Consciousness: A Logical Model

J. Andrew Ross *
a.ross@sap.com

Abstract

Consciousness is not personal but subjective. The subject structures an input stream of qualia into a dynamic unity. The unity synthesizes an evolving series of centered virtual worlds that represent pairwise contradictory epistemic and ontic states. Each world is symmetric relative to alternative possible successors. The proposed model is platform independent and can support personal identity.

Keywords

Consciousness – Philosophical logic – Philosophical psychology – The self

Introduction

Before we can build theoretical foundations for a science of consciousness, we need a general model for the concept of consciousness. The preliminary model suggested by common sense is a useful place to start, but it leads quickly to the problem of deciding how much we can reasonably presuppose. For example, consider the model David Chalmers has in mind in this quotation:

The job of a science of consciousness, then, is to connect first-person data to third-person data; perhaps to explain the former in terms of the latter, or at least to come up with systematic theoretical connections between the two.[1]

* To all the friends and colleagues who helped to keep the ideas expressed here alive during the many years they took to make sense, I extend heartfelt thanks.

[1] From "First-person methods in the science of consciousness" by David Chalmers, *Consciousness Bulletin*, The University of Arizona, Fall 1999. The sentence is highlit in the original.

That model presupposes an understanding of the contrast between first-person and third-person data. However, we may well argue that it is one of the tasks of a science of consciousness to explain that contrast, since we may hope to understand the concept of a person partly in terms of a person's ability to serve as a center of consciousness. Therefore, it is worth going deeper.

Subject and object

To avoid presupposing the concept of a person, we can start by using instead the more general concept of a subject. The opposite of a subject is an object. In principle, subjects and objects come in pairs: to every subject, there is an equal and opposite object, and vice versa. For practical purposes, of course, the vast majority of such pairs are of rather one-sided interest.

Consciousness is a relation between subjects and objects: a given subject is conscious of a domain of objects. Only a very small fraction of all subjects and objects instantiate this relationship, and for those that do instantiate it, a more detailed characterization of how they do so can presumably take a variety of forms. But it is essential to the relationship that one subject be conscious of an unspecified number of objects. Also, a conscious subject persists in time. In general, the objects of consciousness may relate to the subject either serially or simultaneously, or both.

The objects of consciousness may be spatiotemporal objects with detailed properties and extensive relations with other objects or they may be degenerate items of immediate experience. Let such items of immediate experience be called *qualia*. Consciousness as humans experience it is always, by definition, consciousness of qualia, although it is hardly ever so raw and unstructured that it is experienced merely *as* consciousness of qualia. It is difficult to give a satisfactory account of the general conditions under which qualia can be further characterized in terms of spatiotemporal location and involvement in complex objects. David Chalmers calls this problem of giving an account of qualia in the "third-person" terms of science the "hard problem" of consciousness.[2]

The phenomenology of consciousness is the logical study of how reality seems to a conscious subject. Reality seems like a changing manifold of qualia. The task of phenomenology is to take this initial characterization far enough to connect with the data and laws of normal science. The normal science of human

[2] David Chalmers has made the hard problem famous. He developed it at length in his big book [CHALMERS 1996] and discussed it with numerous critics in the anthology [SHEAR 1997]. Under different guises, the problem has a long history in philosophy, going back at least to Descartes.

consciousness is modern psychology together with the collection of disciplines known loosely as the brain sciences.[3]

But consciousness is not necessarily human consciousness. The assimilation of the phenomenology of consciousness to the brain sciences presupposes the reductionist premise that consciousness as we understand it is a property or a product of appropriately functioning cerebral tissue. It is not unreasonable to anticipate that this will be an outcome of a future science of consciousness, but it is unreasonable to stipulate it at the outset as an axiom. It is analogous to stipulating a thousand years ago that the task of a future science of astronomy is to explain how the heavenly bodies orbit the Earth.

A future science of consciousness may be expected to explain:

- Personal consciousness. Each and every normally functioning human being is conscious, regularly and routinely. This is the gross fact that any science of consciousness must explain.
- Interpersonal consciousness. Before consciousness is ignited in an organism, the organism may require some special kind of socially mediated personal interaction. Consciousness may be a phenomenon manifested in a society of reciprocating organisms but impose only basic requirements on the cerebral architecture of those organisms. If some kind of interaction is essential, a science of consciousness must explain the societal prerequisites.
- Transpersonal consciousness. It is conceivable that human consciousness is able to transcend its personal bounds and experience other lives or oceanic states. It is certain that human consciousness can *seem* to do these things.
- Impersonal consciousness. Subjective consciousness of a domain of objects may be possible independently of persons. Animals lacking the concept of a person may be conscious. Machine consciousness may be developed without recognizable personality.

This paper outlines a logical model of consciousness that is sufficiently general to accommodate at least three of these items.

---

[3] Psychology and the brain sciences enjoyed a big boost in the 1990s (dubbed the "decade of the brain") and now face newcomers with a daunting mountain of required reading. For example, [CHURCHLAND 1986] is a modern classic, [CLARK 1997, COTTERILL 1998] emphasize cognitive science, [CRICK 1994] is by the DNA Nobelist, [DAMASIO 1999, RAMACHANDRAN 1998] emphasize clinical aspects, [EDELMAN 1992] is by the neuroscience Nobelist, [GAZZANIGA 1998] is a handsome course text, [GREENFIELD 1995, SCOTT 1995] are elegant monographs, [METZINGER 1995, ROSE 1998] are reliable anthologies, and [PINKER 1997] is a long but light overview.

Truth unfolds

Subjective consciousness of a domain of objects can be seen to a first approximation as analogous to optical reflection. Images that represent objects in some way are juxtaposed within the unified scene reflected by the subject. The optical medium within which the reflection occurs allows information about the objects to be transmitted by light beams. The information transmitted is limited to those properties that affect the light beams.

In general, we can regard the information about the objects that surfaces in consciousness as giving a logical characterization of those objects. A logical characterization need not be based on optical images. A logical characterization is based on information that we can specify consistently and perhaps completely by means of a suitably defined formal language.

A linguistic specification of the relation between subject and object allows that the conscious reflection of various spatial parts or temporal phases of the objective domain can be true or false. The *epistemic state* of the subject can be specified in terms of a set of statements of the formal language. Let us call the state of the objective domain confronting the subject the *ontic state*. The ontic state may either match or not match the epistemic state of the subject. If and when they match, the epistemic statements that claim to specify the ontic state are true. Where the states do not match, the statements are false.

This soon leads to the elementary logic of the propositional calculus (PC) (Box 1). Propositions that express the epistemic state of a conscious subject are either true or false, depending on how that epistemic state compares with the ontic state confronting the subject. If we use the same formal language to describe both the epistemic and the ontic state, and if that language is a fragment of English, then we can characterize truth using Tarskian theorems of the form:

"Snow is white" is true if and only if snow is white.

Such true propositions express facts. Wittgenstein developed an outline theory of facts in his earlier philosophy.[4]

The ontic state confronting a conscious subject may be called a *world*, where a world is a totality of facts. The epistemic state of a conscious subject may also be represented as a world. Assuming that the subject is in a consistent state, such a world must be *possible.* By contrast, the ontic state confronting the

[4] Wittgenstein expounded his earlier philosophy in [WITTGENSTEIN 1922], a brief and difficult text based on the logical philosophy of Frege and Russell. Later philosophers substantially enriched that foundation using ideas in semantics due to Alfred Tarski and others, until in the late 20th century it became mainstream (Anglo-American) technical philosophy.

subject may be called the *actual* world. The epistemically possible world that appears as a reflected image in the subject may or may not be isomorphic to the ontically actual world that is being reflected. If they are isomorphic, then everything believed by the subject is true, but in general they will differ.

The linguistic characterization of the actual world is hardly ever perfect, for obvious reasons that defeat any language we can devise. But the linguistic characterization of an epistemically possible world is perfect, by definition. Thus possible worlds are as fundamentally different from the actual world as rational numbers are from real numbers. A diagonal argument shows that the real numbers outrun the rational numbers,[5] and a similar argument, based on a recursion over the sentences of a PC language, shows that the actual world can differ from any epistemically possible world.

Truth and falsity depend on meaning. The sentences of a language can only be classified as bivalent (that is, determinately either true or false) when their meaning has been sufficiently clarified. Their meaning can then be specified in terms of their truth conditions using Tarskian theorems of the form:

"Snow is white" means that snow is white.

Meaning can only be spelled out in detail in terms of patterns of usage in the relevant speech communities. Wittgenstein described this anthropological view of meaning in his later philosophy.[6]

The ongoing pursuit of science generates new meanings for old sentences as well as new concepts and new sentences. Meanings change and unfold, and truth assignments grow with them. Any language used in earnest to describe the epistemic course of a conscious subject must find its evolution reflected in continuing extensions and revisions of the sets of epistemic and ontic states that provide it with a semantic foundation.[7]

---

[5] Cantor's diagonal argument to prove that the infinity of reals has a higher cardinality than the infinity of rationals is the basis for transfinite set theory. For a readable introduction to the higher infinities, see [RUCKER 1982].

[6] Wittgenstein expounded his later philosophy most clearly in [WITTGENSTEIN 1958]. His anthropological view of meaning radically transformed his earlier views and helped clear the way for Noam Chomsky and others to develop a science of linguistics. Modern linguistics is introduced entertainingly in [PINKER 1994].

[7] Before the focus shifted to language, Karl Popper stressed the importance of repeated cycles of corroboration and falsification in an evolutionary epistemology that he refined over several decades [POPPER 1972]. His views were defended and developed by the contributors to [LAKATOS 1979] in response to Thomas Kuhn's sociological view of normal science and paradigm shifts [KUHN 1971].

In general, the continuing experience of a conscious subject can be described as an ongoing dialectic of epistemology and ontology, where the confrontation of each new epistemic state with the actual world generates falsehoods that are corrected in the next epistemic state. Alternatively, following Daniel Dennett,[8] the conscious subject maintains an ongoing narrative about its role in the world, and this narrative goes through multiple drafts as new experience prompts revisions and reappraisals. In general, truth unfolds in the actual world, and an epistemic subject must keep moving to track it.

Things change

Propositions are linguistic items that can be decomposed into subjects and predicates. Thus analyzed, propositions say of objects that they fall under concepts. They may say that individual objects have certain properties, that sets of objects have certain properties, or that various objects stand in various relations to each other. In each case, the propositions express a movement from an *initial state* to a *final state*. In the initial state, certain existing objects are simply denoted. In the final state, the objects are further specified as having the properties or standing in the relations asserted by the proposition. This movement between epistemic states is what makes a proposition informative.

Typically, a conscious subject whose successive epistemic states are represented by sets of informative propositions continually reidentifies many of the same old objects. New things are said about those objects, and old falsehoods are corrected. But for any objects, certain properties are more essential than others. The essential properties are needed to ensure success in denoting those objects. We can distinguish *names* from *definite descriptions.* Names are rigid designators that continue to track changing objects through modifications of their more essential properties, whereas definite descriptions do not, and instead denote whatever happens to satisfy their descriptive predicates.[9]

The use of names to ensure successful denotation does not obviate the need for objects to have essential or criterial properties, but it does enable us to be

---

[8] Daniel Dennett "explains" consciousness with a *pandemonium* model of the mind in which our conscious states result from political upheavals within a society of cognitive demons that compete for key roles in the ongoing drama of our lives [DENNETT 1991].

[9] Saul Kripke reinvigorated the discussion of names and definite descriptions with his lectures [KRIPKE 1980], where he described names as rigid designators. Bertrand Russell analyzed definite descriptions in 1905 thus: "The big bag is full" means "There is an x such that x is a big bag, and for all y such that y is a big bag, $y = x$ and y is full." For a fuller discussion of this and related issues, see [DUMMETT 1973].

more relaxed about their criteriality. Smooth changes can be tracked even when they result over time in outright contradictions compared with earlier situations. The history of science reveals many such contradictions. Scientists can agree on what objects they are talking about even when they disagree on what to say about them.[10]

Denotation can succeed despite changes in the criterial properties of objects. Denotation does succeed routinely as object accumulate determinacy in the continuing course of epistemic advance. For example, as time passes and new facts come into existence, most objects feature as denotees in larger and larger sets of informative statements.

If things change in this way, complicated propositions about them need to be handled with care. For example, quantified propositions can only be given determinate truth conditions when the domains over which the quantifiers range are specified exactly. And quantifiers can be hidden in the semantic foundations of simple and apparently unquantified propositions, such as definite descriptions. For this reason, when parsing any propositions that purport to state facts, it is wise to relativize explicitly any quantifiers involved to definite ontic or epistemic states, or at least to impose definite limits on which states may be invoked for those propositions.

The quantificational calculus (QC) extends PC by admitting quantification over a domain of objects (Box 2). The objects need not all have names in the language. If all the objects in the domain can be named, then PC propositions can be obtained from QC propositions by replacing universal and existential quantifiers thus:

"For all x, $F(x)$" $\Rightarrow$ "$F(a)$ and $F(b)$ and $F(c)$ and …"

"For some x, $F(x)$" $\Rightarrow$ "$F(a)$ or $F(b)$ or $F(c)$ or …"

Here a, b, c, … are the names of all the objects in the domain of quantification and $F(\ )$ is a predicate. Infinite domains of objects cannot all be named in finite languages, or uncountable domains in countable languages, so we need QC in mathematics.

---

[10] In any active field of science, scientists freely go different ways without sacrificing the denotation of their terms. For example, in modern cosmology, Alan Guth proposes that the early universe went through a brief period of exponential inflation powered by vacuum decay and repulsive gravity [GUTH 1997], Lee Smolin suggests that our universe was borne from a black hole and is tuned to reproduce them [SMOLIN 1997], and Brian Greene reports that our universe may have 11 dimensions with a randomly fluctuating topology at the Planck scale [GREENE 1999]. It seems unlikely that all three stories are true (although it is conceivable), yet they all clearly denote the same universe.

Any intelligent subject is likely to perform computations that involve serious mathematics. Most mathematics that remains less powerful than the arithmetic of natural numbers, for example, the sort of finite math that a pocket calculator can handle, can be represented as tautologies in PC or QC. However, the formal theory of arithmetic (AT) goes beyond pure logic, as it involves postulating the existence of infinitely many natural numbers (Box 3).[11]

The dialectical picture of ontico-epistemic advance presented so far presupposes that each epistemic state is internally consistent. If a state is not consistent, then simple PC computations inside that state can generate utter confusion. However, the consistency of an epistemic state that is closed under AT computations cannot be not guaranteed unconditionally. Such a state runs a small but nonzero risk that a contradiction can be generated by apparently valid computation from evidently true premises.

Gödel's incompleteness theorem for AT illustrates the risk.[12] The formal metatheory of AT can be expressed in a simple language based on QC. Gödel coded all the sentences of this language into the natural numbers. The natural numbers form the domain over which the theorems of AT are interpreted, so the Gödel coding allows the sentences of the language of AT to be interpreted as metatheoretic statements about AT. For each sentence s of the metatheory, let s be coded into the Gödel number $G(s)$. Gödel constructed an open AT sentence g with this interpretation in the metatheory:

The sentence with Gödel number x is not a theorem of AT.

Now consider the closed sentence g^* obtained from g by substituting $G(g)$ for the free variable x. As interpreted in the metatheory, sentence g^* says of itself that it is not a theorem of AT. If g^* is a theorem of AT, then AT is inconsistent. If g^* is not a theorem of AT, then AT is incomplete, since there are truths in the language of AT that are not provable.

Gödel's second incompleteness theorem (based on the first) states that if AT is consistent, then the consistency of AT is not provable in AT, but only in

---

[11] Formalization can seem redundant in arithmetic. But it is necessary in such fields as transfinite set theory. Even in discrete math (where uncountable infinities are banished and constructive methods prevail), formal methods track levels of constructivity and enable us to write programs to automate proofs [GRIES 1993].

[12] Gödel's incompleteness theorem, first published in 1931, is presented formally in numerous texts, for example [MENDELSON 1979]. It requires that AT be ω-consistent. In 1936, J. B. Rosser elaborated the theorem slightly to require only that AT be consistent. Douglas Hofstadter discusses Gödel's theorem in the context of a fascinating meditation on art, music, brains, and computers in [HOFSTADTER 1979].

another theory, say ET, that extends AT. But the extended theory ET stands more in need of a consistency proof than AT, so nothing is gained.[13]

Gödel's theorems show that there is a residual risk involved in allowing a computing subject to use the full power of AT to manipulate the propositions that constitute an epistemic state. An ontic domain reflecting such a state may satisfy sentences like *g**, yet such sentences may still count as falsehoods in the corresponding epistemic state (due to their unprovability), and hence generate contradictions for that subject. The subject can always move into new states by accepting sentences like *g** as axioms in extended theories ET, but then the argument can be applied again to ET. In this way, a series of theories ET can be used to generate an epistemo-ontic dialectic.

Gödel's theorems show that in mathematics, truth outruns provability. Alan Turing extended the argument to show that, very roughly, computability is not a computable concept. More exactly, the set of programs that run for a finite time to produce a definite output cannot itself be defined by such a program.[14] More generally, truth outruns computability and ontology outruns epistemology.

Sets cohere

The picture developed here of an open-ended series of logically defined states invites set-theoretic treatment. Subjects and objects can always be represented as sets. The relation between a subject and its objects can be represented (arbitrarily) as the membership relation, so that the sets representing the objects of consciousness are members of the set representing the conscious subject. Successive states of consciousness can be represented as a succession of sets, with some relation between them that we can seek to specify. Successive momentary determinations of given objects can be represented by successive sets, again with some relation to be specified between them. Ontic states and epistemic states can be represented by sets, and whole dialectics of such states can be represented by infinite series of the corresponding sets.

---

[13] Gödel's second theorem is also presented in [MENDELSON 1979]. Gregory Chaitin takes the theme of undecidability in arithmetic further in the context of his algorithmic information theory: by his algorithmic definition of randomness, some truths in arithmetic are random [CHAITIN 1998].

[14] Alan Turing made numerous fundamental contributions to computability theory. The best known may be his invention of the Turing machine, which is an idealized computer. Turing computability proves to be equivalent to several other kinds of computability, so we have good reason to believe that computability is a basic mathematical concept. See [BOOLOS 1980].

The strategy of using sets to represent all the entities comprehended in a theory has the merit that mathematical machinery then becomes available to clarify and extend the theory.[15]

The set-theoretic structure that seems suitable for this task is the cumulative hierarchy of pure well-founded sets, which is the natural or intended model of standard systems of axiomatic set theory such as Zermelo–Fraenkel (ZF) set theory (Box 4).[16] This hierarchy is a mathematical structure analogous to the natural numbers, but much richer and therefore more useful in logic. Just as arithmetic is logic plus the assumption that the natural numbers exist (as defined in the Peano axioms), so set theory is logic plus some assumption about which sets exist (typically formulated as axioms).

Mathematicians developed the cumulative hierarchy as a reaction to the antinomies in naïve set theory discovered by Bertrand Russell and others.[17] The pioneering set theory formalized by Frege allowed sets to be formed as the extensions of any well-defined predicates. Making use of Frege's formal syntax, Russell defined:

$R = \{x \mid x \notin x\}$ (the class of sets x such that x is not a member of x)

The variable x here can stand for any set. If R is a set, then R is a member of R if and only if R is not a member of R. This contradiction shows that we need to restrict the definition of admissible predicates rather carefully.

Ernst Zermelo and others reacted by starting small and building up step by step. Sets were not allowed to be members of themselves and were always constructed from members that had been built earlier; that is, sets were said to be well founded. In pure set theory, we start at stage zero with just the empty set $\varnothing = \{\ \}$. At stage 1, we use just the sets from stage 0, namely $\varnothing$ alone, as elements to form all the sets we can, namely just $\varnothing$ again and its singleton $\{\varnothing\}$, so the universe V_1 of sets at stage 1 is the class $\{\varnothing, \{\varnothing\}\}$. At stage 2, we find that $V_2 = \{\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\varnothing, \{\varnothing\}\}\}$. Generally, at each finite step n in the building process, the nth determination V_n of the universe V has as members all and only the subsets of the previous determination V_{n-1} of V (including $\varnothing$ and

---

[15] Van Quine built a general philosophy [QUINE 1960] on the idea that with sufficient prestidigitation we can reduce any ontology to sets. The idea works well in mathematics, at least [QUINE 1969].

[16] Zermelo–Fraenkel set theory is presented more fully in numerous texts, such as [JECH 1997, MENDELSON 1979, QUINE 1969].

[17] The philosophy of set theory, its history, and its mathematical motivations are discussed in [BENACERRAF 1964, FRAENKEL 1973, QUINE 1969].

V_{n-1} itself).[18] That is, for each n, determination V_{n+1} is the power set $\wp(V_n)$ of the previous determination V_n:

$V_0 = \varnothing = \{\ \}$

$V_{n+1} = \wp(V_n) = \{x \mid x \subseteq V_n\}$

John von Neumann defined a transfinite function V_α that extends this definition to build up the cumulative hierarchy. He represented ordinal numbers α in the hierarchy as sets $\{\beta \mid \beta < \alpha\}$ of all their predecessors:

$\alpha_0 = \varnothing$

$\alpha_+ = \{\beta \mid \beta \leq \alpha\}$ for the successor α_+ of any given ordinal α

$\alpha_\lambda = \bigcup_{i<\lambda} \alpha_i$ at limit stages λ (α_λ is the union of all ordinals less than α_λ)

The von Neumann function for V is defined by transfinite recursion as follows:

$V_0 = \varnothing$

$V_{\alpha+} = \wp(V_\alpha)$ for the successor α_+ of α ($V_{\alpha+}$ is the power set of V_α)

$V_\lambda = \bigcup_{\alpha<\lambda} V_\alpha$ at limit stages λ (V_λ is the union of all sets V_α, for $\alpha < \lambda$)

For any set x, the first V-set V_α in which x appears as a subset gives the *rank* α of x. For example, each ordinal α has rank α.

The axioms of finististic set theory HF assert the existence of sets with finite ranks $n < \omega$, where ω is the first infinite ordinal. The sets comprehended in HF are the *hereditarily finite* sets, which have a finite number of elements, each of which has a finite number of elements, and so on down to $\varnothing$. The natural model of HF is the set V_ω (the first infinite V-set).

The axioms of ZF set theory assert the existence of all sets in the cumulative hierarchy up to the first *inaccessible* ordinal θ, defined as the first ordinal that cannot be reached using just the ZF axioms. The largest ordinals comprehended in ZF are given by Fraenkel's replacement schema, which asserts that for any function F that is definable in the language of ZF, if the domain of F is a set, then the codomain of F is also a set. The natural model of ZF is the set V_θ (the first inaccessible V-set).

Numerous other axioms have been proposed, asserting the existence of sets up to (apparently) higher ordinal or cardinal ceilings.

---

[18] The notion of comprehending all subsets of given sets is not as innocent as it seems. For all infinite sets x, there are uncountably many subsets of x. For any set x, the power set of x has a cardinality greater than the cardinality of x. Cantor's continuum hypothesis is that for countable x (with cardinality $\aleph_0$), the power set of x has the first uncountable cardinality ($\aleph_1$). For details, see any set theory text.

The "thinnest" universe that satisfies the axioms of ZF and related theories is Gödel's *constructible* universe L obtained by comprehending in power sets $\wp(V_\alpha)$ only those subsets of V_α that can be constructed by an explicit recursion from given sets.[19]

The consistency of any theory that comprehends sets of transfinite rank α is increasingly doubtful as α increases. A natural picture is that somewhere in the transfinite hierarchy, the theory simply becomes incoherent. How much of the hierarchy can be described coherently is ultimately a philosophical question. A constructivist answer here is that it depends on how well we have built the syntactic apparatus (notation, proof procedures, and so on).

Each set has two sides. Seen from above, it is an *element* that can be used as a member of further sets. Seen from below, it is a *class*, namely the class of its members. A useful notational extra in von Neumann–Bernays–Gödel (NBG) set theory is to introduce new (uppercase) variables X for classes and deny their equivalence to set variables x by refusing to allow class variables before the membership symbol (so "$x \in X$" is allowed but "$X \in x$" is not).[20] Thus we can discuss classes without assuming they exist as elements. Some classes are then *proper* classes, which means they cannot be consistently regarded as sets. For example, the universe V of sets is a proper class, distinct from its momentary determinations V_α, since otherwise V would be a member of itself, contradicting the requirement that all sets be well founded.

At each stage α in the determination of V, the elements comprehended at that stage are all the sets of rank less than α and the proper classes at that stage are the sets of rank α. A formal logic PC or QC can be defined over the class V_α (PC for finite α, otherwise QC), such that the elements of rank less than α are the objects comprehended in PC/QC and the classes of rank α correspond to (one-place) predicates defined over them.

Returning to consciousness, we can represent a subject by a proper class V and the objects that the subject comprehends by elements in V. The V-sets are successive momentary epistemic states in the ongoing life of the subject, and later (but corresponding) V-sets serve as the ontic states by reference to which those epistemic states are evaluated. The subject is thus embodied as successive V-sets and comprehends an accumulating domain of elements.

---

[19] Gödel constructed his constructible universe L as a minimal model of the ZF axioms in order to prove the consistency of the axiom of choice and the continuum hypothesis with those axioms. Decades later, Paul Cohen proved the independence of the axiom of choice and the continuum hypothesis from the axioms of ZF [COHEN 1966].

[20] The von Neumann–Bernays–Gödel axiomatization of set theory is described in [BERNAYS 1968, MENDELSON 1979, QUINE 1969].

A paradox of consciousness is that the inner life of a conscious subject is invisible from outside, whereas the outer form of a subject is invisible from inside. A paradox in set theory represents the situation: the proper class V is invisible from outside (unlike its momentary determinations, which are visible a moment later), whereas the proper element $\varnothing$ is invisible from inside (since it has no inside). Setting the class V and the element $\varnothing$ "back to back" to form a single entity $V|\varnothing = \infty$ is impossible within set theory, since in any suitably formalized regular set theory *ST* we have:

It is a theorem of *ST* that for all x, $x \in V$

It is a theorem of *ST* that for all x, $x \notin \varnothing$

However, closing the cumulative hierarchy into a (transfinite) loop by setting $V|\varnothing = \infty$ (imagine a closed relativistic universe with a time loop, as in Gödel's solution to Einstein's cosmological equations) gives a model of consciousness that should be radical enough to satisfy even a Zen master (Douglas Hofstadter may appreciate it, anyway).[21]

Worlds evolve

Epistemic and ontic states represent worlds. Worlds can also be represented in set theory as V-sets. More accurately, momentary stages in the evolution of worlds can be modeled using V-sets. The fine structure of the cumulative hierarchy can then be used to micromap the ontic evolution of worlds and the epistemological process of reflecting them with ever increasing precision in consciousness.

The actual world that serves as the notional referent of all true propositions outruns its determinations just as V outruns individual V-sets. Thus characterized, the actual world is reminiscent of Kant's transcendental world in that it lies beyond its phenomenal manifestations.[22]

---

[21] Hofstadter made great sport in [HOFSTADTER 1979] with loops and self-reference, and my Zen loop invites similar sport. I first imagined the ∞ loop in 1974, while trying to represent Hegelian logic in set theory. For more reflections on self-loops, including some nice ones by Raymond Smullyan, see also [HOFSTADTER 1981]. Gödel's solution to Einstein's equations is discussed in [HÁJEK 1996].

[22] Immanuel Kant contrasted the transcendental and phenomenal worlds in his classic *Kritik der reinen Vernunft* (first published 1781). In his critique of pure reason, Kant presents a rigorous post-metaphysical analysis of the *a priori* architecture supporting the phenomenology of consciousness. My reading of his analysis in 1972 led in due course to this essay.

Box 1 The propositional calculus (PC)

Atomic propositions A, B, C, … (true T or false F)

\+ Boolean connectives $\neg$ (not), $\wedge$ (and), $\vee$ (or), …

= Compound propositions P, Q, R, … (T or F)

P	Q	$\neg$ P	P $\wedge$ Q	P $\vee$ Q	P $\rightarrow$ Q	P $\leftrightarrow$ Q
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

Rule of inference MP: P, P $\rightarrow$ Q $\vdash$ Q

Box 2 The quantificational calculus (QC)

QC = PC +

Names a, b, c, …
Object variables x, y, z, …
Unary predicate letters A( ), B( ), C( ), …
n-ary predicate letters P(…), …
…
Existential quantifier $\exists$ (for some)
Universal quantifier $\forall$ (for all)

Open sentences: A(x), P(a, x), Q(y, z), … (x, y, z free)
Closed sentences: $(\exists x)B(x)$, $(\exists y)(\forall z)R(y, z)$, … (x, y, z bound)

QC rules of inference = MP +

$\exists$ introduction: $A(a) \vdash (\exists x)A(x)$
$\forall$ introduction: $A(x) \vdash (\forall y)A(y)$ for free x not bound elsewhere
$\exists$ elimination: $(\exists x)A(x) \vdash A(a)$ for name a not used elsewhere
$\forall$ elimination: $(\forall x)A(x) \vdash A(a)$

Leibniz: $a = b \dashv\vdash$ For any A( ), $A(a) \leftrightarrow A(b)$

Box 3 The formal theory of arithmetic (AT)

N = the set of natural numbers

$S(x)$ = the successor of x

Proper axioms of AT: For all $x, y, z \in N$,

$x = y \to (x = z \to y = z)$	$x + 0 = x$
$x = y \to S(x) = S(y)$	$x + S(y) = S(x + y)$
$0 \neq S(x)$	$x \star 0 = 0$
$S(x) = S(y) \to x = y$	$x \star S(y) = (x \star y) + x$

Principle of mathematical induction: For any AT predicate $A(\ )$, $A(0), (\forall x)(A(x) \to A(S(x))) \vdash (\forall x)A(x)$

Box 4 Zermelo–Fraenkel set theory (ZF)

1 Extensionality (defines identity for sets)
$(\forall x)(\forall y)(x = y \leftrightarrow (\forall z)(z \in x \leftrightarrow z \in y))$

2 Regularity (every set has a rank)
$(\forall x)(x = \varnothing \vee (\exists y)(y \in x \wedge y \cap x = \varnothing))$ $\quad x \in V$

3 Pairs $(\forall x)(\forall y)(\exists z)(\forall u)(u \in z \leftrightarrow u = x \vee u = y)$ $\quad \{x, y\} \in V$

4 Union $(\forall x)(\exists y)(\forall u)(u \in y \leftrightarrow (\exists v)(u \in v \wedge v \in x))$ $\quad U(x) \in V$

5 Power set $(\forall x)(\exists y)(\forall u)(u \in y \leftrightarrow u \subseteq x)$ $\quad P(x) \in V$

6 Null set $(\exists x)(\forall y)(y \notin x)$ $\quad \varnothing \in V$

7 Infinity $(\exists x)(\varnothing \in x \wedge (\forall y)(y \in x \to y \cup \{y\} \in x))$ $\quad \omega \in V$

S Separation schema (Zermelo)
For any ST predicate $A(\ )$ and set x, $\quad x \cap \{u \mid A(u)\} \in V$

R Replacement schema (Fraenkel)
For any function F with dom D and cod C, $D(F) \in V \to C(F) \in V$

$ST(1 \ldots 7) \vdash R \to S$

As represented here, in set theory, the actual world is conflated with the representation of the conscious subject. This is harmless: a conscious subject first realizes itself as a separate inhabitant of its environment in the act of cohering a determinate landscape as its ontic reflection, at which point it is represented by a definite *V*-set.

The possible worlds that serve as the set-theoretic correlates of epistemic states have so far been assumed to have the feature that they purport to match the actual world. Such worlds remain possible unless or until they somewhere come into conflict with the actual world. When a contradiction appears between such a possible world and the subject's view of the actual world, that possible world ceases to be possible. Epistemic progress then becomes a matter of progressively pruning the tree of such possible worlds.

A second kind of possible world is familiar in modal logic, namely one that is *counterfactual*. A counterfactual world exemplifies a state of affairs that is clearly alternative to the state that prevails in the actual world. Such a world may be possible in the sense that it obeys all the basic laws of science and is contingently similar to the actual world in various respects, but it differs from the actual world in some specified way. Here, possibility is contrasted not with actuality but with *necessity*. Certain features of the world are regarded as necessary, and variations in all other features count as possible.

Using modal logic, we can theorize about possible worlds independently of whether they are counterfactual. A useful relation is that of *relative* possibility. In terms of the first (epistemic) kind of possible world, world *B* is possible relative to world *A* if, starting from an epistemic state satisfied by world *A*, we can realize an epistemic state satisfied by world *B*. In terms of the second kind, worlds *A* and *B* satisfy the same necessary truths and differ only contingently.

In modal logic, modal operators modify propositions as follows:

"Necessarily *P*" is true in *A* iff, in all possible worlds (relative to *A*), *P*.

"Possibly *P*" is true in *A* iff, in some possible world(s) (relative to *A*), *P*.

Modal logic is still a rather open frontier in logic.[23]

Epistemically possible worlds that are still candidates to determine the actual world can be ranked in terms of *probability*. Given a set of epistemically possible worlds, we can theorize (on the basis of more or less solid science, as the case may be) that they each have some definite probability of being realized,

[23] Kripke's work is the inspiration behind much work in modal logic in recent decades. He proved the completeness of various axiomatizations of modal logic using what are now known as Kripke structures, which are sets of possible worlds plus accessibility relations defined over them. For an introduction to the field, see [POPKORN 1994].

such that the sum of the probabilities over the full set of alternatives is normalized to one. Then a step forward by the subject can determine which of the alternatives is realized, and the process can repeat itself. This process is evolutionary. A steadily better fit between the current epistemic state and the actual world evolves as successive generations of unrealized alternatives are winnowed out.[24]

With probability comes the concept of *entropy*. As generations of outcomes are realized that are probable relative to their predecessor states, the process of epistemic evolution may quickly become unidirectional. Earlier states may cease to be effectively recoverable from later states. As new states unfold, the traces of past states get blurred over. In terms of conscious states, the lost information sinks into unconsciousness and is forgotten.

Typically, the dimension along which epistemic evolution occurs is *time*. In the pure cumulative hierarchy, the ordinal dimension has no obvious interpretation as time, but in all its more concrete incarnations the evolutionary process is somehow temporal.[25] Ultimately, the epistemological concept of time may turn out to be deeper than that defined by physical clocks.

Reality is centered

A world reflects a subject. No sense has been given in this model to the notion that a world could exist without a subject. And no sense needs to be given to that notion. Worlds are centered in the sense that they are structures ultimately constructed from qualia sets, or information. The qualia must be qualia for a subject. The subject must evolve in lockstep with the world it inhabits. At each stage, subject and world reflect each other.

The actual world of contemporary science is a big-bang universe filled with fermions and bosons and sprinkled with DNA organisms.[26] This world is the notional target of ongoing epistemic investigation by numerous scientists. For

---

[24] Evolutionary epistemology is a Darwinian competition in which candidate theories proliferate and suffer attrition under testing. See [DENNETT 1995]. Such evolution can even work in computers [MICHALEWICZ 1996].

[25] Before we go ahead and identify the dimension of epistemic evolution with time, we need to analyze the physical concept of time, in particular its unidirectionality (entropy), its relativistic absorption in spacetime, and its quantum properties. For example, see [ATMANSPACHER 1997, DEUTSCH 1997, FLOOD 1986, PENROSE 1989].

[26] The actual world of contemporary science is a multiauthorial palimpsest, a moving target rather than a defined entity. For me, it is the world described by the last few hundred issues of *Scientific American* (www.sciam.com).

this world, the subject is not human, yet it is defined in its outlines by human subjectivity. The subject of this world can only readily be characterized in terms of its objective reflection. The consistency and coherence of that reflection constitute what Kant might have called the synthetic unity of apperception of the transcendental subject located beyond the phenomenal self. In human terms, the subject of the universe may be seen as a highly schematic envelope pushed out at each point by the activities of different human scientists.[27]

Worlds, as defined here, are informatic constructs. The formalized language used to define them need bear no simple relation to any natural language, but it is always a symbolic structure. It can always be represented in terms of bits of information. For example, it may be a visual code whose ultimate elements are colored pixels, generating a movie that depicts an evolving world.

More generally, a world is a multimedia presentation portraying a virtual reality (VR).[28] In the epistemological scenario discussed here, a VR world is a candidate for representing real reality, otherwise known as the actual world. A VR world must be centered on a subject. A VR world is a symbolic construct, and a subject must experience it to realize the symbolism. The symbolism works both ways. The subject is itself realized in the evolution of its VR reflection. In this model, without the external correlate embodied in the VR world, the subject would collapse to the null state.

Returning to modal logic, consider a set of epistemically possible worlds arrayed before a subject inhabiting a VR world *A*. For the subject, world *A* is a transparent and presumably accurate and reliable representation of the actual world. For that subject, world *A is* the actual world. Only new experience that somehow contradicts the facts that constitute world *A* can force the subject to move on. Yet the subject can readily entertain a set of possible VR worlds that would somehow extend or replace world *A*. Since the specification of world *A* is limited (as a particular *V*-set, say), it is *always* possible to define a set of further worlds that are possible relative to *A*. So long as *A* remains a viable VR for the subject, *A* is *symmetrical* with respect to the possible worlds in that set. It can become any one of them.

---

[27] The subject of the universe can be represented by its reflection in the more or less unified core of ideas in cosmology. This core has only settled down recently (if at all), with the long-awaited marriage of general relativity and quantum field theory in some variant of superstring theory, or rather in a generalization of superstring theory called M theory (by Edward Witten). See [GREENE 1999].

[28] David Deutsch has made a highly original analysis of VR worlds based on computability theory [DEUTSCH 1997]. In his view, scientific theories provide the software for generating the VR in cerebral wetware. See also [HEIM 1993].

Theoretical physicists have made intensive use of the concepts of symmetry and spontaneous symmetry breaking, and here is a context where they fit well.[29] Any VR world *A* is symmetrical with respect to all worlds that are epistemically possible relative to *A*. World *A* could evolve into any one of those worlds, either following an epistemic breakthrough on the part of the subject experiencing world *A* or simply following the passage of a suitable increment of time. In the latter case, where time suffices, it is natural to say that the symmetry is broken spontaneously.

In the evolution of VR worlds through their momentary determinations, spontaneous symmetry breaking can occur when there is no way to predict the new determination. In this case, the new determinacy of the subsequent world may be *random*. Alternatively, deep theoretical reasons may emerge later as to why world *B* and not world *C* was realized (in which case the symmetry was not broken spontaneously).

Any VR world *A* embodies only limited determinacy, therefore it may be further determined to become some other world *B* that is epistemically possible relative to *A*. For example, world *A* may be the actual world of contemporary science at time t_1. At time t_1, it is not yet determined whether nucleus *X* in a given laboratory experiment will undergo alpha decay in the near future. At time $t_2 > t_1$, nucleus *X* emits an alpha particle. If world *B* is the actual world of contemporary science at time t_2, then world *B* embodies more determinacy than world *A*. If world *C* is like world *B* in all respects except that in world *C* nucleus *X* did not decay, then at time t_1 world *A* was symmetrical with respect to the future worlds *B* and *C*.

Pursuing the example, imagine that in world *A* the unstable nucleus *X* is hidden from any kind of observation. Then world *A* can persist up to time t_2 in a *superposition* of states *B* and *C*. The superposition collapses only when an observation determines whether nucleus *X* decayed or not, at which time world *A* becomes world *B* or *C*, respectively. More generally, according to quantum theory, such hidden determinacy can evolve superpositionally for arbitrarily long periods of time (this is the moral of the story of Schrödinger's cat), so we can be living in a world that is now a superposition of any number of states of systems that remain unobserved in our world. In fact, our actual world is never in a unique quantum state. We always live in a world that embodies a super-

[29] Symmetry and symmetry breaking are technical concepts in physics. If the universe at time *t* is described by a theory in which changing parameter *p* leaves the physics unchanged, then the universe at time *t* is symmetrical with respect to *p*. See the last lecture in [FEYNMAN 1963]. Symmetry breaking occurs when new order appears during a phase transition, as when water freezes to become ice.

position of states and is symmetrical with respect to different possible outcomes of measurements of those states.[30]

The formal model of consciousness presented here is well suited for interpretation in terms of quantum mechanics. Ontic state *A* evolves into a superposition of quantum states, then something happens and the superposition collapses into state *B*. State *B* evolves into a superposition, and so on. The whole story of evolving states in consciousness can be told in such terms, with the rhythm of the changing states determined by the speed with which the superpositions collapse.[31] In principle, it even seems conceivable that some such quantum story could be told for the evolution of states of electrochemical excitation in human brains, where the collapse times for interneural resonance quanta correspond to moments of specious present (of *now*) in consciousness.[32] However, the formal model is independent of such a brainbound interpretation. The story of collapsing quantum superpositions is interesting at the microscopic level across the whole of physical reality. Indeed, the relativity of determinacy to an observing subject has long been a puzzle for physicists.[33]

Any VR world corresponding to a definite epistemic state reflects a limited subject. The totality of information represented by that world fails to determine any amount of detail that further investigation can reveal. Any such world is centered on that subject and is inconceivable without that center. As epistemic

---

[30] The coexistence of superposed states in quantum mechanics is a tricky concept for the nonspecialist to grasp. Abner Shimony reviews the conceptual foundations of quantum mechanics in [DAVIES 1989]. Readable books on the topic include [FEYNMAN 1985, RAE 1986]. Original ideas appear in [DEUTSCH 1997, LOCKWOOD 1991].

[31] It may seem that collapse times for coherent states in quantum systems must be too short to explain any empirically reasonable rhythm of conscious states. But recall that by the Heisenberg uncertainty principle, the collapse time T for a superposition of states of a photon with energy E is such that the product $ET \approx h$ (the Planck constant). And photon energy $E = hf$, where f is the frequency of the photon. So, for $f = 40$ Hz, the collapse time $T \approx 25$ ms, in rough agreement with the flicker fusion rate. However, the corresponding spatial uncertainty is almost the size of the Earth!

[32] The question of whether quantum phenomena play an important role in brain function has a very unsettled history. Speculation abounds on the role of 40 Hz resonances, the properties of Bose–Einstein condensates, quantum gravity, laser action in microtubules, and more. For example, see [PENROSE 1989, PENROSE 1994, STAPP 1993, WOLF 1981, ZOHAR 1990]. Scepticism is still appropriate.

[33] The Copenhagen interpretation of quantum mechanics features a crucial role for the observer. Among interpretations that play down the observer, Everett's "many worlds" view is popular: see [CHALMERS 1996, DEUTSCH 1997, LOCKWOOD 1991].

agents in an ontic environment that we can only access through the medium of our VR tools (concepts, theories, imaging hardware, brains, and so on), we are doomed to have our own perspective on reality.

I am conscious

I experience an evolving series of VR worlds, therefore I am a conscious subject. These VR worlds are pixelated with qualia and structured with logic. My experience is ordered along a timeline as a series of states of knowledge. These states embody limited determinacy and evolve into their successors in a great variety of ways. Each step I take along this series is a transition between two states, the *before* and *after* states, and in general the states are contradictory, since they are competing representations of the actual world.

During each transition between states, I change. Either I briefly bridge the two states or I transit a null state between them. Alternatively, at each step I briefly relive all the states from the null state to the new state, in an ever-increasing spiral movement that invites description in terms of the apparatus of V-sets. Such stepwise or cyclic movement in a space of VR worlds may be constitutive of consciousness.

So far, the entire description of subjective consciousness here has remained neutral with respect to persons. The whole story could have been told for a Hegelian *Weltgeist*[34] as well as for a normal human being with a personal life, or indeed for a robotic subject embodied in electronic circuitry. In terms of Martin Buber's distinction between *I–it* and *I–you*,[35] the story so far has been the drama of *I–it*. That drama is sufficient to account for the objectivity of mathematics and physical science, in the sense that no relativity to *personal* perspective is required of such theory. But it leaves much unexplored.

The denotation of the term "I" is one of the most deeply puzzling subjects of all. I cannot be an object to myself, any more than in set theory the proper class V can be a member of V. Yet I confront my limits with every passing moment, and those limits enable others to see me as coterminous with an object inside

[34] The *Weltgeist* ("world spirit") seems to be the main subject in large parts of Hegel's *Phänomenologie des Geistes* (first published 1807). This obscure and difficult classic tells an extraordinary story of the evolution of consciousness from sensory immediacy to the "absolute in which all is one". Strongly influenced by Kant, Hegel went further and built a dialectical idealism that used contradictions as stepping stones to the truth. The story is reviewed in [TAYLOR 1975].

[35] Martin Buber's *Ich und Du* (first edition 1923) is a philosophical classic [BUBER 1970]. His analysis of "I" leads on to some theological thoughts.

their world. By analogy, I have learned to represent myself as an object inside my world. This *analog* I is a cultural invention by means of which I become socialized in a public world. If we could not refer to ourselves in this relatively objective way, we would each incorrigibly regard our own self as possessing mystic truth and therefore would be unable to discuss anything rationally at all. But the analog I is not the real me. In terms of set theory, the analog I is a normal set within some V-set, whereas the real me is the epistemic subject, the paradoxical entity ∞.

By accepting that you exist as a subject like me in my world, I learn how I can see my own limits and accommodate them gracefully. By accepting that each human being, however different in knowledge or ability, also exists as a subject, I learn to dissociate the concept of subjectivity from all its contingent entanglements. Conversely, by observing the extensive isomorphism of subjectivity among skilled practitioners of a scientific discipline,[36] I can account for the unity and objectivity of the actual world of contemporary science.

The concept of a logical subject is distinct from the concept of a person. I am necessarily a logical subject but only contingently a person. I can accommodate other persons in my world without difficulty, but accommodating other logical subjects would lead either to a branching of the self (schizophrenia) or to a loss of self (zombiism). It is a condition for the healthy existence of my self that my subjective consciousness is unique. If your personal experiences were somehow piped into my brain, they would become my subjective experiences as well. If my personal experiences were simultaneously piped into your brain, we would morph into one logical subject (or flip out).

Consciousness is a state that I can define subjectively rather easily: it is the state I am in, now. Defining it objectively is another matter. It can naturally be defined in some detail in terms of awareness, alertness, and so on, but there is always the residual question of whether such a definition can really exhaust the possible complications.[37] Observation of patients reporting lucidly on their own mental states comes closest to providing criterial evidence of consciousness from the outside, but the catch here is the term “lucidly”. I regard your speech as lucid when I *understand* it, and that occurs only when I can put my own subjectivity behind your words, so to speak. My subjective consciousness is thus

---

[36] The mathematician G. H. Hardy once said that all mathematicians are isomorphic. He meant they all think the same way. I would say that mathematicians deal not only with mental constructs but also with the same (platonic) realm, so they are of *one mind.*

[37] That we face complications when we try to define how humans instantiate consciousness will not surprise readers of [DAMASIO 1994, DAMASIO 1999, RAMACHANDRAN 1998, WEISKRANTZ 1997].

projected through your speech output, much as yours is through mine when you understand my words here.

Our predicament as logical subjects is quite stark: we are one. However closely I identify with you, however freely I grant personhood to all the intelligent organisms on the planet, my experience is mine, and I accept only on faith that yours is yours. The moment I *know* your experience, it becomes mine too. If we share our experience, we become one. As human society becomes more integrated and more pervasively intimate, we shall probably cease to see our separate personal selves as separating our logical subjectivity. We shall each feel for all of us, as integral parts of a global lifenet. Consciousness will be globalized, and the assertion "I am conscious" will take on a new meaning.

We are conscious

In future, engineers may implement consciousness in computational hardware, for example in artificial neural networks or even in a global network like the Internet.[38] The creation of artificial consciousness will surely help us to understand and appreciate human consciousness more fully.

Consciousness is now defined ostensively in terms of how it is manifested in humans. Roughly, human are conscious when they are capable of perceiving their environment and making sense of it in some suitable way. Researchers have explored all this in some detail, albeit without much theoretical guidance. Because theoretical efforts have so far fallen short, there is a great temptation in consciousness research to let it drift into the field of biology and to forget the wider landscape in which our work can really bear fruit.

That said, the biological research program in consciousness studies is clearly the initiative with the best short-term prospects of achieving a breakthrough. Once we know exactly which structures and processes in human brains are responsible for consciousness, we can deepen that knowledge and give it the theoretical foundations that have so far been missing. This will surely smooth the way for any future project of building conscious machines.

The biological search for the neural correlates of consciousness is exciting.[39] Spatiotemporal maps of electrochemical activity in the cerebral neuronets of

[38] For a technical introduction to artificial neural networks, see [ROJAS 1996]. The idea that humans online are like neurons in a global brain is part of our *Zeitgeist*.

[39] I sensed this excitement at the conference "Neural Correlates of Consciousness: Empirical and Conceptual Questions" held in Bremen, Germany, June 19–22, 1998. This was the second conference organized by the Association for the Scientific Study of Consciousness (ASSC – http://assc.caltech.edu).

experimental subjects clearly show detailed and rather exact correlations with the introspected conscious experience of those subjects. As new imaging technology enables us to generate more focused maps, with better spatial and temporal resolution, we can reasonably hope that the correlations will become steadily more fine-grained and illuminating.

However, such maps alone cannot reveal the mechanisms of consciousness. For example, we still cannot explain:

- The binding problem. Large numbers of individual neural excitations are somehow bound together into unified mental images. Introspection reveals qualitatively variegated images, not a blizzard of blips, but how the blips merge into such images is still a puzzle.
- The unity of consciousness. In a brain where billions of neurons are firing in rapid and extremely complex rhythms, fleeting images pass over the screen of the inner Cartesian theater like scenes in a movie. How does this inner theater work, and where am I in it?

As for the binding problem, Wolf Singer and his team have emphasized the possible relevance of resonances between neurons firing synchronously at frequencies of about 40 Hz in explaining how disparate neural excitations form unified mental images.[40] The physical details of this process are still unknown. The electrical waves may trigger chemical changes that stabilize neural groups, which then always fire together to give a qualitatively unique experience at a rather complex level. Alternatively, some hitherto unknown quantum effect may play an essential role.

As for the unity of consciousness, my own view is that successive states of excitation become fused in an ongoing excitation loop that can be analyzed formally in terms of the model presented here. The loop creates a VR world that evolves through a series of momentary determinations as the inbound stream of image data is processed. Each determination can be modeled as a V-set that freezes a single scene in the ongoing movie. The neural hardware (the wetware) that implements the loop creates a look and feel that prompts the movie theater metaphor. The ongoing show is the way the loop appears to itself. Somehow, the loop is transparent enough to show the last scene in its entirety, but not transparent enough to spoil the view with simultaneous images of previous scenes. Yet the loop also has access to a large repository of previous scenes, and can recall them from memory using an associative mechanism.

---

[40] Wolf Singer reports this work in his chapter in [ROSE 1998]. He also reported it in his lecture at the ASSC conference in Bremen in 1998.

The evolution of consciousness in the history of life on Earth is another field where new insights can be expected. Our present understanding of consciousness suggests that relatively simple neuronets, such as those in all mammalian species, may exhibit some form of consciousness. Given this understanding, the idea that consciousness is a distinguishing mark of human beings is unlikely to survive.[41] However, the highly cultivated forms of consciousness exhibited in human communities are certainly unique to our species, and are certainly responsible for the major role humans now play in the ongoing reconfiguration of consciousness support systems on planet Earth.[42]

Conclusion

The formal model of consciousness presented here can be developed further using the tools of logic and mathematics. The model represents consciousness as the inner transparency of a set-theoretic loop in which an evolving VR world is brought to an experienced focus. Such loops can presumably be implemented in a variety of hardware or wetware architectures and support a corresponding variety of active subjects. When the loop is implemented in a human brain and interacts with its environment over the usual human sensorimotor modalities, the result is consciousness as we all know it.

The "hard problem" of accounting for qualia is circumvented in this model. Qualia are the raw inputs for a conscious subject. Normally, they are processed into complex landscapes before they emerge in consciousness, and they are not distinguished individually within those landscapes. However, regardless of how they appear, my qualia are mine alone in the same sense that the entire universe is mine alone. Qualia do not present a separate problem.

The model suggests that our personal identities are constructs with limited validity. If we are to build a truly shared world, we need to grow beyond them toward a new, more universal identity.

---

[41] Until recently, consciousness was often seen as a special attribute of human beings. Julian Jaynes even argued in [JAYNES 1976] that humans first became conscious in biblical times, but his argument may be interpreted rather as showing that humans first developed modern *personal* consciousness then. The Bible and other early documents can plausibly be read as recording the slow emergence of the concept of a person.

[42] On various philosophical and practical aspects of building a global consciousness, see [RUSSELL 1991, SAGAN 1990, STOCK 1993].

References

Atmanspacher, Harald, and Eva Ruhnau (editors): *Time, Temporality, Now.* Experiencing time and concepts of time in an interdisciplinary perspective. Springer 1997

Benacerraf, Paul, and Hilary Putnam (editors): *Philosophy of Mathematics.* Prentice-Hall 1964

Bernays, Paul: *Axiomatic Set Theory.* North-Holland 1968

Boolos, George S., and Richard C. Jeffrey: *Computability and Logic.* Second edition. Cambridge University Press 1980

Buber, Martin: *I and Thou.* Translated by Walter Kaufmann. Scribner 1970

Chaitin, Gregory J.: *The Limits of Mathematics.* A course on information theory and the limits of formal reasoning. Springer 1998

Chalmers, David J.: *The Conscious Mind.* In search of a fundamental theory. Oxford University Press 1996

Churchland, Patricia Smith: *Neurophilosophy.* Toward a unified science of the mind/brain. MIT Press 1986

Clark, Andy: *Being There.* Putting brain, body, and world together again. MIT Press 1997

Cohen, Paul J.: *Set Theory and the Continuum Hypothesis.* Benjamin 1966

Cotterill, Rodney: *Enchanted Looms.* Conscious networks in brains and computers. Cambridge University Press 1998

Crick, Francis: *The Astonishing Hypothesis.* The scientific search for the soul. Simon and Schuster 1994

Damasio, Antonio R.: *Descartes' Error.* Emotion, reason, and the human brain. Grosset/Putnam 1994

Damasio, Antonio R.: *The Feeling of What Happens.* Body and emotion in the making of consciousness. Harcourt Brace and Co. 1999

Davies, Paul C. W. (editor): *The New Physics.* Cambridge University Press 1989

Dennett, Daniel C.: *Consciousness Explained.* Little, Brown, and Co. 1991

Dennett, Daniel C.: *Darwin's Dangerous Idea.* Evolution and the meanings of life. Simon and Schuster 1995

Deutsch, David: *The Fabric of Reality.* Penguin Press 1997

Dummett, Michael A. E.: *Frege.* Philosophy of language. Duckworth 1973

Edelman, Gerald M.: *Bright Air, Brilliant Fire.* On the matter of the mind. Basic Books 1992

Feynman, Richard P., Robert B. Leighton, and Matthew Sands: *The Feynman Lectures on Physics.* Volume I. Mainly mechanics, radiation, and heat. Addison-Wesley 1963

Feynman, Richard P.: *QED.* The strange theory of light and matter. Princeton University Press 1985

Flood, Raymond, and Michael Lockwood (editors): *The Nature of Time.* Blackwell 1986

Fraenkel, Abraham A., Yehoshua Bar-Hillel, and Azriel Lévy: *Foundations of Set Theory.* Second edition. North-Holland 1973

Gazzaniga, Michael S., Richard B. Ivry, and George R. Mangun: *Cognitive Neuroscience.* The biology of the mind. W. W. Norton 1998

Greene, Brian: *The Elegant Universe.* Superstrings, hidden dimensions, and the quest for the ultimate theory. W. W. Norton 1999

Greenfield, Susan A.: *Journey to the Centers of the Mind.* Toward a science of consciousness. W. H. Freeman and Co. 1995

Gries, David, and Fred B. Schneider: *A Logical Approach to Discrete Math.* Springer 1993

Guth, Alan H.: *The Inflationary Universe.* The quest for a new theory of cosmic origins. Jonathan Cape 1997

Hájek, Petr (editor): *Gödel '96.* Logical foundations of mathematics, computer science and physics – Kurt Gödel's legacy. Springer 1996

Heim, Michael: *The Metaphysics of Virtual Reality.* Oxford University Press 1993

Hofstadter, Douglas R.: *Gödel, Escher, Bach: An Eternal Golden Braid.* A metaphorical fugue on minds and machines in the spirit of Lewis Carroll. Basic Books 1979

Hofstadter, Douglas R., and Daniel C. Dennett (editors): *The Mind's I.* Fantasies and reflections on self and soul. Basic Books 1981

Jaynes, Julian: *The Origin of Consciousness in the Breakdown of the Bicameral Mind.* Houghton Mifflin 1976

Jech, Thomas: *Set Theory.* Second edition. Springer 1997

Kripke, Saul: *Naming and Necessity.* Blackwell 1980

Kuhn, Thomas: *The Structure of Scientific Revolutions.* Second edition. University of Chicago Press 1971

Lakatos, Imre, and Alan Musgrave (editors): *Criticism and the Growth of Knowledge.* Cambridge University Press 1979

Lockwood, Michael: *Mind, Brain and the Quantum.* The compound 'I'. Blackwell 1991

Mendelson, Elliott: *Introduction to Mathematical Logic.* Second edition. D. Van Nostrand Co. 1979

Metzinger, Thomas (editor): *Conscious Experience.* Schöningh/Imprint Academic 1995

Michalewicz, Zbigniew: *Genetic Algorithms + Data Structures = Evolution Programs.* Third edition. Springer 1996

Penrose, Roger: *The Emperor's New Mind.* Concerning computers, minds, and the laws of physics. Oxford University Press 1989

Penrose, Roger: *Shadows of the Mind.* A search for the missing science of consciousness. Oxford University Press 1994

Pinker, Steven: *The Language Instinct.* The new science of language and mind. Allen Lane 1994

Pinker, Steven: *How the Mind Works.* W. W. Norton 1997

Popkorn, Sally: *First Steps in Modal Logic.* Cambridge University Press 1994

Popper, Karl: *Objective Knowledge.* An evolutionary approach. Oxford University Press 1972

Quine, Willard Van Orman: *Word and Object.* MIT Press 1960

Quine, Willard Van Orman: *Set Theory and Its Logic.* Second edition. Harvard University Press 1969

Rae, Alastair I. M.: *Quantum Physics: Illusion or Reality?* Cambridge University Press 1986

Ramachandran, V. S., and Sandra Blakeslee: *Phantoms in the Brain.* Human nature and the architecture of the mind. Fourth Estate 1998

Rojas, Raul: *Neural Networks.* A systematic introduction. Springer 1996

Rose, Steven (editor): *From Brains to Consciousness?* Essays on the new sciences of the mind. Allen Lane 1998

Rucker, Rudy: *Infinity and the Mind.* The science and philosophy of the infinite. Harvester Press 1982

Russell, Peter: *The Awakening Earth.* The global brain. Revised edition. Arkana 1991

Sagan, Dorion: *Biospheres.* Metamorphosis of Planet Earth. Arkana 1990

Scott, Alwyn: *Stairway to the Mind.* The controversial new science of consciousness. Copernicus 1995

Shear, Jonathan (editor): *Explaining Consciousness – The 'Hard Problem'.* MIT Press 1997

Smolin, Lee: *The Life of the Cosmos.* Oxford University Press 1997

Stapp, Henry P.: *Mind, Matter, and Quantum Mechanics.* Springer 1993

Stock, Gregory: *Metaman.* Humans, machines, and the birth of a global super-organism. Bantam Press 1993

Taylor, Charles: *Hegel.* Cambridge University Press 1975

Weiskrantz, Lawrence: *Consciousness Lost and Found.* A neuropsychological exploration. Oxford University Press 1997

Wittgenstein, Ludwig: *Tractatus Logico-Philosophicus.* Routledge and Kegan Paul 1922

Wittgenstein, Ludwig: *Philosophical Investigations.* Blackwell 1958

Wolf, Fred Alan: *Taking the Quantum Leap.* The new physics for nonscientists. Harper and Row 1981

Zohar, Danah: *The Quantum Self.* Bloomsbury 1990