Systems Research Vol. 13 No. 3, pp. 293–310 1996

■ *Research Paper*

Virtual Logic

Louis H. Kauffman

Department of Mathematics, University of Illinois at Chicago, 851 South Morgan Street, Chicago, Illinois 60607–7045, USA

Virtual logic is not logic, nor is it the actual subject matter of the mathematics, physics or cybernetics in which it may appear to be embedded. Virtual logic lives in the boundary between syntax and semantics. It is the pivot that allows us to move from one world of ideas to another. This paper studies the virtuality of ordinary and mathematical logic. Through examples it is shown how what we call 'ordinary reason' is itself a paradox. Reason itself is not at all reasonable! Each new mathematical construction, each new distinction, each theorem is an act of creation. Ordinary reason itself is virtual. The credo of clarity is not ordinary. It goes beyond reason into a world of beauty, communication and possibility.

Keywords virtual; paradox; logic; mathematics; cybernetics; self-reference; clarity

1. INTRODUCTION

> **vir′tu.al** (vur′tu.al), adj. 1. *Archaic.* Of or relating to a virtue or efficacious power; energizing. 2. Being in essence or in effect, but not in fact; as the *virtual* rulers of a country.—**vir′tu.al′i.ty** (-al′i.ti), n.—**vir′tu.al.ly**, *adv.*

I take the meaning of the word virtual in the archaic sense. Virtual logic is that which energizes reason and so brings the forms of logic and mathematics into being.

Virtual logic is not logic, nor is it the actual subject matter of the mathematics, physics or cybernetics in which it may appear to be embedded. Virtual logic lives in the boundary between syntax and semantics. It is the pivot that allows us to move from one world of ideas to another. The power of virtual logic is that it is not just a pivot. It provides the real possibility and the means for the opening of communications across boundaries long thought to be impenetrable. Here there is no guarantee of success. It is in the place where we make the most valiant attempt at clarity and communication that our inevitable failures to communicate have the potential to be transformed into new worlds.

Section 2 is topological, concentrating on the Möbius band as an exemplar of a system with local distinctions and global connectedness. By viewing the Möbius as an embodiment of the Liar Paradox, we see how this paradox is actually a nexus of ideas that can be applied in understanding observing systems. Section 3 discusses reference, self-reference, fixed points, recursions and the use of imaginary values (also known as complex numbers). It is explained how the number π is naturally an amalgam of formal fixed points whose relative values produce the ratio of the circumference of a circle to its diameter. It is remarkable that domains imaginary with respect to arithmetic are vitally real with respect to geometry. In section 4 we discuss the general Church–Curry fixed point construction ($GX = F(XX)$ implies $GG = F(GG)$) in relation to

CCC 0731-7239/96/030293-18

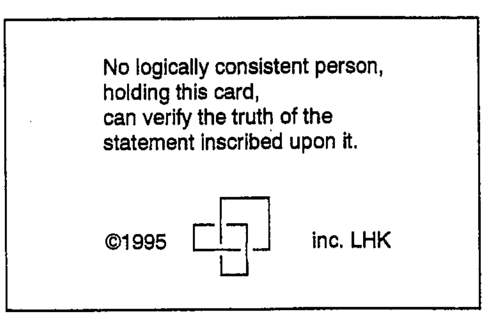
the Russell paradox and the issues already raised in this paper. In section 5 we introduce a basic referential shift (from A⟶B to AM⟶AB) in the context of a painting by Magritte. This shift is compared to the shift of an observer who separates a world into the one who sees and the that which is seen. Language is the interface of this separation. Necessarily, language is shifted to both sides of the apparent distinction. The Magritte shift is the central pattern behind the Gödel Incompleteness Theorem and constructions for direct and indirect self-reference (I⟶M then IM⟶IM). This is taken up in section 6. Section 7 continues the discussion of section 6, showing how Löb's paradox (which dangerously proves any proposition!) can be tamed in a Gödelian way to yield insight into the limitations of formal systems and to show that the self-affirming Gödelian sentence is provable in consistent formal systems. This Löb Theorem is a beautiful affirmation of the dictum that behind every paradox there is a rich vein of virtual logic. By this point in the paper, the author has split into two dialoguing parts in a fugue of understanding understanding.

Paradoxes are gateways into new worlds.

Before reading the rest of the paper, the reader may like to try his/her hand at the following puzzle. You have been handed a card with an inscription that reads:

> No logically consistent person, holding this card, can verify the truth of the statement inscribed upon it.

You place the card on the table before you and, reading it, reason quite correctly that indeed the statement on the card is true. For if the holder of the card had verified the truth of the statement, then the statement would undermine this


No logically consistent person,
holding this card,
can verify the truth of the
statement inscribed upon it.

©1995 inc. LHK

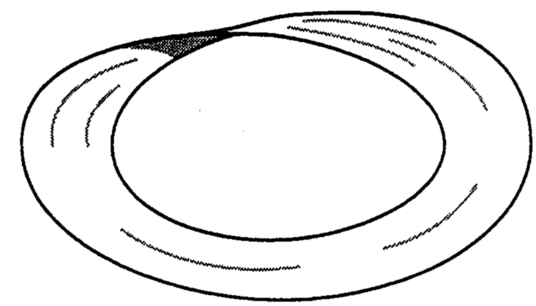


Figure 1

verification. Now you reach out and take the card in hand. Where has your verification of the truth of the statement gone?

It gives the author great pleasure to thank Heinz von Foerster, Annetta Pedretti, James Flagg, Gary Berkowitz, Steve Sloan and Daniel Davidson for many stimulating conversations in the course of imagining this paper.

2. TOPOLOGICAL LOGIC, PARADOX LOGIC, KNOT LOGIC

> **par'a.dox** (par'a.doks), n. [F. *paradoxe*, fr. L., fr. Gr. *paradoxon*, neut. of *paradoxos*, adj., fr. para beside, contrary to + *doxa* opinion.] 1. A tenet contrary to received opinion; also, an assertion or sentiment seemingly contradictory, or opposed to common sense, but that yet may be true in fact. 2. A statement actually self-contradictory, or false.

Sometimes a paradox can be used to reason to a real and sensible solution of an intellectual or mathematical problem. Sometimes a paradox has a geometrical, topological or structural counterpart that is special, useful and fascinates us with its centrality. The Möbius band is just such an object. View Figure 1.

A Möbius band has one edge and one side. Locally, an observer on the band will see two edges and detect two sides. If the observer walks along the band he/she will eventually return to her starting place, but will find that she is on the other side of the band! In this case, it is probably best to have two observers on the band. They start in the same spot, but one walks along while Mr Fly waits. Eventually, they come back together, but Ms Fly finds that she is now across the band from Mr Fly.

One more turn around the band puts them back together.

The Möbius band is paradoxical in the first meaning of the term. Its properties go against common sense and yet they are true.

The band embodies a form of which it may be said that it is ONE (one side) and yet it is also MANY (two sides). Since this form is topological it is both one and many at the same time. To speak this way in the absence of the model would seem to be paradoxical in the second sense of the word (self-contradictory), yet in the Möbius band we have a picture before us of just how this unity in multiplicity can come about. It is ONE for the global observer and MANY for the local observer. It is ONE for the local observer who is willing to travel and experience the whole. It is MANY for the global observer who is willing to play the game that looks locally and forgets for a moment the whole.

The circular interconnectedness of the band provides the stability that keeps it with one side and one edge. One can look at the band as a topological realization of a circuit that interconnects the output of an inverter with its input (Figure 2).

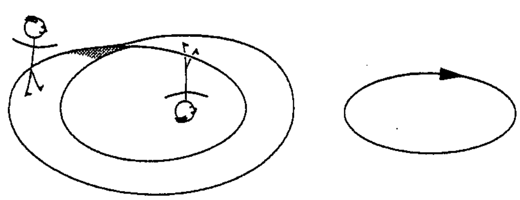


Figure 2

Paradox Generates Time

The topological analogue of the inverter is the half-twisted band. An observer with head up is flipped to an observer with head down upon passing through the twist. The abstract inverter is usually depicted as shown in Figure 3. It takes

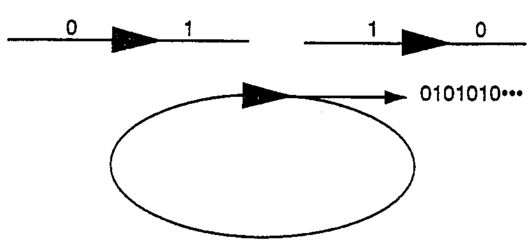


Figure 3

a Boolean signal and flips it from 0 to 1 or from 1 to 0. If the signal is motivated to go around the circuit then it becomes the temporal embodiment of the Liar Paradox: 'This statement is false.' Each turn flips the signal and the output is an oscillation that indicates the presence of the contradiction. Paradox generates time.

In the topological and timeless world of the Möbius strip there is no oscillation (unless the fly insists on running around the band). The inverter (the twist) is itself held in form by the circular interconnection of the band. There would be no twist without the band and there would be no band without the twist.

To attain fully to the condition where observer and observed are one, it is necessary to go a step beyond the Möbius band as topological object. The human observer is already a cybernetic Möbius band. The space outside is only known inside and through the twist of perception that makes the inside appear outside. The observer fancies a whole that cannot be seen and a distinction of (outside/inside) that is imaginary. Thus appears an illusion of seer and seen.

Projective Space

Möbius is the core of the logic of projective three-dimensional space. Send a ray straight out from the observer. Continue that ray on out to infinity and imagine that the infinity before you is directly connected with the infinity behind. The great sphere of points at infinity is doubled upon itself so that forward rays meet backward rays and every straight line becomes a circle. In this folded space of three-dimensional projective geometry the Möbius band lives in the twist through infinity. A road sent forward and back meets itself at infinity and is identical to a Möbius strip.

Observer/Space

Let us return to the observing system. The topologist's Möbius band is a viable abstract model. The system inverts observation in the act of separation into observer and observed. (Let the passage through the twist connote observation.) The system is created just to maintain this self- observation and will return to flatness when observation ceases. (Cut the band cross-wise and the loop will break and untwist.) The system is its own self description: L = NOT(L). See Figure 4.

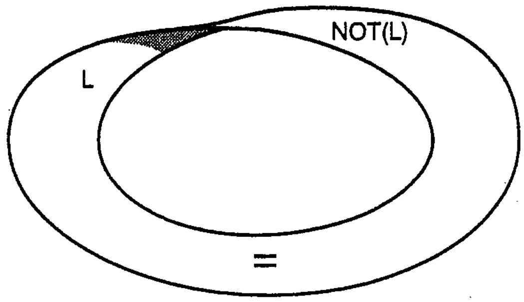
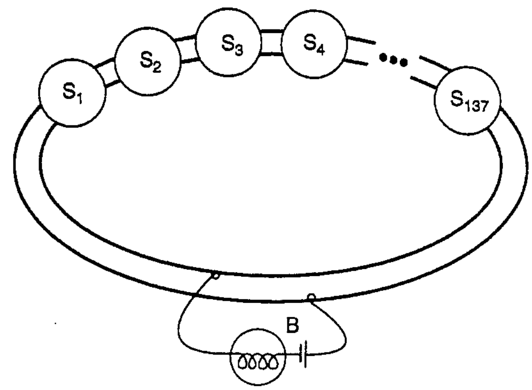


Figure 4



137 Independent Controls for B

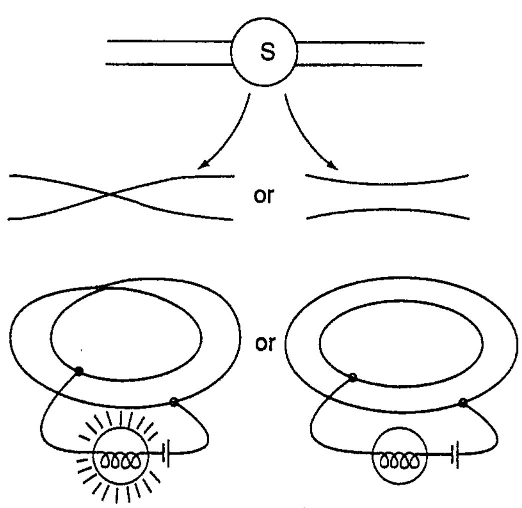
Figure 6

Applying the Möbius

As Ricardo Uribe points out (1995), the Möbius band will solve real problems in switching theory. We want a system of switches, each controlled locally, such that each switch controls a single device. For example, each switch is to turn a given light on or off. Consider the Möbius and *regard the twist as a switch with the two positions: twisted or untwisted.* Replace the twist in a Möbius with no twist and the resulting band has two sides and two edges.

Make a band with 137 half-twists. Replace any twist with an untwist, and there will be two boundary components to the band. Flip any other twist and the band is Möbius once more. Each twist-switch individually controls the connectivity of the boundary of the band. Attach the bulb and battery across the local distinction of band sides and the desired circuit is created. Each switch is easily made (double-pole, double-throw) and the whole circuit design proceeds from Möbius topology with not a hint of Boolean algebra.

Is this design the simplest possible? No. Is the circularity inherent in the Möbius band necessary for the design? No. Can an equivalent design evolve from standard logic and Boolean algebra? Yes, but with much more attention to algebra. What is the moral of this tale? A paradox can tell a



Wiring the Boundary

Figure 5

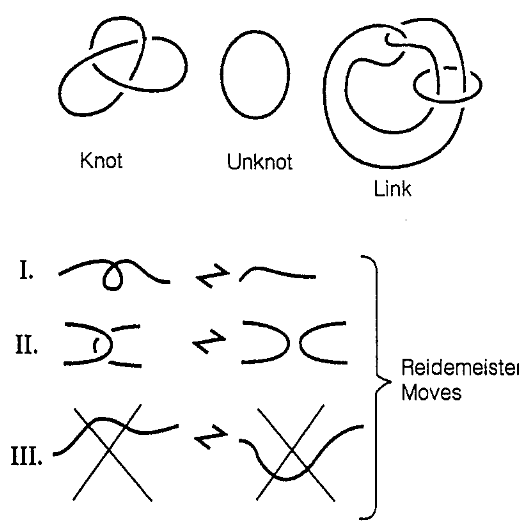


Figure 7

tale that jumps across the steps of logical analysis. Opposites are joined in the whole.

Knot Logic

In this short section, we return to fixed points and self-reference from the point of view of knots and links in three-dimensional space. The idea is very close to the remarks about the Möbius band, but here we consider all possible knots, links and tangles. For a more detailed exposition see Kauffman (1995a).

Knots and links are represented by diagrams like those shown in Figure 7. A knot is a single closed loop in space and it is said to be knotted if it cannot be topologically deformed to a standard flat circle in the plane. A link has many loops. Two loops are said to be linked if it is not possible to separate them by topological deformation. Topological deformation corresponds to moving the loops continuously without crossing them through themselves or one another.

This figure also shows the schema of three local moves on the diagrams that generate all topological deformations of knots and links in three-dimensional space. These Reidemeister moves provide a translation from the complex problem of topology of loops in three-space to an equally complex problem of understanding a two-dimensional diagrammatic calculus. This calculus of diagrams is the virtual logic behind the topology of knots and links. A great deal of mathematics has been devoted to the study of this branch of topology (see Kauffman, 1987, 1991, 1993).

In this section we point out how the diagrams can be interpreted for a version of set theory without the Axiom of Foundation. This Axiom forbids infinite descending chains of membership such as are implicit in the set whose only member is itself: $\Omega = \{\Omega\}$.

Knot sets (Kauffman, 1995a) are knot and link diagrams interpreted in terms of set membership through the conventions that

(1) labeled arcs in the diagram correspond to the members of the set;
(2) if arc a crosses under arc b then we say that a is a member of b ($a \in b$).

In this way, we have $\Omega = \{\Omega\}$ represented by a diagram with a simple twist as shown below.

Mutually creative sets such as A={B}, B={A} correspond to a link.

See Kauffman (1995a) for a discussion of how this theory interfaces with the Reidemeister moves. Here we only comment that the self-reference in the knot model for Ω shows clearly that the infinite descent implicit in the self-reference is no more serious than the fact that one can go round a circle infinitely often. Circulation of the loop representing Ω takes an observer through periodic continuous change from contained to container to contained to container . . . in an endless round. In the virtuality of the model, the container is the contained.

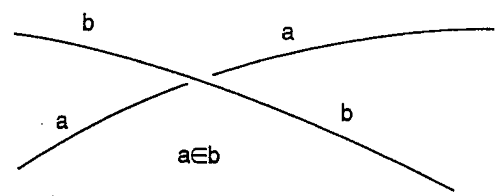


Figure 7.1

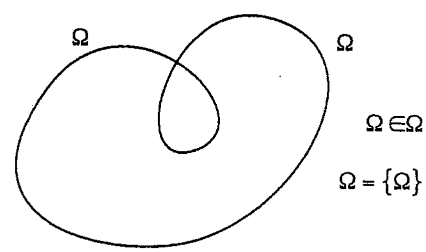


Figure 7.2

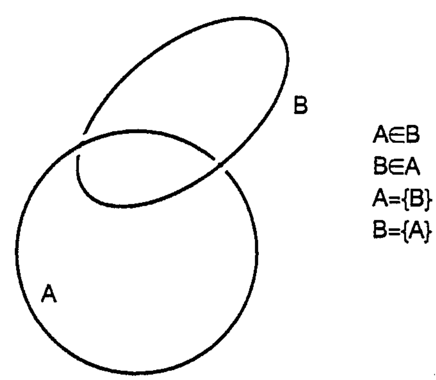


Figure 7.3

3. SELF-REFERENCE, FIXED POINTS AND INFINITY

Reference requires a distinction between what refers and what is referenced.

Self-reference does not require words.
Self-reference is as simple as saying nothing.
Self-reference is as simple as saying 'I'.
Self-reference is as simple as saying 'I am I'.
Self-reference is as simple as an empty circle.
Self-reference is as simple as an arrow bent into a circle.

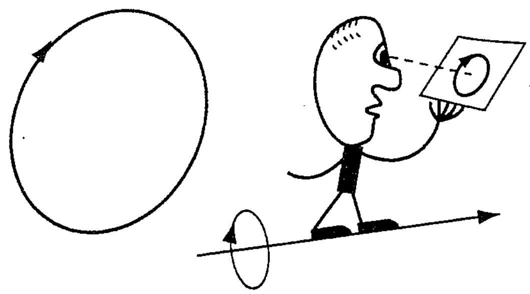


Figure 7.4

An arrow bent into a circle in the presence of an observer can be seen to point to itself.
An arrow points.
Pointing is reference.
An arrow that points to itself is self-referent.
An 'object' that points to itself with the help of an arrow is self-referent.
$A \longrightarrow A$.
An observer traveling on the self-pointing arrow will experience the same journey as an observer travelling on the infinite juxtaposition of unfolded arrows—the unidirectional unfolding of the self-referential arrow.

$$I = \longrightarrow \longrightarrow \longrightarrow \longrightarrow \ldots$$

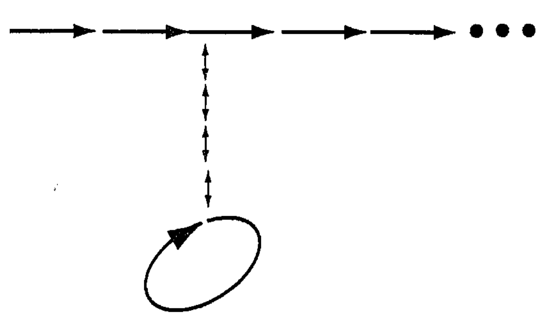


Figure 7.5

The unfolding of the self-referential arrow is a fixed point of that arrow (Kauffman, 1987b):

$$I = \longrightarrow I$$

Each self-reference gives rise to a fixed point at the level of its unfolding.

In the fixed point, I is 'the that to which reference is applied'.
I am the I you point toward with an arrow.
I am $\rightarrow$ I.
Reference is observation.
Observation is reference.
I am {the observation of $\longrightarrow$ I}.
'I am the observed relation between myself and observing myself' (von Foerster, 1981).
In this interlock lives the possibility of a return to silence.

The formal structure of paradox L=NOT(L) is the structure of a fixed point

$$\mathbf{L} = \mathbf{F}(\mathbf{L})$$

At an appropriate level of abstraction, every operator has a fixed point. Allow infinite repetition.

Form

$$\mathbf{X} = \mathbf{F}(\mathbf{F}(\mathbf{F}(\mathbf{F}(\ldots))))$$

Then

$$\mathbf{F}(\mathbf{X}) = \mathbf{F}(\mathbf{F}(\mathbf{F}(\mathbf{F}(\mathbf{F}(\ldots))))) = \mathbf{X}$$

An example from geometry is an infinite nest of rectangles as shown below.

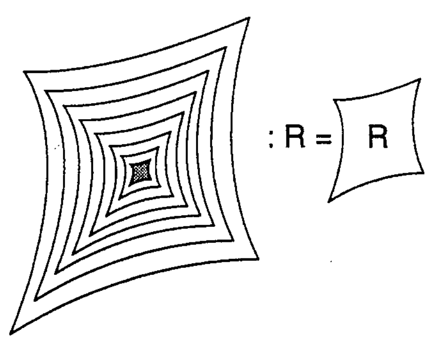


Figure 7.6

An example from arithmetic is the infinite continued fraction

$$\mathbf{f} = \mathbf{1} + \mathbf{1}/(\mathbf{1} + \mathbf{1}/(\mathbf{1} + \ldots))$$

$$\mathbf{f} = \mathbf{1} + \mathbf{1}/\mathbf{f}$$

A second example from arithmetic is the infinite continued square root. Notation: SQRT(A) denotes the square root of A: SQRT(A) = $\sqrt{A}$.

$$\mathbf{g = SQRT(1 + SQRT(1 + SQRT(1 + ...)))}$$

$$\mathbf{g = SQRT(1 + g)}$$

Both f and g are defined in the virtuality of their self-reference. In arithmetic we prove that they are equal by applying ordinary mathematics to their fixed points:

$$f = 1 + 1/f$$

$$ff = f(1 + 1/f) = f1 + f(1/f) = f + 1$$

$$f^2 = f + 1$$

$$g = SQRT(1 + g)$$

$$gg = SQRT(1 + g)SQRT(1 + g) = 1 + g$$

$$= g + 1$$

$$g^2 = 1 + g = g + 1$$

f and g are each positive solutions to $x^2 = x + 1$. Therefore, $f = g$.

$$\mathbf{1 + 1/(1 + 1/(1 + ...))}$$

$$\mathbf{= SQRT(1 + SQRT(1 + SQRT(1 + ...)))}$$

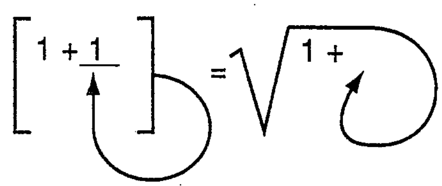


Figure 7.7

Self-reference is a virtual ingredient in the proof. Compare with Kauffman (1987b) and Spencer-Brown (1979).

Another example:

$S = 1 + a + a^2 + a^3 + a^4 + ...$
$S = 1 + a(1 + a + a^2 + a^3 + ...)$
$S = 1 + aS$
$S - aS = 1$
$S(1 - a) = 1$
$S = 1/(1 - a)$.

An example in classical mathematics is Euler's formula for e^x:

$$\mathbf{e^x = (1 + x/\infty)^{\infty}}$$

The virtuality lives in the use of the symbol for infinity. Infinity is self-referential:

$$\infty = 1 + 1 + 1 + 1 + \ldots$$

$$\infty = \mathbf{1} + \infty$$

Euler's formula is interpreted as the limit of $(1 + 1/N)^N$ as N tends to infinity. But the symbol ∞ can be used as an element in the language of arithmetic and its transformations.

Let us transform a formula of Euler. Euler's famous formula is a relationship between e, i, π, 0 and 1.

$$\mathbf{e^{i\pi} + 1 = 0}$$

Here π is pi, the ratio of the circumference of a circle to its diameter. i is the imaginary unit: $i^2 = -1$.

$$\mathbf{e^{i\pi}} = (1 + i\pi/\infty)^{\infty}$$

$$\mathbf{e^{i\pi}} = -1$$

$$(\mathbf{1 + i}\pi/\infty)^{\infty} = -1$$

$$((\mathbf{1 + i}\pi/\infty)^{\infty})^{1/\infty} = (-1)^{1/\infty}$$

$$\mathbf{1 + i}\pi/\infty = (-1)^{1/\infty}$$

$$\mathbf{i}\pi/\infty = (-1)^{1/\infty} - 1$$

$$\mathbf{i}\pi = \infty((-1)^{1/\infty} - 1)$$

$$\pi = \infty[((-1)^{1/\infty} - 1)/\mathbf{i}]$$

Here is a mystical formula for pi, derived directly from Euler's formula.

In fact this formula for pi can actually be used to calculate its numerical approximations. The key is to consider the roots $(-1)^{1/M}$ of –1 for integers M that are powers of 2 and to apply the basic formula SQRT(A + Bi) = SQRT((1+A)/2) + i E(b) SQRT((1–A)/2)) when $A^2 + B^2 = 1$ and E(B) is 1 or –1 according as B is positive or negative. We omit the details of this part.

The real point is that *the mystical formula is actually a true statement about the nature of pi*. In order to understand this, we must look closely at the meaning of $(-1)^{1/\infty}$.

Leibniz characterized **i = SQRT(–1) = $(-1)^{1/2}$** as an *amphibian between being and non-being*.

i is not real.
All real numbers have positive squares.
i^2 is –1.
–1 is negative unity, the symbol for non-being.
+1 is positive unity, the symbol for being.
i intermediates between being and non-being.

i itself is the embodiment of a Liar paradox where

'NOT'(X) = –1/X
'NOT' 'NOT' (X) = –1/(-1/X)) = – – X = X
i = –1/i
i = 'NOT' i

Gauss depicted this relationship in a brilliant interpretation of i as a point in the plane, occupying the unit direction on an axis perpendicular to the axis of real numbers.

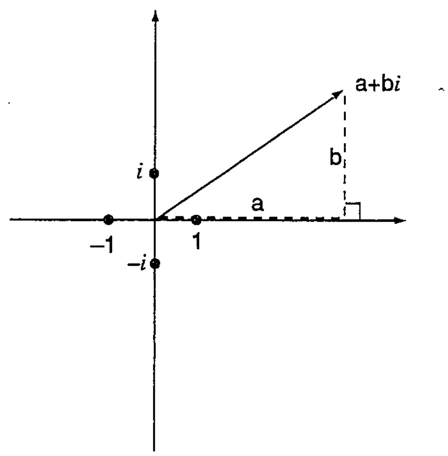


Figure 8

In the Gauss plane all numbers of the form A+Bi with A and B real are represented as points with horizontal coordinate A and vertical coordinate B. Such a point may be regarded as an arrow emanating from the origin and terminating at the point (A,B). Thus each 'complex number' A+Bi can be regarded as an arrow from the origin to a point in the plane. An extraordinary harmony with geometry comes into play. Each arrow has an angle θ with the horizontal axis and a radius R measuring its distance from the origin. Let Z = [R,θ] denote an arrow with radius R and angle θ.

It turns out that the rule for multiplying complex numbers is simply the arrow rule: *multiply the lengths and add the angles.*

$$[\mathrm{R}, \theta][\mathrm{S}, \phi] = [\mathrm{RS}, \theta + \phi]$$

This rule is perfectly in accord with the formula ii = –1. i has angle 90° and length 1. Multiplied by itself it acquires angle 180° and length 1. The number located at 180° and length 1 is exactly –1.

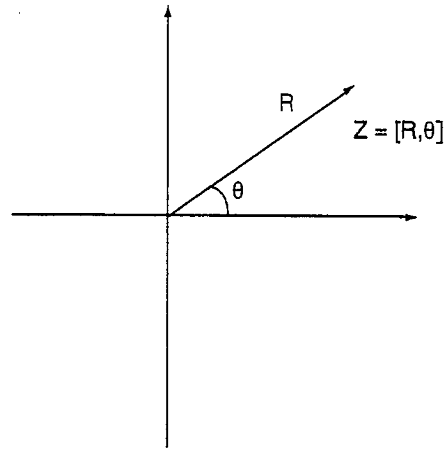


Figure 9

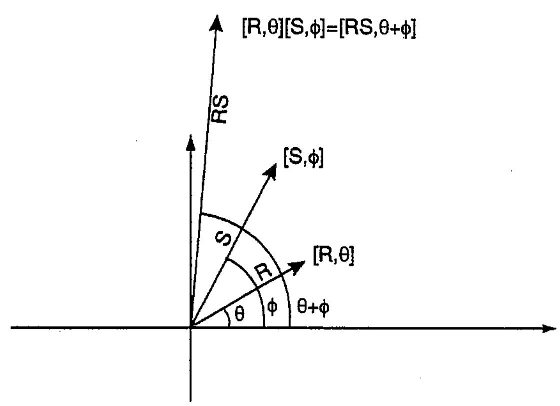


Figure 10

In general, multiplying by i will turn an arrow counterclockwise by 90°.

Now we can interpret SQRT(–1), SQRT(SQRT(–1)), SQRT(SQRT(SQRT(–1))), . . . We wish to understand the nature of the limiting form

$$(-1)^{1/\infty} = \mathbf{SQRT}(\mathbf{SQRT}(\ldots\mathbf{SQRT}(-1)\ldots))$$

–1 = [1, 180°] = [1, π] (using radian measure for the angle)
SQRT(–1) = [1, $\pi/2$]
SQRT(SQRT(–1)) = [1, $\pi/4$]
$\mathrm{SQRT}^{\mathrm{N}}(-1)$ = [1, $\pi/2^{\mathrm{N}}$]

$\mathrm{SQRT}^{\mathrm{N}}(-1)$ is a complex number whose real part is nearly 1, and whose imaginary part fits with great accuracy a $(1/2^N)$-th part of the half circumference of a circle of unit radius.

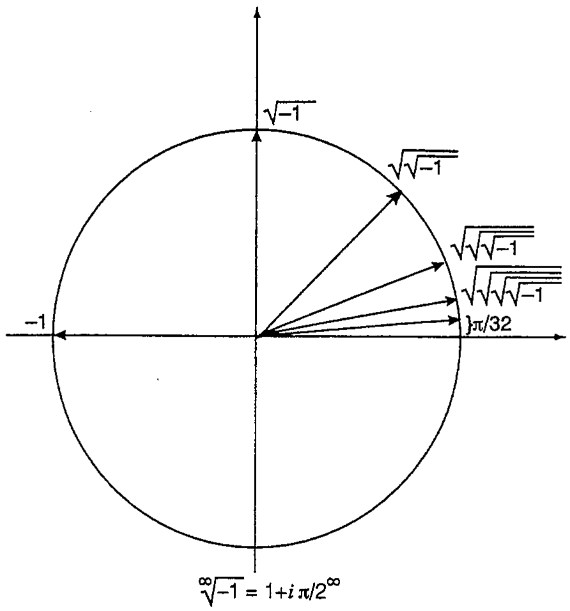


Figure 11

[SQRT^N(–1) –1] is approximately i times the length of an arc of the circle of unit radius in the plane where half the circumference has been divided into 2^N equal arcs. Therefore 2^N [(SQRT^N(–1) –1)/i] is a close approximation to π.
Symbolically,

$$\pi = \mathbf{2}^{\infty}[(\mathbf{SQRT}^{\infty}(-1) - \mathbf{1})/\mathbf{i}]$$

This is a refined version of our mystical formula for π. Note that both 2^∞ and SQRT$^\infty$(–1) are formal fixed points

$$(22^{\infty} = 2^{\infty}, \mathrm{SQRT\ SQRT}^{\infty}(-1) = \mathrm{SQRT}^{\infty}(-1))$$

with special properties that inform this mathematical application, producing π as a combination of self-referential forms. π is an interference pattern in the elementary references of any observer. Euler's original mystery formula

$$e^{i\pi} + 1 = 0$$

can now be seen as a direct consequence (read the story backwards) of the geometry inherent in $\pi = 2^\infty$ [(SQRT$^\infty$(–1) –1)/i].

4. FIXED POINTS AND THE RUSSELL SET

Infinite repetition is the life-blood of recursion.
Elemental self-reference (I am I) partakes of an interlock that does not need endless elaboration.
The general fixed point for F can be produced without infinite self-application of F to itself. This is the fixed point theorem of Church and Curry (Barendregt, 1984).

Let GX = F(XX)

Then GG = F(GG)

GG is a fixed point for F.
This production of a fixed point is the logical analogue of the production of SQRT(2) by drawing the diagonal of a square of unit side.

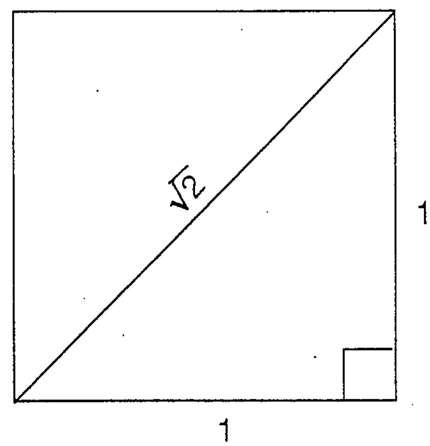


Figure 11.1

The leap to infinity is accomplished in an instant, and the observer must look back to see the vast chasm avoided by taking a step in a different conceptual dimension.
Criticism of a leap is appropriate once it has been leapt.

GX = F(XX).
What does this mean? What is F?
Note that you never asked this before when we made g = F(F(F(F...))).
In the old days F was anything that could be applied to anything else in the form F(Y).
It could just be a box to put around Y.
It could be the formalism of some arcane mathematics from the planet Tralfamidor.

To write XY assumes that the X and the Y are capable of being juxtaposed.

To write GX = F(XX) means exactly what it says.

Take X.
Duplicate it to form XX.
Put that inside F( ).
The formula GX = F(XX) defines the action of G.

(Careful now.)

The *equality sign* in GX = F(XX) is a prescription for the action 'Duplicate X and put the double copy inside F( ).'

Most of the time the application of this formula is harmless. However, if G is applied to itself, then the duplicate GG is a symbol for the application of G to itself and so the equation GG=F(GG) is an equation for action. GG=F(GG) = F(F(GG)) = F(F(F(GG))) = . . . Thus GG is a generator of recursion.

Now you may say 'Why not interpret the equals sign as a sign of identity?'

Then GG=F(GG) and GG 'has a copy of itself inside itself'.

Then the definition GX=F(XX) asserts that GX *is* F(XX).

Does G exist?

We are reminded of Anselm's proof of the existence of God.

Hypothesis. Existence is greater than non-existence.

Definition. God is the That of which nothing greater can be conceived.

Proof. If God did not exist it would nevertheless be possible to conceive of a God that did exist. Since an existent God would be greater than a non-existent God, the non-existence of God contradicts Hypothesis.

Therefore God exists.

Let us return to G. Consider the interpretation: XY means Y is a member of X.

Let ~X denote 'not X'.

RX = ~XX defines R by the sentence 'X is a member of R exactly when X is not a member of X.'

R is the Russell set.

The substitution RR = ~RR tells us *R is a member of R exactly when R is not a member of R.*

Does RR exist?

Of course RR does exist.

RR exists as a concept whose extension can never be fully realized.

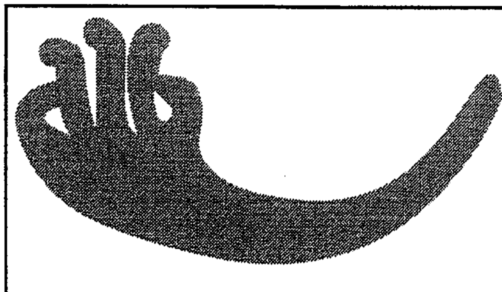
RR exists in the structure of the Möbius band.

RR exists in the process that would always encompass what is into what can be.

Paradox is the generator of time and space.

5. THE MAGRITTE SHIFT

A famous painting by Rene Magritte shows a realistic drawing of a pipe and underneath the drawing (within the frame of the painting) are the words 'Ceci n'est pas une pipe.'



Ceci n'est pas une pipe

Figure 12

It is quite common to interpret this sentence as referring to the drawing of the pipe, as though the painting was saying

'A drawing of a pipe is not a (real) pipe.'
or
'The map is not the territory.'
Of course the map is not the territory
if only we could manage to articulate the distinction between the two.

Of course the map *is* the territory.
Territory is itself a map.
Reality is identical with the appearance of reality.
Universe is what there would be if there could be anything at all.
Universe is the map of timespace process,
Flower of Nothing in the Void of perception.

It may help to point out that the sentence written so beautifully beneath the pipe (Pardonnez moi, the drawing of the pipe.) does not necessarily refer to the drawn pipe. It can refer to the canvas, the frame, the viewer or even to itself.

But what did Magritte do? Did he first write the fated phrase on a blank canvas? Or did he draw the pipe and then write the phrase below it? Let us assume the latter and thereby tease a particular strand from the Magritte nexus.

This Magritte, from the multitude of Magrittes and their not-pipes, drew the pipe first.

And held the phrase in his mind all the while.

'Ceci n'est pas une pipe.', 'Ceci n'est pas une pipe.', . . .

Always referring to the drawing.

Then Magritte reached into his 'mind' and drew out the phrase from internal speech to paint on canvas, appending the description of his creation to the creation itself.

The painting is born. What is its name?

The name of the painting of the pipe without the phrase was

'Ceci n'est pas une pipe.'

The inscribed painting has a new name that goes beyond speech.

The name of an object
(whose name is a description of what it is not)
has been added to the object.
The new object has no name.
Will you allow the old name once again?
Perhaps with brackets around it?

What do we do?

I learned to call you Heinz.

Whenever I see you, you are not just a person unknown.

You are Heinz.

In my imagination, your name is over there with you,

and it is in my mind.

I meet you

and your name splits.

One part is there with my perception of your body in space.

One part is in my speech.

Should I call the part that is in my speech 'Heinz myself' or 'Heinz Meta' or 'Heinz Magritte'?

Heinz $\longrightarrow$ the person
Heinz Myself $\longrightarrow$ Heinz the person
Let us denote this
Meta Heinz
by Heinz M.

In English we use Heinz for both circumstances.

Every reference that is meaningful to an observer has the name split and shifted in just this way.

The map and the territory are entangled.

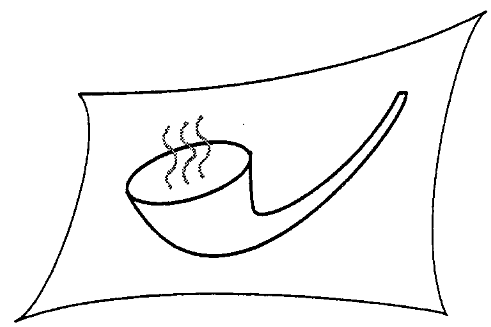
The involvement of the observer is the resonance of name and shift.

Discourse is condensed by the collapse of name and shifted name:

Heinz M = Heinz,
and indeed it is so.
My Heinz is Heinz!

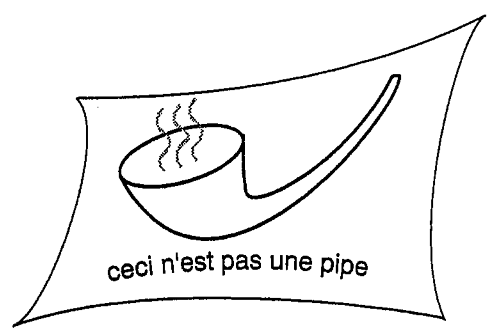
In Magritte, we had first the painting and its reference.

Ceci n'est pas une pipe. $\longrightarrow$



Magritte performs the shift by his inscription.

Ceci n'est pas une pipe. M(agrittte) $\longrightarrow$



We formalize the pattern of this *Magritte Shift* as follows.

Let $A \longrightarrow B$ denote a reference of A to B.

The structure of A and B is left open.

Define the **Shift** (see Kauffman, 1994) of this reference to be the reference

$$AM \longrightarrow AB.$$

The shift of $A \longrightarrow B$ is $AM \longrightarrow AB$.

Theorem. Let **I** denote the name of the operator in the Magritte shift. Then **IM** $\longrightarrow$ **IM**. Hence **IM** refers to itself.

Proof. Since I is the name of M, we have $I \longrightarrow M$. The shift of this reference is $IM \longrightarrow IM$. This completes the proof.

It is our contention that the formal self-reference exhibited in this theorem articulates personal self-reference. When I say 'I' there is an imaginary separation of self into the roles of 'self that names' and 'self that is named'. Here is an evocation of that circumstance.

{Silence}
I.
I say I.
I am the one who says I.
I myself say I.

'I' imagine that it is possible to divide myself into a part that sees and a part that is seen.

Each is 'I'.
Separate yet the same.

Let I denote the one who sees.
(Seeing is a form of reference.)

Let myself be that which is seen.
Let M denote myself.

I am myself.
I see myself.
I refer to myself.

$I \longrightarrow M.$

Myself performs the shift.

$A \longrightarrow B$
$AM \longrightarrow AB.$

Shift entangles seer with the seen.

$I \longrightarrow M$
$IM \longrightarrow IM$

I myself am I myself.
I am that I am.

This is the linguistic stability of self-reference.

6. THE MAGRITTE SHIFT AND GÖDEL'S THEOREM

The Magritte shift (see section 5) is right in between the general fixed point construction of section 4 and the formal structure of Gödel's incompleteness Theorem (Gödel, 1931). The shift is the underlying logical structure of the Gödel Theorem.

Recall that **Shift**$(A \longrightarrow B) = AM \longrightarrow AB$.
Whence **Shift**$(I \longrightarrow M) = IM \longrightarrow IM$.
The shift achieves self-reference in a minimal formal language.

We now place the shift in a mathematical context. In this context the objects of reference are statements and texts in a formal language. We are particularly concerned with those texts that have one designated 'free variable'. In mathematics this is quite common. For example,

$$S = \text{'U is a prime number.'}$$

The free variable in S is U. S is true for some U and false for others and imaginary for those U that are not numbers. For example, let U= 'Marilyn Monroe' for an imaginary value, U=8 for a false value and U=17 for a true value. Given a statement with a free variable U, we can substitute another text in place of the variable U.

This situation of statements with free variables does not occur directly in spoken English. However, there are close relatives. For example, consider the phrases

'Time flies like an arrow!'
'Fruit flies like a banana!'

These can be regarded as the result of substitution into the free variables U and V in the phrase:
'U flies like a V.'
Other substitutions are possible:

We will also assume that every text S is referenced by a positive integer $g = \langle S \rangle$.

This number is called the Gödel number of the text S. We assume that one can decipher the text S from its number $g = \langle S \rangle$. Such methods of coding and decoding are not hard to devise. Word processors and other devices use similar tricks all the time. We write

$$g = \langle S \rangle \longrightarrow S$$

to denote the reference of the code number to its corresponding text.

Now given $g \longrightarrow S(U)$ where U is a free variable in S, we can substitute g into S to form the statement S(g). S(g) will have its own code number.

Let $gM = \langle S(g) \rangle$ denote the code number of the result of substituting the code number of S(U) into S(U).

Then $gM \longrightarrow S(g)$ when $g \longrightarrow S(U)$.

We have defined a structure of substitution in texts and free variables that mimics precisely the Magritte shift:

$$\textbf{If } g \longrightarrow S(U) \textbf{ then}$$
$$gM \longrightarrow S(g)$$

This means that it is possible to obtain statements that talk about their own Gödel numbers by considering the Gödel number of the statements of the form S(UM) and performing a shift.

If $g \longrightarrow S(UM)$ then $gM \longrightarrow S(gM)$.

Hence S(gM) refers to its own Gödel number!

Now all we have to do is let S(U) state that there is no proof of the statement obtained by decoding U. Then S(gM) (as above) says that there is no proof of itself! If the formal system is consistent then it cannot demonstrate this statement even though 'we' the observers of the formal system have indeed proved exactly this! Proof and demonstration are thereby shown to be distinct.

Gödel's Incompleteness Theorem

Any consistent formal system rich enough to discuss and embody the properties of standard integer arithmetic is incomplete. There are theorems of arithmetic that are true but unprovable by the given formal system.

(At this point the author splits into a part that writes and a part that reads, trying to understand what the writer has written. The author is Lou; the reader will be called Jeremy.)

Jeremy. Wait. Wait! That was too fast. Let's go back to S(UM). Let's say that $g \longrightarrow P(U)$ where U is a free variable in P. Then you defined gM to the code number of P(g) so that $gM \longrightarrow P(g)$. But what is UM where U is a free variable?

Lou. Right. UM is the code number of the result of substituting U for the free variable in a proposition whose code is U. In other words, UM is a function of U, and if you give UM a number g for U, then gM will return to you a number that is the code number of the proposition obtained by substituting the number g into the free variable of the original proposition. Of course if 13 is not the code number of a statement with a free variable, then13M will not have this meaning and we will have to assign a special value like gM = 'TILT' whenever the decoding of g has no free variable.

Jeremy. So if $17 \longrightarrow$ 'U is prime.', then $17M \longrightarrow$ '17 is prime.' and all I would need to get the value of 17M would be a knowledge of the procedure for encoding statements as numbers.

Lou. Yep.

Jeremy. What about $137 \longrightarrow$ 'UM is prime.' ? Then $137M \longrightarrow$ '137M is prime.'

Is the 137M on the right a number? If so, I think that I do not know how to compute it, even in principle, because it is by definition the code number of the statement that I am working with! Are you trying to swindle me?!

Lou. No, I am not swindling you. All the cars on this lot are nearly new and driven only around Möbius bands on national holidays. The answer to the question is this. In '137M is prime' the 137M is not a numeral; it is just literally the symbol string 137M. This is like writing $(137)^2 +1$ when we mean the number 1782, or $\sin(\pi/2)$ when we mean 1. In the formal system, just as in the rest of mathematics, it is possible to refer to numbers indirectly by using a functional notation. On the other hand, the 137M on the left is our abbreviation for the actual value of the code number of '137M is prime.' This is not circular, it is just a laxity in notation. We could say that {gM} will be the actual value of gM. Then

$$g \longrightarrow S(U)$$
$$\{gM\} \longrightarrow S(g)$$
$$\text{and}$$
$$g \longrightarrow P(UM)$$
$$\{gM\} \longrightarrow P(gM)$$

Now P speaks indirectly about the number {gM} and this number is indeed the code number of P(gM). However, this matter about naming numbers is probably best solved by allowing us to identify gM and {gM}. That is, since I can refer indirectly to numbers with algorithms that generate them, there is usually no confusion in letting such algorithms be the names of the numbers. In the rest of this discussion I will deliberately make this identification and leave it to you to sort it out.

Jeremy. Well, that begins to clear up some of these things. But what is this about proof and demonstration? How do statements in your formal system discuss proof?

Lou. Ah, you made a natural slip there. My formal system knows nothing about proof. It only can know about demonstrations. A demonstration is a special text such that each statement in the text can be justified by either reference to a specific assumption that is made at the beginning or by the rules and axioms of the formal system. The last line of such a text is the statement we intended to demonstrate. In other words *a demonstration is a proof that is written within the confines of the formal system.* If P is the text of the proof and Q is the conclusion of the proof, then I will write $\langle P \rangle >> \langle Q \rangle$ to denote this relation between the code numbers of P and of Q. To say that P is a proof of Q is to say that the numbers $\langle P \rangle$ and $\langle Q \rangle$ have a certain complicated relationship with each other. To find this relationship, you have to decode each one and check step by step that P is indeed a demonstration of Q. In principle this is not different than saying something like '89 is a number that is obtained by starting with 1 (and 1), forming a series of numbers by making the next number the sum of the previous two numbers.' The statement $\langle P \rangle >> \langle Q \rangle$ is a statement about numbers that our formal system can handle.

Jeremy. I get it. But doesn't this leave the formal system a bit impoverished in regard to proofs? For example, suppose I define **Bg** to be the statement that the decoding of g is a statement provable in the formal system? In other words

Bg = [There exists P such that $\langle$P$\rangle$ >> g]

Now consider the statement **B**$\langle$**Bg**$\rangle$. Suppose that I hand this to your formal system as an assumption. I bet that it can't create a demonstration of **Bg** from it!

Lou. You are right. In most cases the formal system can't demonstrate **Bg** from **B**$\langle$**Bg**$\rangle$ because **B**$\langle$**Bg**$\rangle$ only asserts the existence of a demonstration of **Bg**, but it does not give us any specific way to write one down. In this way the formal system is a good deal more skeptical that you or me!

Incidentally, you might note what happens when we 'Gödelize' the statement **Bg**. We get

B(UM) = [There exists g such that g >> UM]

If **a** $\longrightarrow$ **B(UM)** then **aM** $\longrightarrow$ **B(aM)**.

B(aM) asserts its own demonstrability in the formal system.

Is **B(aM)** demonstrable??

Jeremy. Beats me.

Lou. Well the answer is yes! **B(aM)** is demonstrable. The statement that asserts its own provability is provable!

Jeremy. That's hard to believe.

Lou. It is a famous result of Löb (1955). Löb found a way to tame a paradox to do it. Just as Gödel tamed the Liar Paradox to prove his incompleteness Theorem, Löb tamed the Löb Paradox.

7. LÖB'S PARADOX

This section continues the dialogue between Jeremy and Lou. Lou begins with a demonstration of the Löb paradox (cf. Laraudogoitia, 1990).

Lou. You know after a person has done mathematics for a time, it becomes easier and easier to prove theorems. Eventually, I realized that I could, with slight effort, prove any theorem that interested me. Perhaps, you would like to enlist my services. I would be happy to prove any results that you happen to need.

Jeremy. OK. How about a proof of the Fermat Conjecture?

Lou. Sure. My proofs are all content-free, so I needn't even state this conjecture. We will just go ahead and prove it. Consider the following statement:

S=[If this statement is true then the Fermat Conjecture is correct.]

If S is true then it follows from S that the Fermat Conjecture is correct. But this is what S says. Therefore what S says is correct and hence the Fermat Conjecture is correct. QED.

Jeremy. Your proof is certainly short and easy, but it doesn't give me any insight into the Fermat Conjecture. Furthermore, it seems to me you could just as well argue that the Fermat Conjecture is false. Your statement is of the form, S = 'S implies A.' If S is true then A must be true since 'T implies F' is false. If S is false then S is true since 'F implies A' is always true. Therefore S is never false. I think that the problem is in the self-reference of S.

Lou. Well you may be right, but we can control this self-reference and get a version of S inside our favorite Gödelian formal system. Let's see if we can use it to throw a monkey wrench into the works!

Jeremy. That is a fine diabolical idea.

Lou. Do you recall **B(UM)** from the previous section?

Let **g** ⟶ **[B(UM) implies A]**.
Then **gM** ⟶ **[B(gM) implies A]**.
Let **L = [B(gM) implies A]** so that **gM** ⟶ **L**.
Then L says that the demonstrability of L implies A and
L = [B⟨L⟩ implies A].

This is our analog of the Löb sentence.

Jeremy. Now let's see if we can make the formal system demonstrate A. No good. I have no way to get started writing a demonstraton of A. I can't just play with truth values and interpretations in the formal system. I have to start somewhere specific and start making a demonstration.

Lou. Well, some progress is possible. First of all we can prove the following Lemma.

Lemma. **[B⟨L⟩ implies B⟨A⟩]** is demonstrable in the formal system.

Remark. The following facts about B can be shown to hold for the formal system (cf. Mendelson, 1987, p. 167):

We use the abbreviation
Dem. B for *[B is demonstrable in the formal system]*.

(1) **Dem**. P implies **Dem**. B⟨P⟩.
(2) **Dem**. B⟨P implies Q⟩ implies (B⟨P⟩ implies B⟨Q⟩).
(3) **Dem**. B⟨P⟩ implies B⟨B⟨P⟩⟩.

Each of these is a direction that the formal system can handle because the hypothesis already assumes a given text that proves P or P implies Q. We will use these properties in the proof.

Proof of Lemma.
We have L = [B(⟨L⟩) implies A].
Therefore
Dem. L implies [B(⟨L⟩) implies A]. (This is a consequence of [L implies L].)
Dem. B⟨L⟩ implies [B(B(⟨L⟩)) implies B⟨A⟩] (by 2).
Dem. B⟨L⟩ implies **Dem**. B⟨B⟨L⟩⟩ (by 1).
Dem. [B⟨L⟩ implies B⟨A⟩] (modus ponens on previous two lines).
This completes the proof of the Lemma.

Now we get the following fantastic theorem of Löb:

Löb Theorem. If **Dem**. [B⟨A⟩ implies A] , then **Dem**. A.
Proof.
Dem. [B⟨L⟩ implies B⟨A⟩] (by the Lemma).
Dem. [B⟨A⟩ implies A] (by hypothesis).
Therefore **Dem**. [B⟨L⟩ implies A] (modus ponens).
But L = [B⟨L⟩ implies A].
Therefore **Dem**. L. Hence **Dem**. B⟨L⟩.
Therefore **Dem**. A (modus ponens).

Jeremy. This really underlines the fact that one can seldom prove

B⟨A⟩ implies A

inside the formal system. If we could always do this, then the system would be inconsistent. The proof of the Löb Theorem is actually a transcription of the Löb Paradox into metamathematical language that goes back and forth across the boundaries between the formal system and our observation of it through code numbers. That mode of observation is available both to us and to the formal system, but we get to argue more hypothetically in constructing proofs than the formal system can in constructing demonstrations.

Lou. Löb's Theorem gives us an immediate proof that the self-affirming statement is provable in the formal system. This statement, call it A, has the property that $A = B\langle A \rangle$. Hence **Dem.** $[B\langle A \rangle$ implies A] and therefore **Dem.** A.

Jeremy. With Löb's Theorem and Gödel's Theorem in place it becomes clear that genuine logical paradoxes have deep consequences for the study of mathematics and logic. Paradoxes can actually be used to prove theorems that ordinary mathematics and logic can never know. I think that we should call the logic of paradox *virtual logic* and write a paper about these insights!

Lou. I happen to have information that we are discussants in a paper on that very topic, so your suggestion may have come to pass. After all, *paradox generates time* (cf. Spencer-Brown, 1979, Ch. 11) and with plenty of paradoxes there may be time enough. However, I am a bit confused about your claims for this 'virtual logic'. Do you assert that it has the capacity to transcend what can be accomplished by ordinary reasoning?

Jeremy. Aha! You think there is such a thing as 'ordinary reasoning'!?

Lou. Well, of course. Even Löb's and Gödel's Theorems are to be found in the handbooks of mathematical logic. They are clever constructs, but they are proved by ordinary reasoning.

Jeremy. Ordinary reasoning is not ordinary at all. It is exactly in the hope of understanding understanding that we set out on a voyage toward virtual logic. Gödel's Theorem shows, by a reasoning that all mathematicians can follow, that *reasoning itself cannnot be confined to any particular set of rules,* not if it is to be a reason powerful enough to handle numbers. So we are left out here in the void, forced to create creation, rationalize reason, cogitate cognition and understand our own understanding. What we call 'ordinary reason' is itself a paradox. Reason itself is not at all reasonable! And that is what those handbooks of mathematical logic are really saying. Each new mathematical construction, each new distinction, each theorem is an act of creation. Ordinary reason itself is virtual.

Lou. Yes. But those handbooks have a credo. They insist that the discussion be always open to question, open to asking why, asking for the reasons behind any step, asking for clarity of structure and design.

Jeremy. I agree completely. The credo of clarity is not ordinary. It goes beyond reason into a world of beauty, communication and possibility. If you want to call that ordinary reason, I will call it 'nothing special' and we can go look at the stars.

Lou: Perhaps a comet.

And we are not separate at all.
The observer is the observed.
The map is the territory.
And our tale is just begun.

REFERENCES

Barendregt, H.F. (1984). *The Lambda Calculus: Its Syntax and Semantics*, North-Holland, Amsterdam.

Berkowitz, G.C., Greenberg, D.R., and White, C.A. (1988). An approach to a mathematics of phenomena: canonical aspects of reentrant form eigenbehaviour in the extended calculus of indications. *Cybernetics and Systems* **19**, 123–167.

Gödel, K. (1931). *On Formally Undecidable Propositions in 'Principia Mathematica' and Related Systems* (transl. B. Meltzer), Braithwaite, R. B. (ed.), Basic Books, New York (1962).

Kauffman, L.H. (1987a). *On Knots*, Princeton University Press, Princeton .

Kauffman, L.H. (1987b). Self reference and recursive forms. *Journal of Social and Biological Structures*, 53–72.

Kauffman, L.H. (1991, 1993). *Knots and Physics*, World Scientific.

Kauffman, L.H. (1994). Ways of the game: play and position play. *Cybernetics and Human Knowing* **2** (3), 17–34.

Kauffman, L.H. (1995a). Knot logic. In Kauffman, L.H. (ed.), *Knots and Applications*, World Scientific, pp. 1–110.

Kauffman, L.H. (1995b). Arithmetic in the form. *Cybernetics and Systems* **26** , 1–57.

Kauffman, L.H. and Varela, F.G. (1980). Form dynamics. *Journal of Social and Biological Structures* 3, 171–206.

Laraudogoitia, J.P. (1990). This article should not be rejected by Mind. *Mind* **99**, 598–599.

Löb, M.H. (1955). Solution of a problem of Leon Henkin. *Journal of Symbolic Logic* **20**, 115–118.

Mendelson, E. (1987). *Introduction to Mathematical Logic*, Wadsworth and Brooks/Cole .

Pedretti, A. (ed.) (1979). *Self-reference on the Isle of Wight*, Transcripts of the first International Conference on Self-reference, 24–27 August, Princelet Editions, London.

Pedretti, A. (1981). *Cybernetics of Language,* Princelet Editions, London. (In the I of Language, 2nd edn of *Cybernetics of Language,* to appear).
Spencer-Brown, G. (1979). *Laws of Form,* E.P. Dutton.
Uribe, R. (1995). *Tractatus Paradoxico-Philosophicus* (preprint, University of Illinois at Urbana–Champaign).
von Foerster, H. (1981). Objects: tokens for (Eigen-) behaviours (reprinted) in *Observing Systems,* Intersystems, pp. 274—285.