expect that variations of our development will provide mathematical foundations for the study of evolution driven by *selfish replicators*, both genetic and mimetic (Dawkins, 1979, Blackmore, 1999).

Appendix: Axioms of Set Theory

We use the following axioms of set theory.

Existence: $\exists z$.

Extensionality: $\forall z(z \in a \leftrightarrow z \in b) \rightarrow a = b$.

Pairing: $\exists z[a \in z \,\&\, b \in z]$.

Union: $\exists z(\forall x \in a)(\forall y \in x)(y \in z)$.

Comprehension: $\exists z \forall x[x \in z \leftrightarrow x \in a \,\&\, \varphi(x)]$.

Here φ can be any formula in which the variable z does not occur free.

Except for the axiom of existence these axioms along with the Axioms of Infinity, Collection, Power Set and Choice can be found in Aczel (1988). We do not state the latter four axioms since we don't use them. Note that Aczel uses the name Axiom of Separation for the Axiom of Comprehension. The

Axiom of Foundation: $\exists x(x \in a) \rightarrow (\exists x \in a)(\forall y \in x)\neg(y \in a)$.

is not included in the original Z-F list. It was proposed by Von Neumann. We don't use the AF, and we replace it by the

Anti-Foundation Axiom: *Every graph has a unique decoration.*

The AFA, due to Aczel, is central to our development.

Glossary[1]

Terminology

Experience/primary experience...a set *x*/primary layer when there are layers of experience
Consciousness...$\mathcal{K}x$, where $\mathcal{K}$ is a consciousness operator. See Semantic Thesis in §3.2
Awareness....$\mathcal{K}x$, where $\mathcal{K}$ is a consciousness operator. See Semantic Thesis in §3.2
Graph....a collection of nodes with certain pairs of the nodes specified as edges
Directed graph....a graph in which the nodal pairs are ordered (edges are directed)
Pointed graph....a directed graph with a distinguished node, the point
Accessible pointed graph (apg)....a pointed graph, every node of which is reachable from the point by a chain of directed edges
Decoration....the unique assignment (specified by (2.2)) of sets to the nodes of an apg

[1] For convenience, some of the definitions listed here are abbreviated. In such cases more complete definitions are found in the text.

Picture of a set....the pointed graph in whose decoration, the set corresponds to the point
Labeled graph....a graph with an arbitrary assignment of sets (the labels) to the nodes
Labeled decoration....a labeling dependent decoration of a graph (specified by (4.0))
Histogram.....construct replacing a collection by a set as domain of a set valued function
M-Z equation....specifies the labeled decoration of a graph arising from neural networks
Hebb's rule....specifies the synaptic weight change in a model neuron
McCulloch-Pitts equation....specifies the output of a model neuron

Set types
Collection....a set as defined by Cantor
Naïve set....another name for a collection
Set....a primitive construct, the subject of the Z-F axioms
Bona fide set....a set, emphasizing its being specified as a primitive defined by Z-F
Pure set....a set whose elements are sets
Path....a sequence of nodes (finite or infinite) linked by edges
Well-founded picture....a graph whose paths are finite (in particular one without loops)
Non well-founded picture....a graph with an infinite path
Well-founded set.... a set whose picture is well-founded
Non well-founded set....a set whose picture is non well founded
Normal set....a set whose elements do not contain themselves
Abnormal set....a set with an element that contains itself
Platonic set....a not physical set, a not computable set, a set located in a virtual space

Classes
Class....a collection of sets with a common property
Proper class....a class that is not a set
$\mathcal{U}$....the universe of sets
$\mathcal{A}$....the class of abnormal sets
$\mathcal{N}$....the class of normal sets

Fundamental Operators
$\mathcal{E}$....elimination
$\mathcal{I}$....identity
$\mathcal{B}$....brace, singleton
$\mathcal{R}$....Russell
$\mathcal{T}$....anti-Russell
$\mathcal{D}$....duality operator
$\mathcal{C}$....a particular consciousness operator

Types of Operators
O....a generic operator
$\mathcal{K}$....a generic consciousness operator
$\mathcal{K}_{\text{\textlambda}}$....a special class of consciousness operators parameterized by a set A
$\mathcal{K}_x^{diag}x$....diagonalization of $\mathcal{K}_{\text{\textlambda}}$ acting on a set x
Selectors....operators O with the following property: $x \subseteq y \Rightarrow Ox = x \cap Oy$

References

Aczel, P. 1988. *Non-Well Founded Sets*. CSLI Publications.
Aleksander, I., Dunmall, B. 2003. Axioms and Tests for the Presence of Minimal Consciousness in Agents. In *Machine Consciousness*, O. Holland, ed., Imprint Academic
Aristotle. 1961. (Ross, D. Ed). *De Anima*, Oxford at the Clarendon Press.
Barwise. J. in the forward to Aczel 1988.
Bernays, P. 1954. A System of Axiomatic Set Theory, VII. *J. Symbolic Logic* 1981-86.
Blackmore, S. 1999. *The Meme Machine,* Oxford Univ. Press.
Cantor, G. 1895. Beiträge zur Begründung der transfiniten Mengenlehre, 1. *Mathematische Annalen* 46;481-512.
Courant, R., Robbins, H. 1941. *What is Mathematics.*
Dawkins, R. 1976. *The Selfish Gene.* Oxford University.
Descartes, R. 1637. *Discours sur la Methode.*
Dym, H, Mckean, H. 1972. *Fourier Series and Inteegrals.* Academic Press.
Fränkel, A. 1922. Zu den Grundlagender Cantor-Zermeloschen Mengenlehre. *Mathematische Annalen 86*;230-237.
Frege, G. 1893. *Grundsetze der Arithmetik, begriffsschriftlich abgeleitet,* Vol. 1. Jena Volume 2 published in 1903.
Gödel, K. see Jech
Haykin, S. 1999. *Neural Networks a Comprehensive Foundation.* Prentice Hall.
Hilbert, D. 1900, Address before the International Congress of Mathematicians in Paris. See also Kaplansky. 1977. *Hilbert's Problems.* University of Chicago.
Hofstadter, D. 1979. *Gödel Escher Bach; An Eternal Golden Braid.*
Hrbacek, K., 1999. *Introduction to Set Theory.* M. Dekker.
Jech, T. 2002. *Set Theory*. Springer.
Kanger, S. 1957. *Provability in Logic.* Univ. of Stockholm: Almquist and Wiksell. Stockholm Studies in Philosophy.
Penrose, R. 1989. *The Emperor's New Mind.* Oxford Univ. Press.
Plato. 360 BCE. *The Republic.* Translated by B. Jowett.
Russell, B. Whitehead, A., 1910-1913. *Philosophiae Naturalis Principia Mathematica.*
Schrödinger, E. 1958. *Mind and Matter*. Cambridge Univ. Press
Solms, M. 1995. Chromosomes on the Couch. *Psychoanalytic Psychotherapy*, 9(2); 107-120.
Von Neumann, 1925. Eine Axiomatisierung der Mengenlehre. Jurnal für reine und angewandte Mathematik 155:219-240.
Whitehead, A. see Russell.
Zermelo, E. 1908. Untersuchung uber die Grundlagen der Mengenlehre, I. *Mathematische Annalen* 65;261-281.