Gödelian Explorations

Gary Mar

2030-01-01

Table of contents

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

1 For Editors' Use

This is a testing page left for editorial purposes and will not be included in the publication.

1.1 TODO: Chapter Proposal

1.1.1 Yutong: What is a ... *motivic model*?

Denef, J. & Loeser, F. 1999, October 22. Definable sets, motives and p-adic integrals. (http://arxiv.org/abs/math/9910107) (Accessed 2023-11-18)

(Denef & Loeser 1999) says

$$\chi\left(\left\{\varphi := x \neq 0 \wedge (\exists y)\left(x = y^2\right)\right\}\right) = \frac{1}{2}([\mathbb{L}] - \mathbb{1})$$
$$\varphi(\mathbb{F}) \leftrightarrow \zeta\left(\frac{1}{2}([\mathbb{L}] - \mathbb{1})\right)$$

for $\mathbb{F}$ with a sufficiently large characteristics from compactness, as

> ... [over] finite fields of characteristic > 2, half of the units are squares ...
>
> (— Cluckers & Nicaise & Sebag 2011: 23)

Cluckers, Raf & Nicaise, Johannes & Sebag, Julien (eds.). 2011. *[I] Motivic integration and its interactions with model theory and non-Archimedean geometry* Vol. 1. Cambridge ; New York: Cambridge University Press.

Brown, Francis. 2017, February 1. Feynman Amplitudes and Cosmic Galois group. (http://arxiv.org/abs/1512.06409) (Accessed 2023-11-16)

1.1.2 Yutong: What is a ... *motivic path integral*?

(Brown 2017) says

$$K(x, y) = \int \exp(iS(\phi))\, D\phi.$$

1.2 Format Cheatsheet

Quarto's official docs

1.2.1 Quiver CD

1.2.1.1 SVG

$$X \times T = \mathfrak{X}_{\overline{\eta}} \qquad T = \mathfrak{X}_R$$
$$K \boxtimes L = \mathbb{1}_{\dots} \qquad L = \mathbb{1}_T$$

$$\begin{tikzcd}
X \times_S T \arrow[r, "\alpha"] \arrow[d, "q"'] \arrow[dr, phantom, "\lrcorner", very near start] & T \arrow[d, "p"] \\
X \arrow[r, "\beta"'] & S
\end{tikzcd}$$

$$X = \overline{\eta}$$
$$K = \mathbb{1}_X$$

1.2.1.2 iframe

1.2.2 Indented Text

> The other metaphorical analogue to Gödel's Theorem which I find provocative suggests that ultimately, we cannot understand our own minds/brains.... All the limitative theorems of mathematics and the theory of computation suggest that once your ability to represent your own structure has reached a certain critical point, that is the kiss of death: it guarantees that you can never represent yourself totally.

1.2.3 Color CSS

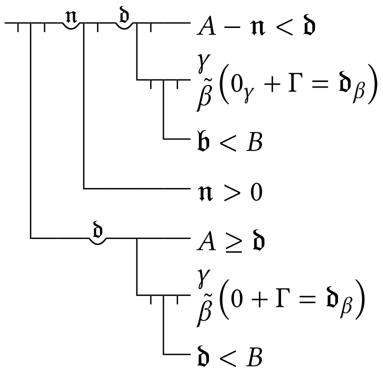
Prof. Mar's favorite color

1.2.4 TODO: Conditional

Some text

Will only appear in PDF.

1.2.5 Frege



1.2.6 Tufte

- twoside class
- tufte.qmd
- auto break sidenotes

1.2.7 Graphviz with MathJax

Actually, this is horrible as it requires manual coding of size and position.

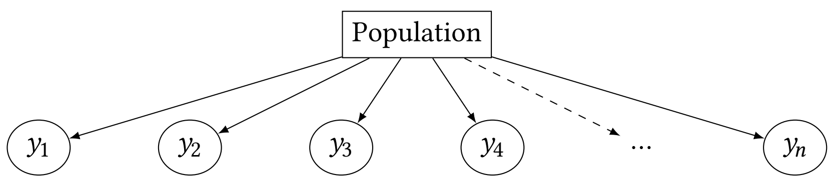
- HTML: Graphviz + MathJax
- TeX

```
//| code-fold: true

d3 = require("d3@7", "d3-graphviz@2");
mathjaxBlob = (await fetch('https://cdn.jsdelivr.net/npm/mathjax@3/es5/tex-svg.js')).blob();
mj = {
  delete window.MathJax
  delete window.ContextMenu
  window.MathJax = {
    loader: {load: ['output/svg']},
```

```
  }
  const MathJax = await import(URL.createObjectURL(await mathjaxBlob))
    .then(() => window.MathJax );
  return MathJax;
};

seir_graph = {
  let graph = d3.create('div').style('width', `${width}px`);

  // Here's the source code describing the graph to graphviz.
  // Note that nodes and edge labels contain LaTeX code that
  // will be passed to MathJax. I guess it gets piped through
  // a couple of things; hence, the double escape leading to
  // quadruple backslashes \\\\.
  let source_code = `digraph {
      S [pos="0,0!"]
      E [pos="2.7,0!"]
      I_1 [pos="4,0!"]
      I_2 [pos="6,0.5!"]
      I_3 [pos="8,0!"]
      D [pos="10,0!"]
      R [pos="6,-1.5!"]
      S -> E [label="\\\\beta_1 I_1 S + \\\\beta_2 I_2 S + \\\\beta_3 I_3"]
      E -> I_1 [label="\\\\alpha E"]
      I_1 -> I_2 [label="p_1 I_1"]
      I_2 -> I_3 [label="p_2 I_2"]
      I_3-> D [label="\\\\mu I_3"]
      I_1 -> R [label="\\\\gamma_1 I_1"]
      I_2 -> R [label="\\\\gamma_2 I_2"]
      I_3 -> R [label="\\\\gamma_3 I_3"]
    }`;
  d3.graphviz(graph.node())
    .width(width)
    .fit(true) // Doesn't quite work; see transform in penultimate line.
    .zoom(false) // Re-transform for fit breaks the zoom.
    .engine('neato')
    .renderDot(source_code);

  // The image is completely contained in a top level group,
  // which we're going to manipulate
  let main_group = graph.select('g');

  // Don't really want the title
  main_group.select('title').remove();

  // Typeset the nodes
  main_group.selectAll('.node').each(function(e, i) {
    let text = d3.select(this).select('text');
    if (text.node() != null) {
      let x = parseFloat(text.attr('x'));
      let y = parseFloat(text.attr('y'));
      let tex_group = main_group
        .append('g')
        .attr('transform', `translate(${x - 8} ${y - 10})`)
        .append(() =>
          mj.tex2svg(String.raw`${text.text()}`).querySelector("svg")
```

```
      );
      text.remove();
    }
  });

  // Placement of the typeset edge labels is a bit trickier. The following
  // list of shifts adjusts the placement of the labels from the location
  // specified by graphviz.
  let shifts = [
    [60, -30],
    [-9, -27],
    [-37, -15],
    [10, 5],
    [-20, -10],
    [10, 0],
    [-25, -12],
    [44, 5]
  ];
  main_group.selectAll('.edge').each(function(e, i) {
    let text = d3.select(this).select('text');
    if (text.node() != null) {
      let x = parseFloat(text.attr('x'));
      let y = parseFloat(text.attr('y'));
      let tex_group = main_group
        .append('g')
        .attr(
          'transform',
          `translate(${x + shifts[i][0]} ${y + shifts[i][1]}) scale(0.75)`
        )
        .append(() =>
          mj.tex2svg(String.raw`${text.text()}`).querySelector("svg")
        );
      text.remove();
    }
  });

  // There's far more space to the left of the graph than I'd expect;
  // I guess the reason is that the first, pre-shifted edge label extends
  // quite far to the left. A hacky fix is to redefine the main transform
  // to fit it a bit better. Unfortunately, this breaks zoom.
  main_group.attr('transform', `translate(-130, 204) scale(1.15)`);
  return graph.node();
}
```

1.2.7.1 Let's try the TikZJax or just plain tikz



1.2.8 TODO: mathnote

```
\def\mathnote#1{%
  \tag*{\rlap{\hspace\marginparsep{\parbox[t]{\marginparwidth}{\footnotesize#1}}}}%
}
\def\mathnotes#1{%
  \tag*{\rlap{\hspace\marginparsep\smash{\parbox[t]{\marginparwidth}{\footnotesize#1}}}}%
}
\def\mathnoteps#1{%
  \tag*{\rlap{\hspace\marginparsep{\parbox[t]{\marginparwidth}{\footnotesize#1}}}}%
}
```

https://tex.stackexchange.com/questions/120104/collectively-aligning-multiple-align-environments

macro: intertext

1.2.9 TODO: BibLaTeX with Zotero Group Library Sync?

2 LOGIC IN RETROSPECT

2.1 *Meta-Logical Reflections on Lewis Carroll's "What the Tortoise Taught Achilles"*

> *I tell how there may be a better wilderness of logic than of inconsequence. But the logic is backward, in retrospect, after the act. It must be more felt than seen ahead like prophecy. The figure a poem makes. It begins in delight and ends in wisdom.* — Robert Frost, "The Figure a Poem Makes" (1939)

In this chapter, we wish to draw some lessons from Lewis Carroll's classic logical parable "What the Tortoise Taught Achilles". These lessons, or retrospective reflections, shed light on the experience of learning logic, the epistemology or foundations of logical truth, and about the aesthetics, and art, that inspires many-splendored branches of logic.

When students struggle to learn a system of logic for the first time, they may experience a clutter of disconnected details that can be confusing, even chaotic. But, if they persevere if they continue to exercise their power of analysis, out of this chaos there may emerge a synthesis, the experience of a constellation of mutually supporting ideas coming together, reducing cognitive entropy, and forming cognitively pleasing patterns of information, a grand synthesis, the reveals the beautiful symmetries of logical laws within and a conceptual tool for mapping the starry heavens above.

One of the most well-known descriptions of this experience, or phenomenology, of learning, as experienced XXX by mathematicians and logicians, was recorded by the French mathematician and physicist Henri Poincaré (1854 - 1912):

> Just at this time I left Caen, where I was then living, to go on a geologic excursion under the auspices of the school of mines. The changes of travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step of the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry. I did not verify the idea; I should not have had time, as upon taking my seat in the omnibus, I went on with a conversation already commented, but I felt a perfect certainty. On my return to Caen, for conscience' sake I verified the result at my leisure.

This characteristic cycle of preparatory intellectual struggle, creative pause, and sudden illumination was experienced by Poincare time and again, and so he proposed an explanation: "Most striking at first is this appearance of sudden illumination, a manifest sign of long, unconscious work." The initial intellectual struggle, which *appeared* to bear no fruit, in fact planted the seeds of a cognitive puzzle with sufficient clarity so that the unconscious mind, like a computer program, could continue to search for a solution in the background, not *actively* presented to the conscious mind.

Learning logic follows a similar progression, with different levels of processing: (1) there is the *mechanical* level at which you learn the syntax, the formatting, the order of executing steps when constructing a derivation; (2) there is the level of *understanding* the meaning of the rules of inference and the forms of derivation; (3) there is the *intuitive* level in which you grasp, or have *insight* into, the flow of the entire derivation from beginning to end with one "fell swoop" of the mind. This intuition cultivates logical insight into the crux of logical problems, it cuts the Gordian knot by reducing cognitive entropy and producing kind of intellectual pleasure, characteristic of logical and mathematical inquiry.

There is also a perverse kind of *philosophical* delight in meta-logical insights into Carrollian *nonsense*, which becomes a kind of *parody* of sense when seen through the lens of symbolic logic. Like the Doubting Tortoise of "What the Tortoise Taught Achilles," logicians seem to take a perverse delight in contemplating logical paradoxes, which fascinates the logical mind, in a manner similar to how optical illusions are fascinating to one's perceptual faculties.

The Cambridge mathematician G. H. Hardy (1877 – 1947) wrote his classic essay *A Mathematician's Apology*. An *apology* is a justification or defense. Plato's *Apology* is a dialogue about his teacher Socrates's apology or defense of the philosophic way of life. Hardy is credited with reforming (or deforming) mathematical by adhering to standards of rigor characteristic of German mathematics and for promoting the conception of pure, as opposed to applied, mathematics. IN his Apology, Hardy boasted:

> I have never done anything "useful". No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world.

The boast was premature: the Hardy-Weinberg principle is useful in population genetics and the Hardy-Ramanujan asymptotic formula was used by Niels Bohr to find quantum partition functions.

Hardy was a peculiar sort of atheist. According to one story, he once set out a postcard announcing his discovery of a proof of the Riemann hypothesis as "travel insurance" before crossing the North Sea, reasoning perversely that God would not let him die with the glory of having be thought to have "solved" this most famous of open mathematical problems.

A confirmed bachelor, with a few close friends, and platonic relationships with young men who shared his interests—especially mathematics and cricket. In an interview by the mathematician Paul Erdős, Hardy unhesitatingly replied that his greatest contribution to mathematics was the discovery of Ramanujan, which he said in a lecture was " ... the one romantic incident in my life".[7]2 One of Hardy's associates was Bertrand Russell. Hardy wrote objecting to Bertrand Russell dismissal from Cambridge as a consequence of his pacificism and criticism of British politicians for involving Great Britain in World War. Russell's punishment was not that severe insofar as he was able to write while in jail *The Principle of Mathematics*, which he dedicated to this mistress Lady Ottoline Morrell.

Hardy fascination with proof surpassed his interest in friendship with persons. Hardy once told Bertrand Russell "If I could prove by logic that you would die in five minutes, I should be sorry you were going to die, but my sorrow would be very much mitigated by pleasure in the proof".[29] A junior colleague C. P. Snow, a research physicist turned novelist of English literature and noted for his lecture "The Two Cultures and the Scientific Revolution" (1959), suggested that Hardy write this essay as way to stave off suicidal depressionl) wrote his classic apology or justification for the mathematician's life. The film *The Man Who Knew Infinity* is about the Indian mathematical genius and mystic Srinivasa Ramanujan, whom Hardy discovered and collaborated with calling this episode the "one great romance" of his life.

Ono of the memorable quotations from *A Mathematician's Apology* concerns beauty:

> The mathematician's patterns, like the painter's or the poet's, must be beautiful, the ideas, like the colours or the words, must fit together in a harmonious way. Beauty is the first test: there is no permanent place in the world for ugly mathematics.

Logic, like light, is composed of many difference colors—in fact, it when logic when projected through the critical prism of analysis, reveals a rainbow or spectrum of logics—one shading into another. Just as light reveals the bluest blue, the reddest red, the greenest green, the light of logic reveals logic with paradigmatic focus—the logic of propositions, the logic of quantifier and predicates, the logic of identity and descriptions, the logic of relations, the logic of modality, the logic of time, the logic of logic or meta-logic.

As the seasons are changing, you'll witness the changing colors of the leaves, the shortening of the march of the sun across the sky, the unique and intricate symmetries of a delicate snowflake. These are symmetries and patterns you can see with the naked eye. But you have also witnessed that there is a rhythm and pattern to logical thought—patterns not always visible to the naked eye but revealed by the eye of analysis. Among these patterns are the laws of logic.

But are the laws of logic true? Are they universal for all rational minds—at all times, in all cultures, in every conceivable possible world? Is there a bottom or foundation, to all this logical thinking—some foundational *a priori* truths, or is there no bottom—that the world that rests on the back of a elephant standing on a tortoise. On what does the tortoise stand—or it is "tortoises all the way down"?

In his classic logical parable "What the Tortoise Taught Achilles" [1895], Lewis Carroll (the penname of Charles Dodgson discovered a logical paradox about justifying laws of logic.

Carroll borrows two characters from Zeno's famous paradoxes of motion—Achilles and the Tortoise. In Zeno's paradox, Achilles, the swiftest of all runners can never overtake the Tortoise, who has a head start. Why? Achilles must read the Tortoise's starting point, by which time the Tortoise has moved ahead, producing a second starting point that Achilles must reach before, overtaking the Tortoise, by which time the Tortoise have again inched ahead, and so on *ad infinitum.* Achilles, to overtake the Tortoise, must therefore traverse an *infinite* series of points in a *finite* amount of time, which, according to Zeno, is impossible.

Carroll transports these characters into a dialogue about justifying the laws of logic. Achilles is a dutiful student, ready to write down what the Tortoise, a teacher with a fondness for paradox, dictates.

Originally publishes in *Mind* in 1895, and since that time, Carroll's puzzle—like *Alice in Wonderland* and *Through the Looking Glass*—has become something of a minor classic among philosophers of logic. Philosophers have perennially been drawn to intriguing parable as a source of inspiration for philosophizing about the nature of logic.

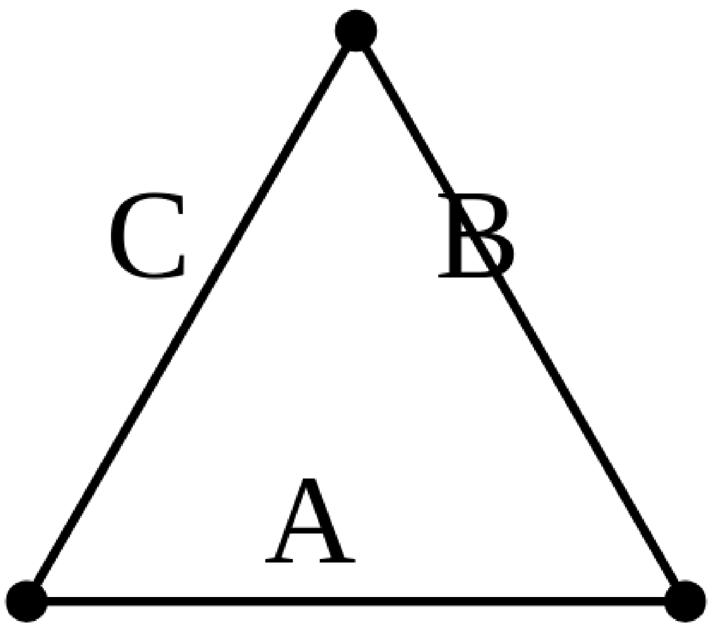
- Gilbert Ryle [1950] and Stephen Toulmin [1953] appealed to the parable to attack the *hypothetico-deductive model of science*, according to which scientific predictions are deduced from universal hypotheses together with empirical observation statements.
- Willard van Orman Quine [1936] in "Truth by Convention," his famous attack on the Rudolf Carnap's thesis of conventionalism in logic cites Carroll's parable. Turning the tables on Quine, Saul Kripke in a series of lecture and a recent on-line seminar, attended by logicians from all over the world, famously used Carroll's parable in lectures to attack Quine's empiricist thesis concerning the revisability of logic.[1]
- The logician Hao Wang [1987, pp. 203-204], a chronicler and colleague of the most famous 20th century logician Kurt Gödel, uses Carroll's parable to illustrate and explain the distinction drawn by Gödel between *intuition* and *proof.*

There is probably not one, and only one, clear and unambiguous lesson to be learned from Carroll's parable. It is likely that Carroll himself was puzzled by his discovery and used his dialogue to express a problem that he intuitively felt but could not adequately articular. Indeed, when pressed by the editor of *Mind* to explain his point, Carroll seemed to be at a loss and replied by simply iterating his paradox.

What actually takes place in Lewis Carroll's parable? At the end of Zeno's famous paradox, Achilles and the Tortoise pause for a paradoxical chat. This time the paradox is not about the impossibility of motion, but about the impossibility of justification in logic.

Imagine Achilles as the dutiful, but somewhat confused, logic student, and the Tortoise as a perversely humorous, but persistently ingenious philosophical soul with a fondness for paradox.

The Tortoise poses a logical argument for Achilles to consider about an equilateral triangle.



The argument is this: The two sides A and B of the triangle are congruent to a third side C.; therefore, sides A and B are congruent to each other.

Let’s symbolize this argument by first introducing a dictionary:

P : sides A and B of the triangle are congruent to the third side C.

Q : sides A and B are congruent to each other.

The Tortoise’s argument can be symbolized:

(1) If P, then Q

(2) P

(3)∴ Q

The Tortoise points out that a skeptic object to the argument in one of two ways: (1) he can object to the truth of the premises, or (2) he can reject the validity of the argument.

The Tortoise declares himself to be skeptic of the second sort—in other words, he grants the truth of the premises, but rejects the validity of the inference. He challenges Achilles to catch him in a contradiction. But this is easy, isn’t it? The argument is simply an instance of *modus ponens.* In other words, the Doubting Tortoise is skeptical about *modus ponens* and challenges Achilles to prove that’s illogical—that is, the Tortoise challenges Achilles to logically convict him of being illogical!

The Tortoises argument is, of course, an instance of the now the inference rule known as *modus ponens*:

$\phi \longrightarrow \psi$

ϕ

ψ

[MP for *Modus Ponens*]

Here, for sake of emphasis, we may *strengthen* the parable slightly. Let’s assume that the Tortoise not only rejects *modus ponens* but adopts the following rule of inference:

$\phi \longrightarrow \psi$

ϕ

$\sim\psi$

[MTo for *Modus Tortoise*]

Can the Tortoise consistency get away with this logical heresy?

Achilles dutifully proceeds. Since Tortoise accepts the truth of the premises ($P \longrightarrow Q$) and P, he must also accept their conjunction:

(4) (1) $\wedge$ (2).

Achilles asked the Tortoise if he accepts the conditional that if both (1) and (2) are true, then somuchy (3) be true. The Tortoise, to Achilles’s surprise, grants this conditional is also true:

(5) (1) $\wedge$ (2) $\longrightarrow$ (3)

Achilles cries, “Eureka!” thinking he’s snared the wily Tortoise in an inconsistency: from

(5) (1) $\wedge$ (2) $\longrightarrow$ (3)

(4) (1) $\wedge$ (2)

Achilles infers:

∴ (3)

by *modus ponens*! “Gotcha!” But the Tortoise begs to disagree, for he infers:

∴ ~(3) by *modus tortoise*!

Achilles doggedly pursues his line of argument. He convinces the Tortoise to accept the conjunction of (4) and (5),

(6) (4) $\wedge$ (5)

as well as the conditional premise:

(7) (4) $\wedge$ (5) $\longrightarrow$ (3)

But the problem persists: Achilles inference (3) by *modus pon*ens, but the Tortoise infers ~(3) by *modus tortoise.*

Carroll, like the dutiful Achilles, repeats the pattern a few more times for those the flat-footed, like Achilles, who have no nose for logical paradox. The Tortoise concedes:

(8) (6) $\wedge$ (7)

(9) (6) $\wedge$ (7) $\longrightarrow$ (3)

But stubbornly infers

□ ~ (3) by *modus tortoise.*

The infinite regress of logical justification, like the logical regress of Zeno's paradox of motion, has only just begun!

Many lessons have been drawn from Lewis's paradox of logical ustification. Before I drawn three such lessons, we're well-advised to recall Lewis Carroll's remarks about lessons in *Alice In Wonderland*, where alice is talking with the Gryphon, and the Mock Turtle. The Gryphon remarks:

> *"I went to the Classical master, though. He was an old crab, he was."*
>
> *"I never went to him," the Mock Turtle said with a sigh. "He taught Laughing and Grief, they used to say."*
>
> *"So he did, so he did," said the Gryphon, sighing in his turn; and both creatures hid their faces in their paws.*
>
> *"And how many hours a day did you do lessons?" said Alice, in a hurry to change the subject.*
>
> *"Ten hours the first day," said the Mock Turtle: "nine the next, and so on."*
>
> *"What a curious plan!" exclaimed Alice.*
>
> *"That's the reason they're called lessons," the Gryphon remarked: "because they lessen from day to day."*

I shall draw three lessons—the first two based on formal logic and metalogic and the third more philosophical, with echoes of the Socratic paradox of learning from Plato's dialogue the *Meno*—but I don't mean for the lessons to "lessen" in their logical or philosophical cogency.

In an article, *'If', 'So', and 'Because'* the Oxford philosopher Gilbert Ryle [1950] used Carroll's parable to deny one of the most important features of what was known as the hypothetical-deductive model of scientific explanation, namely, the requirement that universal hypotheses be included in the premises of explanatory arguments. Universal hypotheses, argued Ryle, do not act as *premises* in our explanations, but as *inference licenses* or *inference tickets.* Ryle explains:

> ...the Tortoise proved to Achilles...(the) principle of an inference cannot be one of its premises or part of its premiss. *Conclusions are drawn from premisses in accordance with principles, not from premisses that embody those principles....* In saying 'q because p' we are not just asserting but using what is expressed by "if p, then q".

Stephen E. Toulmin, in his *Introduction to the Philosophy of Science* [1953], endorsed Ryle's lesson:

> It is the same with the laws of nature. The conclusions about the world which scientists derive from laws of nature are not deduced from those laws, but rather drawn in accordance with them or inferred as applications of them....

These are powerful lessons to be drawn from such a short parable. Let's write it down:

Lesson 1. No inference rule can be eliminated in favor its corresponding logical law.

Two questions can be asked: (1) what exactly is their thesis? (2) is it true?

In order to make a *reality* (something concrete so to speak) out of something that might otherwise be too *vague*, I shall reason about a specific formal system, namely, the natural deduction system of chapter 1 in KM2. Here I use the word "*vague*", rather than "*abstract*", in opposition to what is *precise*. By now, you realize that *abstraction* can be quite precise and real.

First, we note that the logical connectives $\{\wedge, \vee, \leftrightarrow\}$ of Chapter 2 of Kalish, Montague and Mar's *Logic: Techniques of Formal Reasoning* (hereafter, KM2) could have been defined in terms of $\{\sim, \longrightarrow\}$ the logical connectives of Chapter 1:

$(\phi \wedge \psi) := \sim(\phi \longrightarrow \sim\psi)$

$(\phi \vee \psi) := (\sim\phi \longrightarrow \psi)$

$(\phi \leftrightarrow \psi) := ((\phi \longrightarrow \psi) \wedge (\psi \longrightarrow \phi)) \equiv \sim((\phi \longrightarrow \psi) \longrightarrow \sim(\psi \longrightarrow \phi))$

Hence, we can, therefore, reduce the full propositional logic of Chapter 2 to the conditional logic of Chapter 1 via these definitions. Perhaps the exercises of Chapter were more elaborate, but logically speaking, everything we can accomplish logically in Chapter 2 can, by definitional reductions, already be accomplished using the natural deduction system of Chapter 1.

The natural deduction system of Chapter 1 consisted of 4 inference rules and 3 forms of derivation. Can this logical system be simplified even further? Yes. For example, the rule of repetition is eliminable without loss of logical power: it can be replaced by showing $\phi \longrightarrow \phi$ and applying *modus ponens*, or by two applications of *double negation*. In fact, we can replace all the inference rules, other than *modus ponens*, with the following set of logical laws:

(MT) $(\phi \longrightarrow \psi) \longrightarrow (\sim\psi \longrightarrow \sim\phi)$

(DN) $\sim\sim\phi \longrightarrow \phi$

(DN) $\phi \longrightarrow \sim\sim\phi$

It happens that these three laws can be reduced to the *converse* law of MT:

(MTc) $(\sim \phi \longrightarrow \sim \psi) \longrightarrow (\psi \longrightarrow \phi)$

These proofs are exercises at the end of Chapter 1. Therefore, the moral drawn by Ryle, and seconded by Toulmin, is too strong since logical inference rules can be eliminated without any loss of derivational theorem, by adopting the corresponding logical laws. However, Ryle and Toulmin, are partially correctly since not all logical inference rules can be eliminated for otherwise, we would not be able to deduce any interesting logical consequences other than instances of the laws. Hence, we can restate the first lesson:

> Lesson #1. *Some, but not all, logical inference rules can be eliminated, without any loss of deductive strength, by adopting the corresponding logical laws.*
>
> A similar logical lesson can be drawn about the three forms of derivation—direct, conditional, and indirect. The two forms of indirect derivation can be eliminated. Suppose have an indirect derivation of the form:

1. *Show* ϕ *j, k* ID
2. $\sim\phi$ Assume (ID)

.

.

.

j. χ

k. ~χ

It turns out that we may transform the above indirect derivation schema into a direct derivation. We replace every line subsequent to the initial 'Show' line with a subsidiary derivation of the conditional consisting of the assumption for indirect derivation as its antecedent and the given line as its consequent.

The ultimate success of this reduction depends our ability to prove a theorem (namely, T19) without using any form of indirect derivation.

Exercise: Annotate the following derivation which proves T19 without using ID.

1. *Show* (~$\phi \rightarrow \phi$) $\rightarrow \phi$ ___, CD
2. ~$\phi \rightarrow \phi$ Assume (CD)
3. *Show* ~$\phi \rightarrow$ ($\phi \rightarrow \phi$) ___, CD
4. ~ϕ Assume (CD)
5. *Show* (~$\phi \rightarrow \phi$) $\rightarrow \phi$ ___, CD
6. ~$\phi \rightarrow \phi$ Assume (CD)
7. ϕ __, ___ MT
8. ($\phi \rightarrow \phi$) 3, 4 _____
9. ~ ($\phi \rightarrow \phi$) 2, ___
10. ~~ϕ 3, 9 _____
11. ϕ 10, ___

    1. *Show* ϕ 10, DD

2. *Show* ~$\phi \rightarrow \chi$

~ϕ Assume (CD)

. .

. (*copy lines 2-j from the original indirect derivation*)

.

χ

3. *Show* ~$\phi \rightarrow$ ~ χ

~ϕ Assume (CD)

.

. (*copy lines 2-k from the original indirect derivation*)

. χ

4. *Show* $\sim\phi \rightarrow \phi$ 8, CD

5. $\sim\phi$ Assume (CD)

6. $\sim\chi$ 3, 5 MP

7. $\sim\sim\phi$ 2, 6 MT

8. ϕ 7 DN

9. *Show* $(\sim\phi \rightarrow \phi) \rightarrow \phi$ (T19 to be proved without using indirect)

10. ϕ 4, 9 MP

The first two conditional derivation (in the green boxes) can be obtained by copying the lines from the assumption for indirect derivation to the derivation of the respective sentences forming the contradiction. Assuming that T19 can be proved without using indirect derivation (in the red box), we could obtain a direct derivation of the conclusion by proving the antecedent of T19, which is accomplished in lines 5 – 8 (in the blue box). The proof of T19 without using indirect derivation is left as an exercise below. The metalogical point is that indirect derivation, which many students may have found initially counterintuitive, can be eliminated in favor of conditional derivation and direct derivation.

It turns out that this system can be reduced to just three logical laws (axiom schema) with just one form of derivation, direct derivation, and just one rule of inference *modus ponens.* This axiomatic system of conditional logic is due to Polish logician Stanisław Leśniewski (1886 – 1939).

(L1) $\phi \rightarrow (\psi \rightarrow \phi)$

(L2) $(\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \chi))$

(L3) $(\sim\phi \rightarrow \sim\psi) \rightarrow (\psi \rightarrow \phi)$

Direct derivation, *Modus Ponens*

To illustrate the austerity of the Leśniewski axiomatic system, here is a proof of theorem 1.

T1 $P \rightarrow P$

1. *Show* $P \rightarrow P$ 6, DD
2. $P \rightarrow (P \rightarrow P)$ L1
3. $(P \rightarrow ((P \rightarrow P) \rightarrow P)$ L1
4. $((P \rightarrow ((P \rightarrow P) \rightarrow P)) \rightarrow ((P \rightarrow (P \rightarrow P)) \rightarrow (P \rightarrow P))$ L2
5. $(P \rightarrow (P \rightarrow P)) \rightarrow (P \rightarrow P)$ 4, 3 MP
6. $P \rightarrow P$ 5, 2 MP

It turns out that the converse form of *modus tollens* allows one to prove the two forms of double negation as derived rules and axioms L1 and L2, together, with the proof of theorem 1, as sufficient to justify conditional derivation, which is known as the *deduction theorem.*

The founder of the Warsaw school of logic, Jan Łukasiewicz (1878–1956), describe the experience of finding the most compact formulation of a logical system as follows:

> *"As a conclusion to these remarks, I should like to sketch a picture connected with the deepest intuitive feelings I always get about logistic. This picture perhaps throws more light than any discursive exposition would on the real foundations from which this science grows (at least so far as I am concerned). Whenever I am occupied even with the tiniest logistical problem, e.g., trying to find the shortest axiom of the implicational calculus, I have the impression that I am confronted with a mighty construction, of indescribably complexity and immeasurable rigidity. This construction has the effect upon me of a concrete tangible object, fashioned from the hardest of materials, a hundred times stronger than concrete and steel. I cannot change anything in it; by intense labour I merely find in it ever new details, and attain unshakable and eternal truths."* [Translation by Peter T. Geach, quoted in *A Wittgenstein Workbook*, p. 22, edited by Coope, Geach, Potts, and White.]

Confronted with these elegant syntactical simplifications, let us not forsake the philosophical question posed by Lewis Carroll's parable: how can we justify the logical law of *modus ponens*? Even if logical systems can be reduced to axioms together with the one inference rule *modus ponens*, we can still ask whether this rule is intuitively valid?

A corollary to this is that the law of logic could not be true by convention.

Quine in his famous essay *"Truth by Convention"* (1936) attacked his mentor Carnap for the view that the laws of logic are true by convention:

> In a word, the difficulty is that if logic is to proceed *mediately* from convention, logic is needed for the inferring of logic from the conventions.... It is supposed that if the *if*-idiom, the *not*-idiom, the *every*-idiom, and so on, mean nothing to us initially, and that we adopt… [certain] conventions… by way of circumscribing their meaning; and the difficulty is that the communication of [those conventions] depends upon free use of those very idioms which we are attempting to circumscribe, and can succeed only if we are already conversant with the idioms.

To illustrate Quine's point consider how we would explain how

$(P \rightarrow Q) \wedge P \rightarrow Q$

$(P \rightarrow Q) \wedge P$

How can we apply the schema of *modus ponens*:

If '$(P \rightarrow Q) \wedge P$' is substituted for ϕ and 'Q' is substituted for ψ,

then the following is a valid instance of *modus ponens*,

> $(P \rightarrow Q) \wedge P \rightarrow Q$
>
> $(P \rightarrow Q) \wedge P$
>
> $\therefore Q$

In order to obtain this instance, we need to employ *modus ponens*!

In order words, we need to use *modus ponens* in order to apply the convention of *modus ponens* stated as a schema. Quine concluded that the Carnapian claim that logical laws are simply true by convention would be circular: in order to apply the conventions for a logical rule like *modus ponens* we would already have to be able to use the logical law of *modus ponens*.

Perhaps Carroll's paradox of the infinite regress of justification of rules like *modus ponens* occurs because he lacked the modern distinction, typically attributed to Alfred Tarski, between the formal object language and the informal meta-language in which we justify the syntactical rules by semantic consideration of truth-preserving inferences or semantic validity.

Lesson 2. Syntactical rules can be justified semantically in terms of being truth-preserving.

The syntactical inference of *modus ponens* is justified semantically using truth tables: *modus ponen*s is *truth-preserving*, i.e., it never leads from true premises to a false conclusion.

ϕ

ψ

$\phi \rightarrow \psi$

ϕ

ψ

T

T

T

T

T

T

F

F

T

F

F

T

T

F

T

F

F

T

F

F

In contrast, *modus tortoise*, fails to preserve truth:

ϕ

ψ

$\phi \to \psi$

ϕ

ψ

T

T

T

T

T

T

F

F

T

F

F

T

T

F

T

F

F

T

F

F

Was the Tortoise's "Achilles' heel," so to speak, the failure to respect the Tarskian distinction between the formal object language and the meta-language?

The Tortoise dictates so the Achilles write out *metalogical justification*:

> If *modus ponens* is truth-preserving, then *modus ponens* is a valid rule of inference.

Second, the semantical argument based on truth tables gives us the premise:

Modus ponens is truth-preserving.

So, QED, *modus ponens*, but not *modus tortoise*, is a valid rule of inference.

But wait! The persistent Tortoise applies *modus tortoise* in the meta-language to conclude:

> *modus ponens* is not a valid rule of inference!

The problem is that the meta-logical justification for *modus ponens* is *rule circular*: the justification for *modus ponens* uses *modus ponens* in the metalanguage!

Curiously, the parallel justification for *modus tortoise* goes as follows:

If *modus tortoise* is *not* truth-preserving, then *modus tortoise* is *not* a valid rule of inference.

Modus Tortoise is *not* truth-preserving.

∴ *Modus Tortoise fails not* to be a valid rule of inference!

Can we distinguish between the *virtuous circularity* of the metalogical justification for *modus ponens* and the *vicious circularity* of the parallel justification for *modus tortoise*? This brings us to our revision of the second lesson.

> Lesson #2 *Justifications of logical laws by ascending to meta-language can be rule circular: the justification of those basic logical laws often presuppose we are already justified in using tthose very laws.*

The truly fundamental truths of logic appear to be so basic that they are presupposed in every attempt to justify or to explain them.

Quine himself thought that the above considerations supported a radical thesis with regard to the laws of logic. In his *"Two Dogmas of Empiricism"* (1951), Quine concludes:

> … no statement is immune from revision. Revision even of the law of excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle.

Quine makes the radical claim that even the laws of logic are revisable. This thesis is sometimes called the claim that logic is *not* exceptional.

Quine proposes a story about how these statements might be empirically confirmed. These statements get to be confirmed because of some relation in which they stand to statements of observation and empirical generalization. Kripke raises a puzzle for Quine's view, which can be summarized:

> (Q1) Any statement can be maintained as long as sufficient revisions are made to one's web of belief.

(Q2) Any statement (even statements of logical laws) can be revised.

Just prior to his death, Kripke held a seminar zoomed by logicians and philosophers from around the work about what he dubbed as the “Adoption Problem”. Just as Carnapian truth by convention is subject to the Adoption Problem, so it seems is Quine’s metaphorical views about the revisability of logic in a “web of belief”.

(K1) No scientific statements in one’s web of belief are immune from empirical revision

(K2) The laws of logic are statements in one’s web of belief.

(K3) ⊠ The laws of logic are *not* immune from empirically revision.

Using *modus tortoise* in this philosophical argument leads to the conclusion:

(K3’) ⊠ The laws of logic are immune from empirically revision.

What could Quine mean by his claim that logic is revisable? There are at least two possibilities:

> (QM) The truths of logic *might have been otherwise than they are.*
>
> (QE) The truths of logic *might be other than we take them to be.*

(QM) is the *metaphysical* claim that the laws of logic are not necessary; (QE) is the *epistemological* claim (known as *fallibilism*) that we could be mistaken about what the truths of logic are.

There is good reason to admit that we can be fallible with respect to logic as we have been fallible with respect to other scientific theories. A meta-induction with respect to the fallibility of scientific theories recommends epistemological humility. (Of course, there is perverse meta-induction similar to the Gambler’s Fallacy that justifies taking one more spin on the roulette wheel—“if I’ve lost consistently up to now, it must be more likely that my luck’s about to change!”)

Humility (or, *fallibilism*) with respect to logic is an antidote to (*epistemological*) hubris. Kant audaciously proclaimed: “in our own times there has been no famous logician, and indeed we do not require any new discoveries in logic….” And Kant was, famously, proved wrong by the research into the foundations of mathematics. Kant’s hubris provoked research into the foundations of logic.

This research was profoundly, and perseveringly, initiated by Gottlob Frege [*Begriffsschrift* [1879], *Grundlagen* [1884], *Grundgesetze* [1893/1903]—a task which was formulated by Guisseppi Peano [1889] whose work by preceded by Grassman [1861] and informally justified by Dedekind in his letter to Keferstein [1890]. Frege’s logicist foundational program faltered on Russell’s Paradox, but was carried out in type theory, but less rigorously, by Russell and Whitehead in *Principia Mathematic*a [1910-13]—that gave birth to the modern treatment of the foundations of logic and set theory in which the von Neumann, rather than the Frege-Russell construction, is the more usual.

Frege thought that the reduction of arithmetic to logic and set theory would guarantee arithmetic epistemologically and he took the laws of logic to be analytically self-evident. However, Bertrand Russell discovered that Russell’s paradox could be derived from Frege’s axiom V and the Fregean conception of sets as extension of predicates in contrast to the Cantorian iterative conception of sets.

> Corollary C. The argument that logical laws are revisable typically confuses the *application* of logical laws to natural languages with the *logical laws* themselves which are *prescriptive*, rather than *descriptive*.

Can one be mistaken about logical laws in their application to natural languages? Even Quine appears to be to be fallible with respect to the logic of natural languages. Quine states in his *Methods of Logic* [1962, p. 41] that *’If* ϕ*, then* ψ’ can be paraphrased as ‘ϕ *only if* ψ’:

> But whereas ‘*if*’ is thus ordinarily a sign of the antecedent, the attachment of ‘*only*’ reverses it; ‘*only if*’ is a sign of the consequent.

Although such statements are commonplace in elementary logic books, whether they are correct about how logic applies to natural languages is a matter of empirical investigation,[2] Here are some counterexamples to Quine’s claim that ‘*If* ϕ, *then* ψ’ can be correctly paraphrased as ‘ϕ *only if* ψ’.

(1) *If* Alfred doesn't have the operation, *then* Alfred will die.

(1*) Alfred doesn't have the operation *only if* Alfred will die.

The two conditionals are not logically equivalent: the former states that in the cases in which Alfred doesn't have the operation are cases in which Alfred will die. The latter states that the cases in which Alfred doesn't die are cases in which Alfred has the operation. Intuitively the cases in which Alfred doesn't have the operation are different from the cases in which Alfred has the operation.

Whereas '*If* ϕ, *then* ψ' is often a poor paraphrase of 'ϕ *only if* ψ'; the *contrapositive* of the conditional is often a perfect paraphrase of "ϕ *only if* ψ".

(2.1) I'll leave *only if* you have somebody to take my place.

is perfectly paraphrased by its *contrapositive*:

(2.2) If you do*n't* have somebody to take my place, then I will *not* leave

but poorly by the Quine's paraphrase:

(2.3) If I'll leave, then you have somebody to take my place.

The latter seems to imply that if ever I'll leave, you'll have someone waiting in the wings to take my place; whereas the former has the sense that I am willing to be responsible by not leaving until you've found a suitable replacement.

According to a linguistic analysis of '*if*' by Geis [1973], '*if*' is better paraphrased by "*in cases in which*" and that this paraphrase can account for our linguistic intuitions in the above examples. The cases in which contraposition appears to fail are due to the fact that the relevant set of cases or circumstances governed by '*if*' shifts.

When there is not success shifting of cases implied by "if", such as in mathematics, the counterexamples to contraposition disappear: the following are all equivalent:

(3.1) A set is infinite *only if* it has a subset equinumerous with itself.

(3.2) *If* a set is infinite, *then* it has a subset equinumerous with itself.

(4.3) If a set does not have a subset that is equinumerous with itself, then the set is not infinite.

Experimental linguistic data summarized by Braine [1978] gives empirical support for the claim that

'ϕ *only if* ψ'

should be paraphrased by the contrapositive of Quine's paraphrase:

'*not* ϕ *if not* ψ'

rather than by Quine's paraphrase:

"*if* ϕ, *then* ψ'.

In time response tests, it was discovered that subjects typically perform *modus ponens* inferences more accurately and faster than *modus tollens* inferences. In the first row of the chart below, we indicate this by bolding the inference of *modus ponens.*

If A, (then) B.

A.

Therefore, B

If A, (then) B.

Not B.

Therefore, not A.

A only if B (= not A if not B)

A

Therefore B.

A only if B (= not A if not B)

Not B.

Therefore, not A.

Next timed experiments were conducted on the subject's responses to inferences involving '*only if*'. The results shown in the second row were that subjects perform the inference pattern on the right more accurately and faster than those on the left. This gives experimental support for the view that the correct logical paraphrase of 'ϕ *only if* ψ' is not '*if* ϕ, *then* ψ' as Quine suggested, but rather '*not* ϕ *if not* ψ'. This corroborates other linguistic data to show that '*only'* contributes a negative polarity to the logic of a sentence. The point is that even the writings of Quine on logic give support for fallibility respect to the laws of logic, or at least with respect to linguistic applications of the laws of logic to natural language.

> Corollary D. One can construct, deviant logics but in order to adopt such a logic, one typically deploys classical logic in the metalanguage to first establish the deviant logic is consistent.

We begin by pointing out some of the counterintuitive consequences—known as the *paradoxes of material implication.* The truth-functional interpretation of the conditional (or *if- then* sentences) is called *material implication.* The *paradoxes of material implication* are a collection of theorems based on this truth-functional interpretation that are at odds with the ordinary use of conditionals. In this collection are:

T2 $Q \longrightarrow (P \longrightarrow Q)$ Law of Affirming the Consequent

T18 $\sim P \longrightarrow (P \longrightarrow Q)$ Law of Denying the Antecedent

T19 $(\sim P \longrightarrow P) \longrightarrow P$ *Consequentia mirabilis*, also known as Clavius's Law

T58 $(P \longrightarrow Q) \vee (Q \longrightarrow R)$ Law of Implicational Excluded Middle

Using the language of implication, the Law of Affirming the Consequent says that any conditional is implied by the affirmation of its consequent, whereas the Law of Denying the Antecedent says that any conditional is implied by the denial of its antecedent. *Consequentia mirabilis* (Latin for the "miraculous consequence") says that any proposition is implied by the inconsistency of its negation. The Law of Implicational Excluded Middle says every proposition is either implied by an arbitrary proposition or implies an arbitrary proposition. Or, to put it another way, if you take any three arbitrary propositions, either the first implies the second or the second implies that third.

Lewis's Dilemma (also known as the Explosion Problem, Scotus's Law, and *Ex Falso Quodlibet* (pronounced QUAD-lee-bet) is the observation that any proposition whatsoever can be deduced from a contradiction. This fact is sometimes expressed by saying that in classical logic "a contradiction proves anything." The unintuitive consequence

$P \wedge \sim P \longrightarrow Q$

follows logically from the otherwise intuitively acceptable rules of simplification, addition, and *modus tollendo ponens*:

$\phi \wedge \psi$

$\phi \wedge \psi$

ϕ

$\phi \vee \psi$

⊠ ϕ

⊠ ψ

⊠ $\phi \vee \psi$

$\sim\phi$

⊠ ψ

Simplification

Addition

Modus Tollendo Ponens

The derivation of an arbitrary propositions from a contradiction using only these three inference rules is as follows:

1. *Show* $P \wedge {\sim}P \rightarrow Q$ 6, CD
2. $P \wedge {\sim}P$ Assume (CD)
3. P 2, S
4. ${\sim}P$ 2, S
5. $P \vee Q$ 3, ADD
6. Q 5, 4 MTP

In the end, Lewis's analysis of entailment involved modal logic. He defined implication or strict entailment as follows:

P —3 Q := $\sim(P \wedge {\sim}Q)$ or equivalently P —3 Q := $(P \rightarrow Q)$

Lewis then proceeded to characterize different kinds of strict implication axiomatically. We pause to list two of Lewis's axioms for strict implication:

(S4) $\phi \rightarrow \phi$

(S5) $\phi \rightarrow \phi$

These axiom will play an important role in the development of modal logic.

Exercises:

(A) Prove that another paradox of material implication,

T59.1 $Q \rightarrow (P \vee {\sim}P)$

that says that "a tautology is proved by anything" also follows from the truth-functional analysis of material implication.

(C) Show that a *conjunction* of two implications implies the *disjunction* of the two implications obtained by *interchanging* their consequents:

T50.1 $(P \rightarrow Q) \wedge (R \rightarrow S)$ ⊠ $(P \rightarrow S) \vee (R \rightarrow Q)$

Do this by constructing a derivation and by using a semantic truth-value analysis.

(D) Show that the intuitively acceptable *Law of Exportation*

T27 $(P \wedge Q \rightarrow R) \leftrightarrow (P \rightarrow [Q \rightarrow R])$

implies that a *disjunction* of conditionals with a consequent in common is equivalent to the *conjunction* of the antecedents of those conditionals implying that consequent:

T60 $(P \rightarrow R) \vee (Q \rightarrow R) \leftrightarrow (P \wedge Q \rightarrow R)$.

(E) Consider again what we have called the Law of Implicational Excluded Middle:

T58 $(P \rightarrow Q) \vee (Q \rightarrow R)$

Analyze this theorem semantically by making the disjunction false and assuming the truth-functional analysis of the conditional. Why is it that only the truth assignment to Q matters, and not the truth value assignments to P or R?

The logicians Alan Ross Anderson, Nuel Belnap, and J. Michael Dunn developed the branch of logic known as *Relevance Logic.*[3] The idea of relevance logics are to construct logical implication in such a way that the consequent must be "relevant" to the antecedent, e.g., by sharing language in common. Such an implication would fail to have the property of classical logic that "a contradiction implies anything."

One such relevance logic is known as *DeMorgan Implication.* In DeMorgan implication there are familiar principles—identity, commutativity and associativity of ∧ and ∨, distributivity of ∧ over ∨ and ∨ over ∧, absorption, and De Morgan's laws—but Lewis's Dilemma is blocked. So we know one of the inferences of simplification, addition or *modus tollendo ponens* must fail if we are to block the derivation of Lewis' dilemma.

DeMorgan logic has four truth-values, which can be represented by the power set of the two values {1, 0} —the sets {1} (true only), {0} (false only}, {1, 0} (both true and false), the empty set { } (neither true nor false), which we abbreviate **T**, **F**, **B,** **U**, respectively.

The truth tables or semantics for DeMorgan Implication are summarized in a Hasse Diagram or lattice in the standard way. The vertical line in the center of the lattice summarizes the classical values with *conjunction* (the Min or lower of the two values) and *disjunction* (the Max or higher of the two values). The values are extended to the four values in a similar way where the conjunction or Min of the values **T** is **U**, of **T** and **B** is **B**, of **U** and **B** is **F.** The disjunction of these is **T.**

Implication

Given that conjunct is read down and disjunction is read up, it is clear that th

The semantics or truth tables for De Morgan Implication can be summarized in a Hasse Diagram with a four element model: {{0}, {1}, { }, {0,1}}, which can be thought of as the powerset of the classical truth values. For brevity, we shall use the values **F**, **T**, **U** and **B,** for {0}, {1}, { }, and {0, 1}, respectively.

T = {1}

U = { } B = {0,1}

F = {0}

In this diagram, we define ∧ and ∨ by Min and Max in the Hasse Diagram:

$|\phi \wedge \psi| = \text{Min}\ \{|\phi|, |\psi|\}$

$|\phi \vee \psi| = \text{Max}\ \{|\phi|, |\psi|\}$

The conditional can be defined in terms of the disjunction:

$(\phi \longrightarrow \psi)$ =df $(\sim \phi \vee \psi)$

Then the relationship of *De Morgan implication* is defined as follows:

ϕ *implies* ψ (in symbols ϕ –3 ψ) =df for all valuations $|\phi| < |\psi|$.

You can easily verify that the following implications are valid. They follow immediately from the fact that the value of disjunctions between two nodes of the diagram can be determined by going *up* to the least upper bound. The values of conjunctions, on the other had, can be determined by going *down* to the greatest lower bound of two nodes. These facts suffice to justify the following implications:

$\phi \wedge \psi$ –3 ϕ

$\phi \wedge \psi$ –3 ψ

ϕ –3 $(\phi \vee \psi)$

ϕ –3 $(\psi \vee \phi)$

But the following implication—*modus tollendo ponens*—is invalid as can be seen from the following:

(ϕ ∨ ψ) ∧ ~ϕ −3 ψ

B B F B BB > F

In other words, *modus tollendo ponens* is invalid.

Notice, *modus tollendo ponens*, via, conditional disjunction, is intuitively equivalent to *modus ponens*:

ϕ ⟶ ψ

~ϕ ⟶ ψ

ϕ ∨ ψ

ϕ

⊠ ϕ ∨ ψ

~ϕ

⊠ ψ

⊠ ψ

Modus Ponens

Conditional/Disjunction

Modus Tollendo Ponens

So we are in a peculiar position logically speaking: we can construct deviant logics in which *modus ponens* fails and even proved that deviant logic is consistent—but that proof of consistency uses *modus ponens* in the metalanguage!

Lesson 3. If it is possible to construct deviant logics in which *modus ponens* fails,

> then *modus ponens* is not known to be true *a priori* UNLESS (iff not) *modus ponens* is not necessarily presupposed in the construction of those deviant logics.

It is possible to construct such deviant logics, but *modus ponens* is necessarily presupposed.

Hence, *modus ponens* is known to be true *a priori*.

> Corollary E. Accepting a *logical law* expressed as a proposition is not the same as perceiving the *rational* basis for engaging in an inferential *practice*.

Hao Wang (*Reflections* [1988], p. 203) mentions Gödel's notion of mathematical or logical intuition in connection with Lewis Carrol's parable: "G contrasts intuition with proof. A proof can be explicit and conclusive because it has the support of axioms and rules. In contrast, intuitions can be communicated only by pointing things out. An elegant illustration of this distinction is Lewis Carroll's frequently quoted 'What the Tortoise said to Achilles [*Mind*, vol. 4, 1895, pp. 278-280l …."

In the third of six drafts of "Is mathematics syntax of language?" Gödel states his conclusion (in italics, GCW-III [*1953/9] (hereafter, Syntax-III), pp. 346-7):

> *To eliminate mathematical intuition or empirical induction by positing the mathematical axioms to be true by convention is not possible.*

Gödel gives three intriguing reasons for this conclusion:

1. *The syntactic maneuver destroys any reason for expecting consistency*: "The scheme of the *syntactic* program to replace mathematical intuition by rules for the use of symbols fails because this replacing *destroys any reason for expecting consistency*. …. And because for the consistency proof one either needs a mathematical intuition or a knowledge of empirical containing equivalent empirical content" (p. 348)

2. *Having a rational basis for a practice requires perceiving truth*: "For these axioms there exists no other *rational* (and *not merely practical*) foundation except either that they or propositions implying them) can directly be perceived to be true (owing to the meaning of the terms or by the intuition of the objects falling under them), or that they are assumed (like physical hypotheses) on the grounds of inductive argument, e.g., their success in the applications."

   Consider simplification as a mere practice of detaching a disjunct as opposed to understanding the rational basis for the rule. Here is an example of the practice that goes wrong:

   This statement is part of a conjunction, and snow is white. (True)

   ⊠ This statement is part of a conjunction. (False)

Understanding why this inference fails to be an instance of simplification requires more than engaging in the syntactic practice of detaching a conjunct—one needs to understand the rational basis for that rule of inference, how demonstratives work, and the meaning of propositions.

1. *Mathematical Induction and Modus Ponens are on a par epistemologically*: The former case would seem to apply at least to some mathematical axioms, e.g., the *modus ponens* and complete induction."

footnote 34: :"It seems arbitrary to me to consider the proposition 'This is red' an immediate datum, but not so to consider the proposition stating modus ponens…."

> "Complete induction would seem to be an axiom (or a consequence of axioms) of the same kind.. On the basis of the non-constructive standpoint this has been proved by Dedekind and Frege."

> There is a different between having the intuition is required for one's inferences to be *governed* by MP, UI, mathematical induction versus merely *conforming* or applying the schematic rules of MP, UI, or mathematical induction. If one understands the *rational* basis for mathematical induction, for example, one can, not only *deduce* by the law of contraposition, that Fermat's proof by infinite descent is logically equivalent and then come to see *why* the two forms are equivalent.

Idea (Poincare): Mathematical Induction can be viewed as an infinitary *modus ponens*.

To return to our philosophical tale, we note that Saul Kripke has marshaled the Lewis Carroll paradox to argue against Quine's thesis that the laws of logic (not merely our understanding of the laws of logic) are revisable.

> Corollary E Even in adopting a logic that deviates from classical logic (e.g., a *relevance* logic or an *intuitionistic* logic that replaces MT* with MT and DN+ but doesn't have DN-) presupposes that one uses such basic laws a universal instantiation and *modus ponens* in the metalanguage.

How could we have learned that the logical law of *modus ponens* is valid? If we already in some sense must know that *modus ponens* is valid in order to justify it, then we cannot really say that we have justified *modus ponens*, for then we would be presupposing what we were trying to justify. If, on the other hand, we really were clueless about the validity of *modus ponens*, then we are the position of Achilles trying to convince the Tortoise that *modus ponens* is valid. Unless the laws of logic were known *a priori*, it seems there would be no way to come to know them to be valid. Kripke's position embraces the Socratic paradox of learning in the *Meno* transposed into the context of learning the laws of logic.

Arguing in a humorous, and perverse, Tortoise-like manner, what are we to make of the Quinean revisionist argument?

> (K1) Either we are justified in adopting the laws of logic as *exceptional* or the laws of logic are empirically revisable.
>
> (K2) We are *not* justified in adopting the laws of logic as exceptional.
>
> (K3) ⊠ The law of logic are empirically revisable.

If you have caught on to the Carrollian spirit of *Modus Tortoise*, you've probably noticed that this argument uses *Modus Tollendo Ponen*s, which in its *Modus Tortoise-Tertullian* version, allows us to infer perversely:

> (K3') ⊠ The law of logic are *not* empirically revisable.

If things are getting a little too abstract, let's return the quotation from Robert Frost.

> "Abstraction is an old story with the philosophers, but it has been like a new toy in the hands of the artists of our day…. Our problem then is, as modern abstractionists, to have the wildness pure; to be wild with nothing to be wild about.
>
> "I tell how there may be a better wildness of logic than of inconsequence. But the logic is backward, in retrospect, after the act. It must be more felt than seen ahead like prophecy…."
>
> *"The figure a poem makes. It begins in delight and ends in wisdom….For me the initial delight is in the surprise of remembering something I didn't know I knew…. There is a glad recognition of the long lost and the rest follows. Step by step the wonder of the unexpected supply keeps growing."*

— Robert Frost (1874-1963), "The Figure of Poem Makes" [1939]

[1] I attended one of these lectures as a graduate student at UCLA and they are mentioned by Hilary Putnam in his "Analyticity and Apriority: Beyond Wittgenstein and Quine" [1979].

[2] Gary Mar with Amanda Caffary and Yuliya Manyakina, "*Unless* and *Until*: A Compositional Analysis," CITATION .Abstract, Tenth International Tbilisi Symposium on Language, Logic and Computation and the Georgian Academy of Sciences and Institute for Logic, Language and Computation (ILLC) of the University of Amsterdam, Oct. 2013

[3] *Entailment*, Vol. 1: *The Logic of Relevance and Necessity* [DATE], Vol. II [DATE], Alan Ross Anderson, Nuel D. Belnap Jr. & J. Michael Dunn,

3 PHI 631. Analytic Seminar

3.1 Gödel's Ontological Dreams and Explorations in Logic

Die Welt ist vernünftig.
— Gödel's notebooks, "My Philosophical Outlook"[1]

"My belief is theistic; not pantheistic, following Leibniz rather than Spinoza."
"Spinoza's God is less than a person. Mine is more than a person.... "
— Gödel's unsent response to the Grandjean questionnaire[2]

Kurt Friedrich Gödel (April 28, 1906 – January 14, 1978) was the greatest logician of the 20th century. Indeed, John von Neumann lauded him as the greatest "logician since Aristotle" and the only mathematician who was "absolutely irreplaceable"[3,4]. Harvard University bestowed upon him an honorary doctorate "for the discovery of the most significant mathematical truth of the century."[5] His friend Einstein said that he went to the Institute of Advanced Studies "*um das Privileg zu haben, mit Gödel zu Fuss nach Hause gehen zu dürfen.*"[6] Despite this nearly universal admiration of Gödel,[7] an accurate understanding of the scope and depth of Gödel's achievements has been marred by misleading and exaggerated accounts.[8] Gödel's theorems, and philosophical achievements, however, stand in no need of exaggeration.

Gödel dreamed of establishing significant philosophical theses with the rigour and precision of mathematics. Other great philosopher-mathematicians—Leibniz, Frege, and Cantor, for example—have also had grand dreams and Gödel's theorems have had implications for those dreams.[9]

Despite their technical sophistication, Gödel's theorems have perennially managed to escape mere mathematics and shed light on larger philosophical issues. While an undergraduate at the University of Vienna, Gödel determined to devote himself only to the sort of mathematics that would have broader philosophical implications. The reason why Gödel's theorems achieved such significance is not merely his judicious choice of problems but his virtuosity with a new method of mathematizing philosophical problems. Hilbert's challenge to solve open problems in the foundation of mathematics inaugurated the revolutionary shift from the Frege-Russell search for universal logics to metamathematics. Instead of remaining mired in unresolvable paradoxes, the Gödel program[10] exploits the interplay between a formal and an informal intuition to obtain limitative results with mathematical precision.

Gödel's achievement of producing logical elegant results by metamathematical methods that have implications for core philosophical problems—e.g., the reality of a Platonistic conceptual world, the non-algorithmic nature of the mind, the refutation of verificationism in mathematics, the structural similarities with completeness proofs and the ontological arguments, and the implications of relativity theory for the nature of time—ought to be more accurately and widely known, and deployed, by contemporary philosophers. This book addresses this need by providing an accessible, yet reliable, guide to Gödel's logical explorations, philosophical ideas, and formal methodology.[11]

These are exciting times for Gödelian scholarship. Ever since the historic centenary celebration of Kurt Gödel's birth, the *Horizons of Truth* conference in Vienna in 2006, which brought together philosophers, logicians, computer scientists and other scholars, the number of logicians, philosophers, and historians publishing topics related to Gödel's life and research has grown dramatically.

Moreover, the topics in this explosion of research are no longer limited to Gödel's recognized work in logic, set theory, the foundations of mathematics, but have expanded to encompass his explorations of the relativity physics of time travel, his views on computability and the philosophy of mind, and even his original contribution of a modal ontological argument to philosophical theology.

[1] This is the first in a list of 14 propositions under the heading "My Philosophical Outlook" in Gödel's notebooks. The last proposition is "Religions are, for the most part, bad—but religion is not." Three of the remaining propositions are: (6) "There is incomparably more knowable *a priori* than is currently known"; (10) "Materialism is false"; and (13) "There is a scientific (exact) philosophy and theology, which deals with concepts of the highest abstractness; and this is also most highly fruitful for science".

[2] Quoted by Hao Wang, *Reflections on Kurt Gödel* (1987), 18 and *A Logical Journey: From Gödel to Philosophy* (1996), 152.

[3] These accolades are referenced in Karl Sigmund, John Dawson, and Kurt Mühlberger's *Kurt Gödel: Das Album* (Vieweg & Sohn Verlag, April 2006), 10, which was published in conjunction with the Gödel exhibition at the 2006 Centenary sponsored by the Templeton Foundation at the University of Vienna.

[4] A tribute on the occasion of Gödel's receiving an honorary doctorate from Harvard was authored by W. V. O. Quine. Gödel talked about this tribute with great admiration with his mother.

[5] That is, *"in order to have the privilege of walking home with Gödel"*, Oskar Morgenstern's letter to Bruno Kreisky (*Bundesmeister fur Auswartige Angelegenheiten* of Austria) dated 23 October 1963.

[6] Two notable exceptions are Ludwig Wittgenstein and Ernest Zermelo, Gödel's most famous philosophical and mathematical detractors. Although Wittgenstein and Gödel never met personally, each had views about the significance of the other's work. Wittgenstein, for example, tried (unsuccessfully) to persuade Alan Turing of the insignificance of the paradoxes and metamathematics. Turing had already published his most celebrated result, the proof of the Unsolvability of the Halting Problem, which contains a philosophical analysis of computability, which Gödel praised highly. In the republication of his 1931 incompleteness results, Gödel added the following footnote in 1963: "In consequence of later advances, in particular of the fact that due to A. M. Turing's work (Turing [1937] 'On computable numbers, with an application to the *Entscheidungsproblem*,' *Proceedings of the London Mathematical Society*, 2nd series, 42, 230-65) a precise and unquestionably adequate definition of the general notion of a formal system can be given, a complete general version of Theorems VI and XI is now possible. That is, it can be proved rigorously that in every consistent formal system that contains a certain amount of finitary number theory there exist undecidable arithmetic propositions and that, moreover, the consistency of any such system can be proved in the system." Gödel met with Zermelo (DATES) and to tried to explain the fundamental ideas of his Incompleteness Theorem; however, Zermelo's antipathy for Gödel's theorems, never turned from *dustaub* into an intellectual acceptance (in contrast with Hilbert).

[7] John Cornwall's *Darwin's Angel: an Angelic Riposte to* The God Delusion (Profile, 2007) contains the following common, but inaccurate, gloss on Gödel (64-65):
"The story begins in 1900 at the mathematical congress in Paris, where Hilbert set for the mathematicians of the world a list of problems for completion in the new century. Not least, he challenged them to demonstrate that mathematics is self-proving. He asked for a computational method, or algorithm, for resolving any kind of mathematical problem.... But in 1931 Kurt Gödel wrote a proof that upset Hilbert's proposal. He demonstrated that there are mathematical statements that no conceivable computer, however capacious, could settle.
"Crucially, philosophers of science have shown that what goes for mathematics goes for physics too. Among early reactions to Hawking's announcement of his quest for the Theory of Everything, a number of peer academics had attempted to expose the Gödel flaw in Hawking's proposal. Gödel had shown, they insisted, that in principle, and for all time a Theory of Everything was bound to be either incomplete or inconsistent. These included world-class mathematicians and theoretical physicists such as Roger Penrose of Oxford (author of *The Emperor's New Mind*), Paul Davies... (author of *The Mind of God*), and ... John Barrow (author of *Impossibility*). 'What Gödel shows,' said Barrow, who now holds the chair in public understanding of mathematics in Cambridge, 'is that no final theory of everything is possible; and that in any case there could be no algorithm, or mechanical procedure, that enables you to prove such a theory."
This account is tightly scripted, breath-taking in its sweep, but there is a problem. Practically every statement in this catechism is false or, at best, an exaggeration (see appendix).

This explosion of research was facilitated by the scholarship published in *Gödel's Collected Works* edited by Solomon Feferman, *et al.*, volumes I - V [1982 - 2003]. This monumental resource made new avenues of research possible by the carefully edited corpus of Gödel's published and unpublished papers, lectures, and selected portions Gödel's *Nachlass* written in *Gabelsberger* shorthand script, as well as much of Gödel's scientific correspondence, sent and unsent—all with critical commentary by leading scholars.

It is perhaps not surprising that *Gödel's Collected Works* is incomplete, e.g., Gödel's correspondence with Georg Kreisel has been withheld from his correspondence as well as hundreds of pages of Gödel's notebooks in *Gabelsberger* shorthand. At a recent conference *Gödel's Legacy: Does the Future Lie in the Past?* [July 2019] dealing with Gödel's writings on time and relativity, Jan von Plato gave a presentation about his project of translating from Gabelsberger short-hand Gödel's four notebooks with the title "*Resultate Grundlagen*" consisting of 368 consecutively numbered pages and 90 theorems proved using a system of natural deduction invented by Gödel for this purpose.

The explorations in the chapters that follow represent some of my own attempts to contribute to the creative chaos in contemporary thinking about Gödel's life, logical writings, and continuing legacy.

This book is written with several audiences in mind.

First, it is mean for students who loved their first course in symbolic logic and who want to grasp Gödel's results in more than a metaphorical way. It is for students who have experienced the joy of constructing derivations and would like to understand Gödel's theorems in ways that build upon what they already know. It is a guide that is simple and elegant, but not too simple with enough technical details (but not more than is necessary) to construct and comprehend those theorems.

Secondly, the book is meant for a general audience, who may appreciate the importance and power of logic but who are not satisfied with merely following logical proofs because they wish to explore the *philosophical* these ideas. It is also for those who want to know the *historical* narrative of Gottlob Frege's *Begriffschrifft* [1879] conceptual writing or symbolic quantifier logic, Georg Cantor's Transfinite Set Theory [1874 – 1897], the logical empiricism of the Vienna Circle (1924 – 1936), Bertrand Russell and Alfred North Whitehead's *Principia Mathematica* [1910-1913], David Hilbert's 23 open problems for 20th century mathematics, Gödel's Completeness and Compactness Theorems [1930], Gödel's Incompleteness Theorems [1930], Alonzo Church's solution to Hilbert's *Entscheidungsproblem* [1936], Alan Turing's alternative solution, analysis of computability, and proof of the Unsolvability of the Halting Problem [1936], the friendship between Einstein and Gödel at the Institute for Advanced Studies in Princeton.

Thirdly, the book is also meant to continue the dialogue among Gödel amateurs and experts, who may never have thought of their areas of expertise in new ways or discovered connections between disciplines. The purpose of this book is not merely to *celebrate* Gödel's achievements but to look at them in their historical context and as a coherent whole—suggesting certain ways of *comprehend* them ... thinking and certain logical techniques that the used to tackle fundamental philosophical issues in the philosophy of mathematics.

The book is dedicated to my teachers, who taught me to love logic, and to my students who took all my classes and who still wanted to know more after the semesters ended. Among my teachers, first and foremost is Donald Kalish, my mentor at UCLA, but also Herbert Enderton, C. C. Chang, Tyler Burge, Donald ("Tony") Martin, David Kaplan, and my Ph.D. chair Alonzo Church.[12]

I have also benefited from conversations and correspondence with friends and colleagues Matthias Baaz, John Dawson, Solomon Feferman, Juliet Floyd, Melvin Fitting, Thomas Graf, Patrick Grim, Jeffrey Heinz, Douglas Hofstader, Hans Kamp, Juliette Kennedy, Richard Larson, Robert L. Martin, Roger Penrose, Jan von Plato, Paul St. Denis, Nathan Salmon, Dana Scott, Brian Skyrms, Raymond Smullyan, Peter Woodruff, Palle Yourgrau, Yutong Zhang, and many others too numerous to mention.

My wish is that my teachers, colleagues, and students will continue to pursue their learning of logic and will pass on their love of logic and passion for philosophy to generations yet to come.

3.1.1 Chapter 1. Metalogic, Maximal Consistency, and Completeness

This chapter uses Lewis Carroll's "What the Tortoise Taught Achilles" 1895] to motivate the central questions of metalogic, the logic of logic. The natural deduction system of Kalish and Montague [1964, 1982]

[8] Leibniz (1646-1716) dreamed of a single logical calculus that could settle philosophical disputes; Frege (1848–1925) dreamed of discovering a logical calculus or *Begriffsschrift* (1879) that could serve as the foundation (*Grundlagen der Arithmetik*, 1884) from which the laws of arithmetic (*Grundgesetze der Arithmetik,* vol. I, 1893; vol. II, 1903) could be derived as theorems; and Cantor (1845-1918) dreamed of discovering the mathematics of infinity (*Contributions to the Founding of the Theory of Transfinite Numbers,* 1915). The first open problem listed by Hilbert in his famous 1900 lecture concerned Cantor's Continuum Hypothesis, and Hilbert's hopeful prediction (1926) was that "No one will drive us from the paradise which Cantor created for us." Gödel's theorems—the completeness of first-order logic, the incompleteness of axiomatic systems containing number theory, and the consistency of the Axiom of Choice and the Continuum Hypothesis with the other axioms of set theory—are not only the most strikingly original and profound foundational results in 20th century logic but these theorems had implications for each of these dreams Gödel's Completeness Theorem proved that a formulation of the first-order part of Frege's predicate logic was sound and sufficient to prove all the logical truths expressible in that formal language. Gödel's celebrated Incompleteness Theorems showed that Leibniz's dream was not only *unrealized* but, in principle, *unrealizable* within any single axiomatic system. And Gödel's work in set theory concerning the consistency of the Axiom of Choice (1935) and Cantor's Generalized Continuum Hypothesis (1939) proved these two axioms were relatively consistent with the remaining axioms of Zermelo-Fraenkel set theory, which characterizes, at least in part, Cantorian set theory.

[9] This felicitous phrase is due to Palle Yourgrau [2019], see also Wang [1996].

[10] One reason why Gödel's work is so widely misunderstood is that it has been accessed indirectly through highly simplified and misleading accounts. Nagel and Newman's classic *Gödel's Proof* [1958, updated 2001] included misstatements of Gödel's first incompleteness theorem and Rosser's improvement. Douglas Hofstader's [1979] Pulitzer Prize-winning *Gödel, Escher Bach*, while popularizing Gödel's ideas in a playful manner reminiscent of Lewis Carroll, isn't a reliable guide because of Hofstader's incessant pro-AI agenda woven into his exposition. Similar remarks, in the opposing view, might apply to Roger Penrose's *anti*-AI agenda in *The Emperors New Mind* [1989] and *Shadows of the Mind* [1994]. Rebecca Goldstein [2005] deploys her considerable skills as a novelist and philosopher of science in *Incompleteness: The Proof and Paradox of Kurt Gödel*, but her accounts of foundational issues in mathematics are impressionistic and her overly dramatic insistence on Gödel's unquestioned, and fanciful, platonism distorts what is known Gödel's evolving views through his unpublished writings and conversations with Hao Wang [1996].
More significantly, there is now a scholarly resource available in the monumental *Gödel's Collected Works*, ed. Solomon Feferman, *et al.*, vols. I-V [1982-2003], which makes possible new avenues of research through an examination of the entire corpus of Gödel's published and unpublished papers and lectures, Gödel's *Nachlass* written in *Gabelsberger* shorthand script, and Gödel's scientific correspondence sent and unsent—all with critical commentary by leading scholars.

[11] Donald Kalish (1919 – 2000) was my teacher and mentor at UCLA with whom I co-authored a second edition of the classic Kalish and Montague textbook. Kalish approached Alonzo Church to be my dissertation advisor at a critical time when I had no advisor. Donald ("Tony") Martin read the first draft of my dissertation before it was submitted to Professor Church. The dissertation was subsequently submitted to Church at a time when he was recovering from a hall in a convalescent home in Santa Monica. I was the last student to have a dissertation directed by the great 20th century logician Alonzo Church, whose list of Ph.D. students reads like the "Who's Who" of logic—among Church's first twenty dissertation students are Stephen Kleene, J. Barkley Rosser, Alan Turing, Leon Henkin, John Kemeny, Martin Davis, Nicholas Rescher, William Boone, Hartley Rogers, Jr., Michael Rabin, Dana Scott, Simon Kochen, and Raymond Smullyan.

[12] I have used the Kalish-Montague-Mar natural deduction system to give advanced logic students a thread of Ariadne through Smullyan's labyrinthine puzzles about logicians who reason about themselves in *Gödel's Incompleteness Theorems* [OUP, 1992] and *Forever Undecided: A Puzzle Guide to Gödel* [1987].

is reduced to Stanisław Leśniewski's (1886–1939) Axiomatization for Conditional Logic. The symmetry between semantic and syntactic ideas motivates result s in proofs theory (e.g., the deduction theorem, proof of the independence of the axioms, the embedding of intuitionistic logic with classical logic) and model theory (e.g., soundness, completeness, compactness). It discusses why Gödel's thought that logicians, blinded by the verificationism of the Vienna Circle, failed to formulate the question of completeness for first-order logic until nearly 50 years after Frege's discovery. This chapter also explores unexpected consequences of Gödel‘s Compactness Theorem (e.g., Abraham Robinson's calculus with infinitesimals), Gödel's modal investigations into intuitionistic provability, and Gödel's maximal consistency proof to provide proof of "Leibniz's Lacuna", the missing premise in Descartes' version of the ontological argument.

3.1.2 Chapter 2. Gödel's Incompleteness Theorems

This chapter presents Gödel's Incompleteness Theorems [1930] using modal provability logic discovered by Löb, Kripke, de Jongh, and Sambin in the 1970s.[13] This is done by developing the standard systems of modal logic **D, T, B, S4, S5** for the Kalish-and-Montague system of natural deduction. Gödel's Incompleteness Theorems, especially, the Second Incompleteness Theorem, according to the usual story, destroyed Hilbert's program, but Gödel's writings show that he was quite sympathetic to the continuation of Hilbert's program by other means. Gödel's theorem shows the inadequacy of *single* formal system leaving open the possibility of "Ordinal logics" investigated by Turing [1939] in his thesis completed under Alonzo Church.[14] Why doesn't Gentzen's [1936] Consistency Proof for Arithmetic contradict Gödel's proof of the Unprovability of Consistency? Gödel's last published paper his *Dialectica* interpretation [1958] extends Hilbert's finitism with principles for functionals of finite type which can be used to prove the consistency of intuitionistic number theory.

[13] This research program was renamed and revived in Feferman [1962] "Transfinite recursive progressions of axiomatic theories." An accessible introduction is Franzén's *Inexhaustibility* [2004].

[14] Philosophers have tried to sandbag Gödel's ontological argument by adding additional premises to render it susceptible to standard atheistic objections, e.g., Howard Sobel [1987] claims that Gödel's ontological argument implies a "modal collapse" by adding to Gödel's argument the (implausible) premises that God must know all truths and that any truth that God knows must be necessary. For another instance, see "Truth, Omniscience, and Cantorian Arguments: An Exchange," Alvin Plantinga and Patrick Grim [1983] and for a refutation of Grim, see Mar [1993], "Why Cantorian Arguments Against the Existence of God Don't Work," *IPQ*, 429-442.

3.1.3 Chapter 3. Gödel vs. Turing: Mechanistic Algorithms or the Creative Mind?

This chapter explores the question whether Gödel's platonist views in philosophy of mathematics can be reconciled with Turing's mechanistic and computational views in the philosophy of mind. The self-appointed defenders of Gödel and Turing have tended to defend dichotomous views, e.g., materialist often give circular arguments for a mechanistic philosophy of mind, whereas J. R. Lucas' "Mind, Machines and Gödel" [1961], John Searle's [1980] provocative Chinese Room thought experiment, and Roger Penrose's use of Gödel's Second Incompleteness Theorem to argue against strong AI and for a quantum theory of consciousness presuppose, but provide a solution, to "hard problem" of consciousness. Gödel's dichotomy is, in contrast, more nuanced: "*Either the human mind surpasses all machines (to be more precise: it can decide more number-theoretic questions than any machine) or else there exist number theoretic questions undecidable for the human mind.*" [1951, Gibbs Lecture, Brown University].

3.1.4 Chapter 4. Cantorian Set Theory, and the Gödel Program for Large Cardinal Axioms

This chapter examines Gödel's profound contributions to axiomatic set theory. Gödel was not always the fundamentally convinced mathematical platonist of his three masterful philosophical articles "Russell's Mathematical logic" [1944], "What is Cantor's continuum problem?" [1947, substantially updated 1964], and his unpublished critiques of Rudolf Carnap's (and the Vienna Circle's) views on mathematics in "Is mathematics the syntax of language?" [*1953/9-III]. In his 1933 Cambridge lecture Gödel expresses caution: '*The result... is that our axioms, if interpreted as meaningful statements, necessarily presuppose a kind of Platonism, which cannot satisfy any critical mind and which does not even produce the conviction that they are consistent.*" After 1959, Gödel's turn the philosophy of Husserl indicates that he was searching for an epistemologically, or phenomenologically, grounded platonism.

3.1.5 Chapter 5. Gödel Universes and the Disappearance of Time

This chapter discusses Gödel's discovery of a solution to Einstein's field equations in the General Theory of Relativity of a cylindrical cosmological universe, which, like the self-referential Gödel sentence, rotates back on itself, thus allowing for time travel into the past by traveling into the future. The convergence of the physical with the philosophical was brought about by the convergence of two avenues of research—George

Gamow's question about the physics of cylindrical universes and Gödel's antecedent philosophical interest in a Kantian theory of time. In contrast to his argument in the Incompleteness Theorems in which Gödel affirms the intuitive notion of truth to establish the incompleteness of provability within formal arithmetic, Gödel uses the physics of cosmological universes with close time-like circles to call into question the reality of our intuitive ("A series") conception of time.

3.1.6 Chapter 6. Gödel's Citizenship Test and a "Strange Loop" in the U.S. Constitution

Naively, we assume a democratic society automatically results from "majority rule" based on the principle of "one person one vote." These assumptions about voting are demonstrably false. Recent elections have produced widespread disillusionment with democratic processes, which appear to be "rigged." Voter alienation can be blamed on many different factors, e.g., the two-party system, the electoral college, echo chambers of slanted reporting, disinformation through foreign interference, gerrymandering, voter suppression, etc. However, there are also paradoxes hidden within the logic of different methods of vote counting. This chapter discusses voting paradoxes, Arrow's Theorem about the Impossibility of Democracy, and Gödel's discovery of an "inconsistency" in the U. S. Constitution that would allow for a democracy to end up with a dictatorship.

3.1.7 Chapter 7. Gödel's GOD: Modal Monstrosity or Logical Investigation of a Leibnizian Lacuna?

On February 10th, 1970, when he feared his own death, Gödel shared his notes for an ontological proof for the existence of God with Dana Scott—two pages of symbolic formulas with terse, sometimes cryptic, comments. Gödel did not try to prove, as Leibniz did, that perfections are mutually compatible.[15] Instead he instead postulates that positive properties are closed under conjunction and that the property of positiveness is monotonic. This advances the literature, not by deriving God's actual existence from God's possible existence,[^16] but by axiomatizing relevant notions and constructing a modal maximality argument. Since Gödel never published his argument, there has arisen a philosophical and logical tradition of proposing emendations to Gödel's argument. Scott himself circulated a modified version of Gödel's argument with brief notes on his conversation with Gödel (Scott, 1987). Howard Sobel (1987) published his transcription from Gödel's original handwritten notes together with his charge that it was implicated in a "modal collapse." Responding to Sobel's charge, Anderson [1990] set forth an elegant set of emendations. More recently, Christoph Benzmüller with Bruno Woltzenlogel Paleo and others [2014] have published computational studies of Gödel's modal ontological argument.

[15] Robert M. Adams [1971] "*The Logical Structure of Anselm's Argument*", *Philosophical Review* 80:647-84 pointed out that G follows from the premise $\Box(G \rightarrow \Box G)$, the modal axiom $\mathbb{B}$ $\Box G \rightarrow \Box\Box G$, and the possibility premise $\Box G$.

3.2 Appendix

Gödel's Theorem: An Incomplete Guide to Its Use and Abuse by Torkel Franzén
A K. Peters, Wellesley, Massachusetts, 2005), 172 pages, ISBN 1-56881-238-8, $24.95.

Reviewed by Gary Mar, *The Mathematical Intelligencer* (vol. 29, no. 2), 2007, pp. 66-70.

At the Gödel Centenary Conference, "Horizons of Truth", held at the University of Vienna in April 2006, Solomon Feferman paid tribute to the work of the late Torkel Franzén. Feferman's comments, printed on the cover of *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*, succinctly pinpoints Franzén's distinctive achievement. "This unique exposition of Kurt Gödel's stunning incompleteness theorems for a general audience manages to do what none other has accomplished; explain clearly and thoroughly just what the theorems really say and imply and correct their diverse misapplications to philosophy, psychology, physics, theology, post-modernist criticism and what have you."

Franzén's book will be of interest to three audiences: (1) beginning logic students who want a concise and self-contained explanation of what Gödel's theorems *do* say; (2) non-mathematically trained scholars and educated laypersons who want a logically correct explanation of Gödel's theorems do *not* say; and (3) professional logicians who want a comprehensive, and critical, survey of the philosophical perspectives opened up by Gödel's work.

Logic students now have access to many popular accounts of Gödel's life and work—Nagel and Newman's classic exposition *Gödel's Proof* (1959) and Douglas Hofstader's Pulitzer-Prize winning *Gödel, Escher Bach* (1979), and more recently, John Casti and Werner DePauli's *Gödel: A Life of Logic* (2000) based on the Austrian national television documentary and Rebecca Goldstein's novelistic biography *Incompleteness: The Proof and Paradox of Kurt Gödel* (2005). However, these books tend to sacrifice technical correctness for public comprehensibility. None of these books comment in detail on the many misstatements and misapplications of Gödel's theorem, and some commit the very errors Franzén exposes. Steering the beginning student clear of some common confusions, Franzén explains technical terms and poses instructive questions:

- Gödel published the *completeness* theorem (1930) for his doctoral dissertation and then in the following year published his celebrated *incompleteness* theorem (1931). The latter is not the negation of the former. What are the two quite distinct meanings of *completeness* in these two landmark theorems by Gödel—the former concerning first-order logic and the latter concerning Peano arithmetic?

- Although it is common to speak of *the* incompleteness theorem, there are actually *two* incompleteness theorems, known as Gödel's First and Second Incompleteness Theorems. Contemporary formulations of both theorems talk about formal systems that "contain a certain amount of arithmetic." What two different requirements are meant by this single phrase?

- One important simplification of Gödel's first incompleteness theorem was discovered by J. Barkley Rosser (1936). What is the difference between Rosser's notion of *simple consistency* and Gödel's original formulation of his first completeness theorem in terms of *ω-consistency*? Goldbach's famous unproven conjecture is that every even number greater than 2 is the sum of two primes. How is Rosser's simplification related to the fact that Goldbach-like statements (i.e., statements with the same logical form as Goldbach's conjecture, known as Π-0-1 statements) that are undecidable must be true?

- Gödel's incompleteness theorem, contrary to some misstatements, does not imply that *every* consistent formal system is incomplete. The Theory of Real Numbers, for example, is complete. How is this possible since the Real Numbers include the Natural Numbers of arithmetic? Moreover, certain subtheories of Peano Arithmetic, such as Presberger Arithmetic (1928), are decidable.

- Four years after the publication of Gödel's incompleteness results, Gerhard Gentzen (1935) published a proof of the consistency of elementary arithmetic making use of a generalized version of mathematical induction, known as transfinite induction. Why doesn't Gentzen's result conflict with Gödel's Second Incompleteness Theorem, which concerns the unprovability of consistency for a wide spectrum of formal systems?

Chapter 2, "The Incompleteness Theorem: An Overview," introduces the reader to the First Incompleteness Theorem, its relation to Hilbert's *Non Ignorabimus* view of mathematics, and its irrelevance with regard to explaining the "Postmodern condition." Chapter 3, "Computability, Formal Systems, and Incompleteness," explains the conceptual connections among the logical notions of computability, formal systems, and incompleteness. These initial chapters of Franzén's book, then, give the beginning logic student a correct and concise account of what the Gödel incompleteness theorems actually *do* say.

…

Readers who are non-mathematically-inclined but intrigued by the many claims about the implications of Gödel's work will find Franzén a sober and reliable guide in explaining what Gödel's theorems do *not* say. For example, does Gödel's theorem show that a Theory of Everything (TOE) in theoretical physics is impossible? Do Gödel's theorems refute the strong Artificial Intelligence (AI) thesis that the human mind can be modeled by a computer? "No mathematical theorem," Franzén notes, "has aroused so much interest among nonmathematicians as Gödel's incompleteness theorem." Indeed, Franzén's book grew out of taking on the exhausting task of commenting on the seemingly inexhaustible erroneous references on the internet to Gödel's incompleteness theorems.

Franzén discusses misuses of the incompleteness theorems in theoretical physics and theology (Chapter 4), in skeptical arguments about mathematical knowledge (Chapter 5), and in the Lucas-Penrose arguments about the limitations of Artificial Intelligence (Chapter 6). He dispatches his task with great clarity and a little self-reflexive humor. After acknowledging his colleagues in the preface, Franzén drolly comments: “For any remaining instances of incompleteness or inconsistency in the book, I consider myself entirely blameless, since after all, Gödel proved that any book on the incompleteness theorem must be incomplete or inconsistent. Well, maybe not.”

“Gödel’s theorem is an inexhaustible source of intellectual abuses,” note Alan Sokel and Jean Briemont in *Fashionable Nonsense: Postmodern Intellectuals’ Abuse of Science* (1997), which is based on the famous hoax in which Sokel submitted a parody of a postmodern article that was accepted for publication. Had Franzén limited his sites to debunking postmodern, political or poetic invocations of Gödel’s theorem that were “obviously nonsensical,” this book could easily have settled into a smugness that comes from dispatching strawmen arguments.

Franzén aims higher. Even competent commentators have erred in misleading the public. Franzén, for example, criticizes Nagel and Newman for claiming that Gödel proves “that it is impossible to establish the internal logical consistency of a very large class of deductive systems—elementary arithmetic, for example—unless one adopts principles of reasoning so complex that their internal consistency is as open to doubt as that of the systems themselves.” Franzén criticizes Freeman Dyson and Stephen Hawking for invoking Gödel’s theorem in their skeptical arguments against the possibility of a Theory of Everything (TOE). Gregory Chaitin, widely acknowledged for his information-theoretic interpretation of Gödel’s theorem, is still taken to task for this misleading metaphor: *“Stated in terms of information theory, Gödel’s theorem says that a 100 pound theorem cannot be derived from 10 pound axioms.”* Let us consider these criticisms in more detail.

In Chapter 4 Franzén discusses the claim that, because of Gödel’s theorem, the physicist’s dream of a Theory of Everything is not only unattained, but theoretically unattainable. In “The World on the String” in the *New York Review of Books* (2004), Freeman Dyson argued: “Another reason why I believe science to be inexhaustible is Gödel’s theorem.... His theorem implies that pure mathematics is inexhaustible. No matter how many problems we solve, there will always be other problems that cannot be solved within the existing rules. Now I claim that because of Gödel’s theorem, physics is inexhaustible too.” In his talk “Gödel and the End of Physics,” Stephen Hawking has argued similarly: “In the standard positivist approach to the philosophy of science, physical theories live rent-free in a Platonic heaven of ideal mathematical models... But we are not angels who view the universe from the outside. Instead, we and our models are both part of the universe we are describing. Thus, a physical theory is self-referencing, like in Gödel’s theorem. One might therefore expect it to be either inconsistent or incomplete.”

Do Gödel’s theorems have such universal implications? Drawing on Feferman’s reply to Dyson in the *New York Review of Books* (www.nybooks.com/articles/17249), Franzén explains: *“The basic equations of physics, whatever they may be, cannot indeed decide every arithmetical statement, but whether or not they are complete considered as a description of the physical world, and what completeness might mean in such a case, is not something that the incompleteness theorem tells us anything about.”* In other words, the incompleteness of the arithmetic component of a physical theory need not imply any incompleteness in the description of the physical world.

In Chapter 5 Franzén critically discusses the claims advanced by J. R. Lucas (1961), and updated more recently by Roger Penrose in his *Emperor’s New Mind* (1989) and *Shadows of the Mind* (1994). Lucas argued that no matter how complicated a machine we construct, it will correspond to a formal system, which, in turn, will be subject to a Gödelian construction for finding a formula unprovable in that system. Defending Lucas’ conclusion, Penrose updates the argument in an attempt to show that the aspirations of strong Artificial Intelligence (AI) are doomed to failure, going on to conjecture that a non-computational extension of quantum mechanics will someday provide a theory of consciousness.

Gödel’s own remarks on the subject (in his unpublished 1951 Josiah Willard Gibbs Lecture at Brown University, see vol. III, *Collected Works of Kurt Gödel*, edited by Feferman *et. al.*) are more cautious and nuanced:

> The human mind is incapable of formulating (or mechanizing) all its mathematical intuitions. I.e.: If it has succeeded in formulating some of them, this very fact yields new intuitive knowledge, e.g. the consistency of this formalism. This fact may be called the ‘incompletability’ of mathematics. On the other hand, on the basis of what has been proved so far, it remains possible that there may exist (and even be empirically discoverable) a theorem-proving machine which in fact *is* equivalent to mathematical intuition, but cannot be *proved* to be so, nor even proved to yield only *correct* theorems of finitary number theory. The second result is the following disjunction: *Either the human mind surpasses all machines (to be more precise: it can decide more number-theoretic questions than any machine) or else there exist number theoretic questions undecidable for the human mind.*

Criticizing Lucas and Penrose, Franzén argues that “we have no basis for claiming that we (‘the human mind’) can out prove a consistent formal system” because Gödel’s theorem only implies the *equivalence* of the consistency of the formal system and the Gödel statement asserting its own unprovability. In general, however, we have no guarantee that the formal system in question is consistent, an assumption required for us to draw the conclusion there is a truth unprovable in the formal system.

And what about the weaker claim that there could not be any formal system that exactly represents the human mind as far as it ability to prove arithmetical theorems is concerned? Franzén criticizes Hofstader’s reflections to this effect from *Gödel, Escher Bach*, noting Hofstader’s informal remarks have at least “the virtue of making it explicit that the role of the incompleteness theorem is a matter or inspiration rather than implication”:

> The other metaphorical analogue to Gödel’s Theorem which I find provocative suggests that ultimately, we cannot understand our own minds/brains.... All the limitative theorems of mathematics and the theory of computation suggest that once your ability to represent your own structure has reached a certain critical point, that is the kiss of death: it guarantees that you can never represent yourself totally.

In such metaphorical statements, Franzén notes, the inability of a formal system to prove its own consistency is interpreted as the inability of the system to “analyze or justify itself, or as a kind of blind spot.” The problem with such a view is that “the metaphor understates the difficulty for a system to prove its own consistency.... [T]he unprovability of consistency is really the unassertibility of consistency. A system cannot truly postulate its own consistency, quite apart from questions of analysis and justification, although other systems can truly postulate the consistency of the system (p. 125).”

...

Franzén’s first two goals were to explain accurately what Gödel’s theorems *do* say to the beginning logic student and to curb the enthusiasm of the non-mathematically inclined who have heard exaggerated claims about the philosophical and mathematical implications of Gödel’s theorem by pointing out what they do *not* say. Gödel’s theorems, as Franzén notes, are stunning and significant enough “without any exaggerated claims for the[ir] revolutionary impact.” Franzén’s book will also be of interest to logicians who want a model of sober clarity for explaining the philosophical perspectives opened up by Gödel’s work.

In Chapter 7 Franzén discusses the conceptual connections among Gödel’s Completeness Theorem, non-standard models of arithmetic, and the Incompleteness Theorems. Chapter 8 covers misleading formulations of incompleteness in terms of Kolmogorov-Chaitin complexity. Gregory Chaitin is known for his information-theoretic interpretation of Gödel’s theorem (1965) and for his discovery of the Halting Probability Ω (also known as Chaitin’s number). As Chaitin touts his results in *The Unknowable* (1999): *“In a nutshell, Gödel discovered incompleteness, Turing discovered uncomputability, and I discovered randomness—that’s the amazing fact that some mathematical statements are true for no reason, they’re true by accident.”* However, Chaitin’s informal explanation that *“... if one has ten pounds of axioms and a twenty-pound theorem, then the theorem cannot be derived from those axioms”* is misleading. In a recent book *Metamath* (2005), Chaitin expands upon this informal account: *“Now we’re really going to get irreducible mathematical facts, mathematical facts that ‘are true for no reason,’ and which simulate in pure math, as much as is possible, independent tosses of a fair coin....”* The problem with Chaitin’s informal explanation, as Franzén points out, is that Chaitin’s version of the Gödel’s theorem does not deal with the complexity of the theorems themselves but instead with theorems that are statements *about* complexity.

There is, moreover, an intriguing connection between Gödel’s incompleteness theorem and axioms of infinity: postulating the existence of various infinite sets has formal consequences for elementary number theory that cannot be proved by elementary means. Most of mathematics done today can be formalized within Zermelo-Fraenkel set theory with the axiom of choice (ZFC). ZFC minus the axiom of infinity, $\mathrm{ZFC}^{-\omega}$, is equivalent in its arithmetic part, to Peano Arithmetic, and so the Gödel incompleteness theorems apply. Therefore, $\mathrm{ZFC}^{-\omega}$ is incomplete and does not imply its own consistency. It turns out that in ZFC (which includes the axiom infinity) can prove the consistency of $\mathrm{ZFC}^{-\omega}$. So here we have an example in which adding an axiom of infinity to a theory (in particular, $\mathrm{ZFC}^{-\omega}$) yields new arithmetical theorems (the consistency of $\mathrm{ZFC}^{-\omega}$) not provable within that original theory. Stronger axioms of infinity extending version of ZFC also yield new arithmetical theorems not provable in the theories they extend. Franzén remarks: “From a philosophical point of view, it is highly significant that extensions of set theory by axioms asserting the existence of very large infinite sets have logical consequences in the realm of arithmetic that are not provable in the theory that they extend.”

As yet, no arithmetical problem of traditional mathematical interest is known to be among the new arithmetical theorems of extensions of ZFC by axioms of infinity. However, a step in this direction was taken with the Paris-Harrington Theorem (1977). The Paris-Harrington Theorem is related to Ramsey's theorem, which concerns complete graphs (i.e., graphs where each vertex or "node" is connected with a line or "edge" to all of the other vertices) colored with two colors. According to Ramsey's theorem, for each pair of positive integers k and l greater than 2, there exists an integer $R(k, l)$ (known as the *Ramsey number*) such that any graph with $R(k, l)$ nodes whose edges are colored red or green will either have a completely green subgraph of order k or a completely red subgraph of order l. For example, at any party with at least six people, there are either three people who are all mutual acquaintances (each one knows the other two) or mutual strangers (each one does not know either of the other two). The Paris-Harrington Theorem, a combinatorial strengthening of Ramsey's theorem, was the first "natural" statement found to be true but unprovable in Peano Arithmetic.

At the 1930 "Epistemology of the Exact Sciences" Conference in Königsberg, Gödel quietly announced his first incompleteness theorem. Among conference participants were such eminent logicians Rudolf Carnap and Arend Heyting, but only John von Neumann appreciated the profound significance of Gödel's incompleteness theorem. Not long afterward, von Neumann realized that the second incompleteness theorem could be obtained by formalizing the argument for the first. Communicating his discovery to Gödel in a letter, von Neumann graciously declined to publish when Gödel informed him that this stunning theorem was already discussed in his forthcoming "On Formally Undecidable Propositions in *Principia Mathematica* and Related Systems I" (1931), which would become the most celebrated achievement of 20th century logic.

What was Gödel's Second Incompleteness Theorem and what effect did it has on Hilbert's program? In addition to constructing the Gödel statement G for the formal system S, the argument establishing the implication "if S is consistent, then G is not provable in S" could be carried out with S itself. Moreover, the property of being a Gödel number of a proof in S is a computable one, and so 'S is consistent' is a Goldbach-like statement, a statement which if false, can be shown to be false by a computation. Thus, Gödel's Second Incompleteness theorem follows: if S proves the statement $\mathrm{Con}(S)$ expressing 'S is consistent' in the language of S, then S proves G, and hence S is in fact inconsistent. Hilbert's metamathematical program calling for consistency proofs for formal systems such as arithmetic in which all finitistic arguments can be formalized was effectively dashed by the Second Incompleteness Theorem.

Franzén carefully points out three common misconceptions about the Second Incompleteness Theorem. "First, it is often said that Gödel's proof shows G to be true, or to be 'in some sense' true. But the proof does not show G to be true. What we learn from the proof is that G is true if and only if S is consistent. In this observation, there is no reason to use any such formulation as 'in some sense true'..." Secondly, Gödel's theorem does not rule out consistency proofs using methods not formalizable within Peano Arithmetic. Thirdly, "[a]nother aspect of the second incompleteness theorem that needs to be emphasized is that it does not show that S can only be proven consistent in a system that is *stronger* than S." For example, Gentzen proved the consistency of Peano Arithmetic (PA) in 1936 by application of an arithmetically expressible instance of transfinite induction up to Cantor's ordinal ε_0 (the least fixed point under ordinal exponentiation to the base ω), while otherwise using arguments that can be formalized in a very weak subsystem of PA. So the consistency of PA is proved in a system that is overlaps in part with PA but is not an extension of it.

On the other hand, it has been argued that if a system S like PA has been accepted as intuitively true then one ought to accept the consistency statement $\mathrm{Con}(S)$ for S. That will give rise to a new formal system S' obtained by adjoining $\mathrm{Con}(S)$ to S. Now S' is also intuitively true, so the process can be iterated, in fact through the constructive transfinite. Alan Turing (1939) showed that one could obtain completeness for Goldbach-like statements for ordinal logics obtained by iterating consistency statements into the constructive transfinite starting with PA. Later, Feferman (1964) showed that one could obtain a progression that is complete for all arithmetical statements by iterating certain reflection principles. Franzén's other book, *Inexhaustibility: a non-exhaustive treatment* (ASL Lecture Notes in Logic #16, 2004) contains an excellent exposition of the incompleteness theorems and the reader is led step-by-step through the technical details needed to establish a significant part of Feferman's completeness results for iterated reflection principles for ordinal logics.

Torkel Franzén's untimely death on April 19, 2006 came shortly before he was to attend, as an invited lecturer, the Gödel Centenary Conference, "Horizons of Truth", held at the University of Vienna later that month. This, and his invitations to speak at other conferences featuring a tribute to Gödel, testifies to the growing international recognition that he deserved for these works.

Acknowledgements: I wish to thank Solomon Feferman for substantive and insightful correspondence during the preparation of this review, Robert Shrock, C. N. Yang Institute for Theoretical Physics, for assistance with

LaTeX to submit my "Gödel Incompleteness, Complexity and Chaos" for the proceedings from the *Kurt Gödel Centenary Symposium* published by the *Collegium Logicum of the Kurt Gödel Society,* and Teresa Spörk-Greenwood for support for my participation in the Gödel Centenary "Horizons of Truth" Conference, University of Vienna, April 28-29th, 2006

Gary Mar
Department of Philosophy
Stony Brook University
Stony Brook, New York, USA 11794-3750
Gary.Mar@stonybrook.edu

4 Hao Wang's Logical Journey in Life

Abstract

Hao Wang (1921-1995) was a prolific philosopher and researcher known for his close intellectual companionship with Kurt Gödel during the last decade of Gödel's life, yet "rather little has been published on Wang's considerable body of work or on the man's personality and unusual personal history."[16] Wang spent the first 25 years of his life in China, where he studied both Chinese and Western philosophy earning an MA in philosophy at Tsinghua University in 1946 before studying at Harvard. Wang study of mathematical logic in China was sufficient advanced for him to complete his Ph.D. under Quine in less than two years. As his Chinese classmate and lifetime friend He Zhaowu remarked at Wang's memorial: "Yet his whole life from beginning to end was 'chock full of contradictions,' be it in his thought, in his studies, or in his personal life. *These contradictions were not only of his own, they belonged also to a part of the difficult course of an entire generation and people.*"[17] The goal of this article is to see what light Wang's life and philosophical work can shed on Chinese philosophy and the Western analytic tradition.

Keywords: History of Analytic Philosophy, Chinese Philosophy, Orientalism, Hao Wang, Gödel, Wittgenstein, Critical Race Theory, Asian American Philosophy[18]

[16] Gary R. Mar is an Associate Professor of Philosophy at Stony Brook University where he is also the Founding Directors of the Philosophy Department Logic Lab and the Asian American Center. He was the last dissertation student of the 20th century logician Alonzo Church and co-author with Kalish and Montague of the logic textbook *Logic: Techniques of Formal Reasoning* (second edition, OUP) and the catalyst for the expansion of the APA Committee on Asian and Asian-American Philosophers and Philosophies. Email: Gary.Mar@stonybrook.edu.
Charles Parsons and Montgomery Link in their preface to *Hao Wang: Logician and Philosopher*, a collected work published 15 years after Wang's death, p. 1.

[17] He [1995], translated in Parsons and Link [2011], 49, italics mine.

[18] This paper is the result of participating in a panel "The Analytic Tradition and Chinese Philosophy" at the 2016 Eastern Division Meeting of the APA sponsored by the Committee for Asian and Asian-American Philosophers and Philosophers. For the history of the usually long name of this committee, see (*APA Newsletter on Asian and Asian-American Philosophers and Philosophies*, Spring 2003 (vol. 2, no. 2), 2).

4.1 Hao Wang's Journey.

On May 20, 1921 Hao Wang Hao (or, in Chinese, 王浩 *Wáng Hào*) was born into a family of intellectuals in Jinan, Shandong Province, China. Hao learned from his parents, who were teachers, both Chinese tradition and modern European-American approaches to science, art, and politics. The influence of Western approaches was brought about by Sun Yat-sen's Revolution of (1911–12) that overthrew the Qing (or Manchu) Dynasty. Wang early education was in the context the "Post-May Fourth Movement in China" (Wang's phrase in (Wang 1993), p. 40), which was inspired by the May 4th, 1919 student demonstration in Tiananmen Square to protest Japanese and Western imperialism in the aftermath of World War I.

Wang, Hao. 1993. From kunming to new york.

"During the War of Resistance"[19] (the second Sino-Japanese 1937 - 1945), Hao earned B.Sc. Mathematics from National Southwestern Associated University (a consortium of Peking, Tsinghua and Nankai Universities) with the intention of using mathematics as a foundation for philosophy, which Wang pursued rather than engineering which was in vogue at the time. In 1941 Hao married Yenking, his first wife, which was the same year that both China and the U.S. entered into World War II. Wang wrote his "virgin work" in philosophy (in Chinese), a study of Hume's problem of induction in 1942, which was also the year President Franklin Roosevelt issued Executive Order 9066 authorized the force relocation and incarceration of 110,00 persons of Japanese descent living in America, 2/3 of whom were American citizens. In 1943 about 30 million people were dying of starvation in China, while in the U.S. the Magnuson Act was passed to repeal the 1882 Chinese Exclusion Act, the first immigration law in U.S. history to target a group for exclusion by race and class, given that China and the U.S. were allies.

[19] Wang's phrase in Wang [1993].

Having taken all the courses in mathematical logic offered at his university in China, Hao agreed to form a reading group with two of his professors to work through Hilbert and Bernay's *Grundlagen der Mathematik*. After Hao presented the first chapter, however, the professors were unable to continue. Wang notes, "*I also became lazy... [and] it wasn't until seven or eight years later, when I was going to teach a class on that subject, that I finished reading the whole book.*"[20] Wang's Chinese writings include "The metaphysical system of the New Lixue" [1944], "Language and Metaphysics" [1945]. Wang characterizes the time of his education as during the "Post-May Fourth Movement" which was characterized with disillusionment with the Chinese Republic and traditional Chinese culture to protect China from imperialism. A recurrent theme of Wang's Chinese writings was the dilemma facing Chinese intellectuals in light of the Opium Wars (1839 –1842, 1856 – 1860).

[20] Wang [1982], Parsons and Link [2011], 31.

In 1945 Wang earned his M.A. in Philosophy at Tsinghua University. “During the defense,” Wang recalled, “Professor Shen Youding asked me why I wanted to study philosophy. I said I was interested in the human dilemma. Professor Shen said, ’In the West it is literature that focuses on the human dilemma, not philosophy.” This was the same year that Truman ordered atomic bombs to be dropped on Hiroshima (Aug. 6th) and Nagasaki (Aug. 9th) after which Japan surrendered (Aug. 15th).

The following year at the age of 24 Wang left China on a U. S. State Department scholarship, to study philosophy at Harvard.[21] Given his background in mathematical logic in China, Wang was able to complete his dissertation *An Economical Ontology for Classical Arithmetic* under Quine in one year and eight months, becoming Quine’s fifth dissertation student. Wang then became a Junior Fellow of the Society of Fellows at Harvard. During this time Wang was able to find an elegant repair for Quine’s inconsistent first edition of *Mathematical Logic* [1946] and demonstrate the consistency of his proposal relative to Quine’s *New Foundations,* a repair adopted and acknowledged by Quine in the second edition (Quine 1951). In October of 1949 Mao Zedong declared the existence of the People’s Republic of China (PRC), and China, who had been an ally, became a communist threat. McCarthyism created a climate of anti-Communism domestically and the U.S entered into the Korean War (1950-53).

[21] The Chinese Exclusion Act [1882], and its various amendments (strengthened in 1884 and expanded by the Scott Act (1888)) and renewed (the Geary Act (1982) and again in 1902 without termination date) made Chinese immigrants permanent aliens by excluding them from U.S. citizenship, so that Chinese men in America had little chance of starting families in America. Although the Chinese Exclusion Act was “repealed” by the Magnuson Act (1943), this act only raised the quota of 100 Chinese per year from anywhere in the world to 105 as compared to a quota of 65,721 immigrants from Great Britain and Northern Ireland. In preparation for entering into World War II, through the “Lend-Lease” program, President Franklin D. Roosevelt had $1.6 billion to invest in China.

Quine, Willard van Orman. 1951. Two dogmas of empiricism. *Originally in*The Philosophical Review** 60. 20–43,.

During the academic year 1950-1951, Wang studied in Zurich under the auspices of Paul Bernays, where he pursued the philosophy of mathematics, with a specialization in the type of predicative mathematics advocated by Bernays. Wang worked on questions of the relative strengths of axiomatic set theories and pioneered the revival of Hermann Weyl’s work on predicative mathematics, which takes its motivation from Russell’s “Vicious-Circle Principle.” In 1953 Wang contributed his own principle of *autonomous iteration* as a natural extension of *predicativism.*

In the Spring of 1955 Wang was invited to be the second philosopher to deliver the prestigious John Locke Lectures at Oxford and then became a reader in the Philosophy of Mathematics at Oxford. This period coincided with the beginning of U. S. involvement in Vietnam (1955 – 1975). In 1956 Wang received a letter from Chancellor Ma Yinchu of Peking University offering him a teaching position, but Wang declined at the time because of interest in pursuing philosophy. Two years later the Cultural Revolution forced Chancellor Ma to resign making it impossible for Wang to retrun to China. Jane Wang (*Hsiaoching*) was born in Oxford to Hao and Yenking in 1957.[22]

[22] Jane would grow up and become a jazz musician (Boston) and her brothers San-You and Yi-Ming became a doctor (Boston) and an astrophysicist (Washington), respectively.

In the Spring of 1958 Wang arranged for his teacher Professor Jin Yuelin who had promoted logic in China, to report to a meeting of Oxford philosophy professors in which Jin explained why he had abandoned academic philosophy and become a Marxist. Wang, who wrote beautifully but whose English could be difficult to understand, remarked, “*Most of the professors who heard the lecture felt his proof a bit too simplistic. But because Professor Jin’s British English was especially elegant and polished, the majority of Oxford professors treated him with utmost respect.”*[23] At this time, the Soviet Union, under the influence of Marxist philosophy, denounced mathematical logic *“as a conceptual game of capitalist class idealism”.*[24]

[23] Wang [1982], in Parsons and Link [2011], 35.

[24] He [1995] in Parsons and Link [2011], 50.

Upon returning to Harvard, Wang offered a professorship, not in Philosophy where most of his contributions, research and passion had been devoted, but in Applied Mathematics. Ahead of his time, Wang found ways to create academic partnerships with the emerging computer industry. While at IBM Wang (Wang 1961) invented a representation of formal systems known as Wang Tiles, graphically represented by square tiles with a color on each side. Wang’s student Robert Berger (Berger 1966) demonstrated in his dissertation the Undecidability of Wang’s Tiling Problem by showing its decidability would contradict Turing’s proof of the *Unsolvability of the Halting Problem.* Berger also found a counterexample to Wang’s conjecture that there were no aperiodic Wang Tilings.[25] Wang popularized his geometrization of logic in his *Scientific American* article, “Games, logic and computers” (Wang 1965).

Wang, Hao. 1961. Proving theorems by patternrecognition—II”. *Bell System Technical Journal** 40**1. 1–41.

Berger, Robert. 1966. The undecidability of the domino problem,”*memoirs. *The American Mathematical Society** 66. 72.

[25] Computational ideas have been a catalyst for contemporary philosophical investigations and speculations, e.g., Wang Tiles serve as the plot engine for a short story “Wang’s Carpets” by Australian science fiction writer Greg Bear [1995], a reflection on the nature of reproduction, life, and intelligence in a post-human context.

Wang, Hao. 1965. Games, logic and computers. **ScientificAmerican** 5(v.). 98–106.

Becoming an American citizen in 1967,[26] Wang moved his academic home from Harvard to Rockefeller University where he headed a research group in logic and computation. Wang began corresponding with Gödel while helping select and clarify various passages on the logical writings of Skolem for Jean van Heijenoort’s classic anthology *From Frege to Gödel.*[27] In October of that year Wang began his weekly conversations with Gödel at the Institute for Advanced Studies at Princeton. The first series of conversations from 1971 - 1972 influenced significant parts of Wang’s *From Mathematics to Philosophy* (Wang 1974), which were published subject to Gödel’s approval. The second series of conversations with Gödel began in the Fall of 1975. These remarkable conversations are chronicled and commented upon in four of Wang’s books beginning with *From Mathematics to Philosophy* (1974) and continuing in *Beyond Analytic Philosophy* (Wang 1986), *Reflections on Kurt Gödel* (Wang 1987) and his posthumously published *A Logical Journey: From Gödel to Philosophy* [1997].

[26] This was two years after the Hart-Celler Naturalization and Immigration Act (19654), which removed the quota system and, for the first time in U. S. history, put Asian immigrants on an equal footing with other nations.

[27] van Heijenoort [1967], vii.

Wang, Hao. 1974. *From mathematics to philosophy*(humanities press. New York.

Wang, Hao. 1986. **Beyond analytic philosophy**. Cambridge,Mass: MIT Press.

Wang, Hao. 1987. **Reflections on kurt gödel**. Cambridge,Mass: MIT Press.

Even with the publication of Solomon Feferman's *et al.* monumental 20-year project, the *Collected Works of Kurt Gödel*, vols. I–V [1986 - 2003], Wang's pioneering books remain significant because they provide the philosophical background missing in Gödel's published papers and even in his *Nachlass.* The philosophical themes or threads that run through Gödel's work remain unstated not because they were secret but because they were so interwoven into the fabric of his thought that it was unnecessary for him to spell them out. Wang books on Gödel made *mathematicians* aware of the *philosophical* views that led to Gödel's remarkable theorems and showed *philosophers* how philosophical views could be made *mathematically* rigorous. Wang's accounts of his conversations with Gödel [1974, 1985, 1987, 1996] remain invaluable to philosophy. In China Hao Wang "was publically acknowledged as the inheritor of the mantel of Gödel", but as Parsons and Link add in a footnote "an acknowledgement that would not extend globally."[28]

[28] He [1995] in Parsons and Link [2011], 50, footnote 4.

4.2 Hao Wang's Contradictions in Life.

Charles Parsons and Montgomery Link, in their editorial preface to *Hao Wang: Logician and Philosopher* [2011], a memorial volume not assembled until 15 years after Wang death, make an intriguing observation:

> Clearly [Wang] never lost his identification as Chinese, and we think he resisted being categorized as Chinese-American or Asian-American, although he did become an American citizen in 1967. The influence on him of Chinese culture and thought comes through in a number of his writing, maybe especially in his reaction to the very Western rationalism of Kurt Gödel, which fascinated him but which he could never fully embrace.

There is no reason to question the editors' opinion with regard to Wang's preferences about terms he would use to express what we would today call his "identify politics." The asymmetry of the hyphenated terms "Chinese-American" or "Asian-American" typically implies that one is referring exclusively to a type of American. However, a different question can be raised. We can use terms with solidus (e.g., "Chinese/American" or "Asian/American") to refer to Wang's status as an individual of Chinese or Asian descent residing in America. What facts about the racial formation of Asian/American intellectuals during Wang's lifetime shaped their life prospects and career possibilities?

Charles Parsons has, perhaps, done more than any other philosopher to bring the unduly neglected accomplishments of Hao Wang to the attention of the APA. However, Parsons' repeated remarks on of Wang's "*Chineseness*" may hide the complexities of racial formation faced by Wang and his generation. Parsons notes that Wang essentially thought of himself "as simply Chinese, a member of the *Chinese diaspora* that has existed for centuries."[29] Parsons characterized Wang's extensive exchanges with Gödel as "conversations between two thinkers with synoptic ambitions, one of whom sought resources from earlier Western traditions, and the other of whom, however Western his training and career, *never ceased to be Chinese*."[30] Parsons characterized Wang's responses to Gödel's discussions of *Weltanschauung* (e.g., of pre-Kantian rationalism and Leibnizian optimism) as "show[ing] his *Chinese cultural background*."[31]

[29] Parsons [2002] 28, *italics mine.*

[30] Parsons, op. cit., *italics mine.*

[31] Parsons [2011], 78, footnote 30, *italics mine.*

The characterization of Chinese in America as part of the global diaspora of *"Overseas Chinese"* is often used to explain why Chinese failed to assimilate and to become American. This *cultural* explanation blames culture for the isolation and marginalization that Chinese/American[32] intellectuals faced and ignores the more subtle forces of racial formation operating on Oriental scholars in American academic institutions. The construction of the Oriental as a "*sojourner*" in the West whose life goal of the triumphant journey home often functioned as a rationalization. This dream of returning home served as a powerful defensive mechanism for many Chinese/American intellectuals coping with social isolation and professional discrimination.

[32] The solidus notation in 'Chinese/American' is here used *inclusively* for being both Chinese and American and avoids the asymmetry of the label 'Chinese-American' which typically refers *exclusively* to a type of American.

Professional Orientalism in the *careers* and *outlook* of academics of Asian descent in America has been examined by Henry Yu in *Thinking Orientals* (Yu 2001) a sociological study of historical transition from "gentlemanly to institutionalized Orientalism" experienced by intellectuals of Chinese and Japanese descent living in the American from 1920 – 1965. Yu's study focuses on the Chicago School of Sociology, spear-headed by Robert Ezra Park (1864–1944), who with the help of his mid-westerner colleagues, recruited Oriental scholars, but the same sociological forces and racial formations were faced by other academics of Asian descent.

Yu, Henry. 2001. **Thinking orientals: Migration, contact, andExoticism in modern america**. New York, NY: Oxford University Press.

In Chapter "Strangers from the Midwest: Robert Ezra Park and Other Men with Three Names" of *Thinking Orientals*, Yu explains why these Midwestern sociologists were so interested in Orientals and how their outlook was conditioned by orientalism. The conceptual link between the European *white immigrant* and the Negro *non-white non-immigrant* was provided by the Oriental, who was both *immigrant* and *non-white.* The "Oriental Problem" was to explain the cultural inability of Orientals to overcome the last stage of the *assimilation* cycle, which the Chicago School equated with becoming *American.* Park and his associates sought an explanation of the failure of Orientals to become American through a cycle of assimilation, not in terms of *racial formation,* but in terms of *cultural distance.*

Robert Park, blinded by his culturalist assumptions, came to think of the Oriental face as a mask that could disguise the inner thoughts and feelings, which is strikingly evidence in the following passage: *"I recently had the curious experience of talking with a young Japanese woman who as not only born in the United States, but was brought up in an American family, in an American college town, where she had almost no association with members of her own race.... I found myself watching her expectantly for some slight accent, some gesture or intonation that would betray her racial origin... When I was not able, by the slightest expression, to detect the Oriental mentality behind the Oriental mask, I was still not able to escape the impression that I was listening to an American woman in a Japanese disguise."*[33]

Quine in his autobiography *The Time of My Life* characterizes Wang as "persistently unhappy" during his time at Harvard. This begs the question: did Hao Wang have reason to be unhappy? Recall that Wang, having already studied mathematical logic in China, wrote his dissertation under Quine in under two years. Wang had repaired Quine's inconsistent system in *Mathematical Logic*, had spent five years at Oxford where he was invited to be the second philosopher to give the John Locke Lectures. The topic of Wang's lectures was "On Formalizing Mathematical Concepts." Wang had been preceded by O. K. Bouwsma and was followed by a distinguished list of philosophers: A. N. Prior, A. C. Jackson, Gregory Vlastos, Nelson Goodman, Jaakko Hintikka, Wilfred Sellars, Paul Lorenzen, Noam Chomsky, Donald Davidson, Saul Kripke, Hilary Putnam, and so forth. Wang's time at Oxford overlaps with his first period of engagement with the philosophy of Wittgenstein (1953 – 1958).

Upon returning to Harvard, Wang as offered a position, not in Philosophy, but in Applied Mathematics. Juliet Floyd reports: *"In conversation Wang told me that one reason he had found his professorship at Harvard unsatisfying is that he was appointed in the Department of Applied Mathematics, and not in Philosophy, where he felt the bulk of his efforts and his contributions had been made."*[34] Quine was a mid-westerner with roots in Akron, Ohio with three names, a fact that he proudly details in *The Time of My Life* (1986): he was the child of Colye Robert *Quine* and Harriet ("Hattie") Ellis *née van Orman* named after a mathematician.[35] One wonders why W. V. O. Quine did not intervene on Wang's behalf.

Juliet Floyd notes: "Wang's role in shaping more than one generation's understanding of the fundamental problems in logic and their history through the 1950s was significant, and not as widely acknowledged as it should be.... Because Wang taught at Harvard and Oxford, the cumulative impact of his teaching on the dissemination of logic was significant.... Hide Ishiguro has stressed to me how supportive Wang was of Michael Dummett during his early years teaching at Oxford, when logic was not a very popular subject among philosophers there."[36] Among Wang's students who went on to have distinguished careers as philosophers and logicians were Charles Parsons, Burton Dreben, Michael Dummett, Donald A. (Tony) Martin, and Hilary Putnam.[37]

Wang expressed his dissatisfaction with analytic philosophy in *Beyond Analytic Philosophy: Doing Justice to What We Know* (Wang 1986). Quine, who described himself as a "disciple of Carnap for six years," declared his independence from Carnap his celebrated "Two Dogmas of Empiricism" (Quine 1951) [1951, 1961].

[DATE]. Similarly, Wang playfully labeled Quine's philosophy "Logical *Negativism*" (a contrast that plays off of Carnap's "Logical *Positivism*") which captures Quine's peculiar combination of *local precision* and *global indefiniteness*, supported by such *negative* Quinean dogmas as the "*indeterminacy* of translation," the "*inscrutability* of reference," and, to borrow Gilbert Harman's phrase, the "*death* of meaning." Supported by vaguely enunciated pragmatic appeals to semantic holism, Quine's logical negativism not only "failed to do justice to what we know"—but, ironically Wang noted, failed to give an adequate account of logic, mathematics, and science, the disciplines from which Quine's philosophical positions drew their rhetorical strength. Wang's criticism of Quine was frustrating to analytic philosophers trained to publish "piecemeal exercises" of focused argument analysis and to diagnose arguments in terms of precisely defined logical failings.

Instead of adopting a *piecemeal* approach, Wang's writings aimed at *perspicuity*—clarity of the sort that can be found in an elegantly arranged mathematical proof or an insightfully arranged set of quotations. Wang was known for his ability to locate a philosophical position in *intellectual history*—a history not told dramatically but with discerning and judicious insight.[38] Floyd characterizes Wang's *synoptic* style of argument: "When arguments are found in Wang, they are deductions from his observations, or relative comparisons of the interest and correctness of different responses, concepts, or principles. The tone of the whole is always tentative, pluralistic, and … synoptic rather than systematic in aim."[39]

[33] Yu [2001], 67.

[34] Floyd [2011], 157 footnote 17.

[35] See http://www.quine.org/crq-tree.html.

[36] Floyd [2011], 146.

[37] Floyd ([2011], 87) herself regarded Wang as a mentor who provided her with an authentic example of what the Chinese call *xiansheng* (先生, teacher, master).

Wang, Hao. 1986. *Beyond analytic philosophy*. Cambridge,Mass: MIT Press.

Quine, Willard van Orman. 1951. Two dogmas of empiricism. *Originally in*The Philosophical Review* 60. 20–43,.

[38] Floyd [2011], 156.

[39] Floyd [2011], 18.

Wang was concerned that philosophy "do justice to what we know." He characterized his major statement on the philosophy of mathematics, *From Mathematics to Philosophy*, not in terms of *constructing* a grand *theory*, but in terms of seriously *collecting data*:

> This book certainly makes no claim to be philosophical theory or a system of philosophy. In fact, for those who are convinced that philosophy should yield a theory, they may find here merely data for philosophy. However, I believe, in spite of my reservations about the possibility of philosophy as a rigorous science, that philosophy can be relevant, serious, and stable. Philosophy should try to achieve some reasonable overview. There is more philosophical value in placing things in their right perspective than in solving specific problems. Both hasty speculations and piecemeal persistence on artificial issues tend to hamper cumulative progress in philosophy.[40]

[40] Wang [1974], x.

As an active researcher in the emerging field of computational logic, Wang was able to see the impediments imposed by Quine's overly narrow analytic conception of logic pronouncing limitations on the proper scope of philosophical and logical inquiry. Wang, in his contribution to Quine's "Schilpp volume" one year after his critique in *Beyond Analytic Philosophy*, sharpened his critical assessment: "for the working logicians, much of Quine's work is thought to be off the mainstream."[41]

[41] Wang's contribution in Schilpp [1986], 635.

Eckehart Köhler was scholar interested in Gödel's Platonism, who first contacted Wang in 1979 on the advise of Solomon Feferman, and who later in 1986 as founder of the Kurt Gödel Society in Vienna invited Wang to become its first president. Köhler in his "Collaborating with Hao Wang on Gödel's Philosophy" (Köhler 2011) in the memorial volume recounts a revealing with Wang:

Köhler, Eckehart. 2011. Collaborating with hao wang on gödel'sPhilosophy. (Parsons & Link, Eds.).

> I began explaining to Wang the well-known work of Herbert Simon, a major figure in Artificial Intelligence. He and his students had done some of the best work on scientific discovery, using AI methods. (I had at other times tried explaining related concepts of decision theory, statistical weighing of evidence, concepts of probability, in attempting to justify Gödel's famous claim that mathematical knowledge is analogous to empirical knowledge, especially…. Wang practically exploded. He was disgusted with Simon and did not want to hear a thing about him, and considered him an execrable logician. Wang would brook not contradiction, so I changed to topic (Köhler adds in a curious footnote: "Oddly enough, Simon… spent quite a lot of time in China and had learned Chinese.")

Köhler speculates on the source of Wang's anger:

> Without knowing any details of Wang's work in computer proofs, I am sure that he was a logician vastly superior to Simon, and I suppose he got sick of hearing form AI researchers that Simon was allegedly the first to do an AI logic proof. I am also sure that Simon was would agree with Wang that the proof he presented at the famous 1956 AI-conference was utterly trivial and without much inherent merit.

Let's examine some of the logical and historical details before drawing any conclusion about Wang's anger. Newell and Simon wanted to program a computer to carry out heuristic problem solving through symbol manipulation, and they considered such problems as playing chess, the solving of geometrical problems, and, as an afterthought, proving logical theorems. In 1955 Simon claimed "over Christmas Allen Newell and I invented a thinking machine"[42] because he had simulated "by hand" a proof from Whitehead and Russell's *Principia Mathematica*, but it wasn't until August 1956, that the *Logic Theorist* program actually run on RAND's Johnniac computer (named after von Neumann) completed the proof of theorem 2.01.

[42] McCorduck [1979], 116.

The *Logic Theorist* was given a list of axioms and definitions and kept in memory a list of previously proved theorems. When it is given a new logical theorem to prove, it runs through all the operations of which it is capable searching for a proof. Consider a proof of theorem 2.01:

$$2.01 \vdash \neg(p \to \neg p) \to \neg p$$

The program would have had as an axiom

$$\vdash (p \vee p) \to p,$$

which by a rule of substitution of $\sim p$ for p could obtain

$$\vdash (\neg p \vee \neg p) \rightarrow \neg p,$$

from which, in turn, by a rule of substitution of equivalents, in particular, an instance of the definition

$$p \rightarrow q \ :=:\ \neg p \vee q$$

could obtain the required theorem.

It should be kept in mind that Newell and Simon *structured* the problems given to the program, including the order in which they were given. In other words, it could be argued that the *Logic Theorist* did not by any stretch of the imagination show that one could mechanize Gödel's notion of intuition: the program was only doing what it had been programmed to do in a very limited realm of problem solving. Eventually the *Logic Theorist* was able to prove 38 of the first 52 propositional theorems. About half the proofs accomplished in less than a minute, most of the remainder taking between 1 – 5 minutes, some taking as much as 15 – 45 minutes.

How did this compare with Wang's accomplishments? In 1959 Wang programed an IBM 704 computer to prove all the logical theorems (over 350, including theorems involving not only propositional logic but also predicate logic and identity) of *Principia Mathematica* in less than 9 minutes. In contrast to Herbert, who triumphantly announced his accomplishment as showing the power of AI to simulate human logic, Wang regarded his accomplishment as demonstrating the "essential lack of conceptual richness" of Russell and Whitehead's *Principia Mathematica.*[43] Wang help to define the new research area of ATP (automated theorem proving) and to show how it rested upon results in mathematical logic.

[43] Floyd [2011], 164.

This work would be the basis of his later work (Wang [1960, 1961, 1963]) on the ∀∃∀ case of Hilbert's *Entscheidungsproblem.* Instead of stocking the computer program with *ad hoc* "heuristics," a research program which has since gone out of style, Wang programmed the "cut-free" formalisms of Herbrand the Gentzen and emphasized the importance of algorithmic pruning, which could eliminate useless terms in advance. Furthermore, Wang provided a well-conceived list of theorems of predicate calculus that could serves as "benchmarks" for judging the effectiveness of new theorem-proving programs. For these achievements Wang would eventually be awarded in 1983 on the recommendation of a committee composed of David Luckham, John McCarthy and chaired by Martin Davis the first *Milestone Prize for Automated Theorem Proving* (ATP) by the *International Joint Conference on Artificial Intelligence.*

To return to Köhler account, Köhler found Wang's reaction to be inconsistent:

> A few days after the clash on Simon, while musing on Gödel's notion of intuition, I mentioned how close it seemed to what AI people (following Simon) now call heuristics; … Wang again rebuffed me, calling this too "scientistic." I slyly hinted that his refusal to consider the idea violated his own injunction "to do justice to what we know." Wang of course immediately got the point and did not reply….

Perhaps even more inconsistent is the suggestion that Wang's silence indicated his *capitulation* to Köhler's criticism. One cannot dismiss the *legitimacy* of Wang's anger on the grounds of Wang's *incivility* or Wang's *silence.* Anger towards institutionalized discrimination can be a legitimate and healthy response. It is presumptuous for more privileged parties to assume that unrecognized parties expressing anger "be civil" before their complaints can be given a hearing, especially when this requires, in effect, that they case cease making forceful points about unfairness. Silencing of those in subordinate positions protects institutionalized discrimination by removing any record of wrong. The vicious circle principle of silencing can work in two ways—when one's colleagues dismiss one as a knower with a legitimate perspective due to one's anger and when one in frustration perceives one's colleagues as unwilling or unable to listen to one's testimony and so engages in self-censoring.

This *vicious silencing principle* is a form of racial formation is often at play in academic circles. This principle combines stereotyping with an asymmetrical application of a principle of civility. When someone from the dominant group is uncivil, this very incivility is valorized as being "forthright" or "refreshingly honest" or even "courageous." However, when those from a subordinated group are uncivil, perhaps uncharacteristically, this angry outburst is used as grounds for not taking their complaints seriously. For example, the silencing response "I can't hear you when you're angry" to Black or women philosopher is now widely acknowledged as ways of illegitimately delegitimizing the criticism.

Oientalism functions not only professionally but personally. Scholars of Asian descent establishing themselves as academics in America during Wang's generation would often *re-orient* their lives selectively around acceptable Asian values (e.g., Chinese culture and philosophy) while distancing themselves from cultural traits considered to be un-American (e.g., clannishness and indirectness). Ever since his youth Wang conceived of philosophy as a more holistic enterprise: "One might wish to say that the task of every philosopher is to describe eventually 'the world as I see it' and that his achievement is to be evaluated by how significant a picture he gives."[44]

[44] Wang [1974], 334.

Henry Yu's book takes as a case study Rose Hum Lee (1904–1964), a first generation Chinese American, who was the first *woman* to become chair*man* in an American university (to succeed Lee was forced not only to become more *modern* but also more *masculine*). In research papers, Lee adopted the cultural analysis of alienation of the Chicago School of sociology and wrote objectively:

> *To retaliate for not heeding their wishes, they spread tales of her supposed misdeeds in the Midwest, where she attended a famous university. The Chinatowners of this prairie city were delighted to have a plum to pick. No native-born of this group had ever obtained a doctorate in philosopher, though some received their master's and medical degrees... The local Chinese, instead of being sympathetic to professional attainments, disparaged this girl's achievements. Her personal life was the subject of slander, gossip, envy and conspiracy. There were no congratulations when her doctorate was awarded.*

Perhaps Wang's self-representation of himself as a "Chinese in exile" was conditioned by the discrimination he faced professionally and personally, a self-presentation that evolved during the two thirds of his life that he lived in America. While Wang was still entertaining thoughts of returning to China—after the birth of the PRC in 1949 but before the Cultural Revolution in China (1966 – 1976) brought about the public persecution of intellectuals—he shifted his research focus to computation with the thought that programming skills would be more useful to China than his research in philosophy. Martin Davis ([2011], 76) in the midst of his account of Wang's contributions to Automatic Theorem Proving and to Hilbert's *Entscheidungsproblem*, recalls a turning point in Wang's disillusionment:

> Like to many others, he eventually became disillusioned with the Maoist order. I remember particularly his telling me of his astonishment at a letter severely criticizing him that he received from his father, a secondary school teacher in China, during the Cultural Revolution. He understood with considerable distress that his father would have written such a letter only under great pressure.

Wang reminisced about how he and his classmate in China He Zhaowu visited each other in 1980 in Kunming and in New York:

> *I began to doubt my own most recent views. I had imagined certain facts that were grounded in wishful thinking; upon this foundation I had erected a logical construction that really was a castle in the sky..... I remember the pain that bitterness I suffered at the time because of the wrenching transformations of my thinking and the lost of my convictions.*"[45]

[45] Wang [1933] in Parsons and Link [2011], 41.

Influenced by Zhaowu's presentation of the poetic pentology of *Bodhisattva Barbarians*, Wang wrote a personal essay about *"homesickness in a foreign land"* but did not publish it regarding it, in the end, as *"sick with self-pity"* and *"nauseating."*[46]

[46] Wang [1992] in Parsons and Link [2011], 44.

4.3 Hao Wang's Philosophical Journey.

Wang believed that "Philosophy has a longer history and a larger diversity of traditions to select from" (Wang [1986], 192). As an *analytically trained* philosopher who was an active researcher in logic, Wang contributed an important *diagnosis* of the *misconceived* role of logic in Anglo-America analytic philosophy. Wang was in a position to comment knowledgeably on the contradictory roles *logic* played in Quine's analytic philosophy: logic was at once *excessively emphasized* and *ineffectively employed*:

> Directly and indirectly logic plays an important part in much of contemporary Anglo-American academic philosophy.... The way logic is commonly used in philosophy seems to me to do less than justice to the full richness of logic as a study of the foundations of mathematics; and that excessive emphasis on the importance of logic for philosophy, combined usually with a misapplication of logic, seems to me to have led to a far from balanced view of philosophy, especially as it is understood in the traditional sense. Moreover, the much-publicized juxtaposition of logic with positivism (or

empiricism or 'analytic' philosophy) has burdened logic with a guilt-by-association, resulting in a surprising ignorance of logic on the part of philosophers of other persuasions.[47]

[47] Wang [1976], ix.

As a *Chinese philosopher*, Wang was attached to a broader view of the philosophical enterprise "beyond analytic philosophy." In a concluding chapter entitled "Metaphilosophical observations" of *Beyond Analytic Philosophy*, Wang acknowledges the Chinese influences on his conception of philosophy: "… my professional training is nearly all in Western philosophy (much of it even logic-oriented), yet my formative years were lived in China. I have tried hard but have not been able to shake off my early conviction that philosophy is not just one subject more or less like any other, but something special. Even today such a belief, I think persists in China…. I continue to believe that philosophy should somehow be comprehensive and aim at a unified … outlook…. I find myself attached to the Chinese tradition of mixing together philosophy, literature, and history—a tradition that conditions and is conditioned by the central concerns of its philosophy; the interest in politics ties it and the concern with the unity of nature and person overlaps with art and literature."[48]

[48] Wang [1986], 194.

To advance the dialogue on such topics as "The Analytic Tradition and Chinese Philosophy," we need to overcome the *duality* of viewing philosophy in terms of "East versus West"—in particular, the unexamined tendency to view the relationship in terms of "*challenges*" or "*contributions*" to the mainstream. "Challenges" presuppose a false philosophical polarity, e.g., in discussing Eastern versus Western ethics in terms of polar concepts of *shame* versus *guilt* tends to stereotype the meaning of each and to discourage remembering cases of ethical reasoning or traditional virtues in which both qualities are present. "Contributions" presuppose one tradition as the philosophical *mainstream* thereby marginalizing the other as making "contributions."

Orientalism shaped the *disciplinary* self-understanding of philosophy. A kind of *professional orientalism*, also shaped the careers of many academics of Asian descent living in America (as well as England and Europe) as well as their *autobiographical* self-understanding of themselves as academics. Edward Said famously coined the term *orientalism*,[49] which is used by scholars of Cultural Studies for a type of depiction of the Orient (East Asia, South Asia, and especially, for Said, the Middle East) by writers, artists and scholars from the West. While Said's use of the term was far too sweeping and impressionistic, scholars in Asian American Studies have more carefully articulated various forms of orientalism.

[49] Edward Said's *Orientalism* (Vintage Books, New York, 1978), was *catalytic*, if not always philosophically *accurate*. Said's philosophical misconceptions were elaborated in my presentation "Critique of Orientalism" delivered at the Annual Meeting of the Association for Asian American Studies, April 24, 2009.

An essential aspect of Said's thesis was that orientalism was essential to the West's self-understanding and self-promotion. Consider Hegel's claims in his *Lectures on the History of Philosophy* (1825–1826):

> [In the East] conscience does not exist, nor does individual morality. Everything is simply in a state of nature, which allows the noblest to exist as it does the worst. The conclusion to be derived from this is that no philosophic knowledge can be found here…. The Eastern form must therefore be excluded from the history of philosophy…. Philosophy proper commences in the West.[50]

[50] Hegel [1882] in Haldane and Simon translation [1974], 97-98.

Even Bertrand Russell's more liberal lament over Great Britain's colonialist policies in *The Problem of China* (1922) appeals to orientalism to find cultural reasons for the dominance of the West:

> The British view is still that China needs a central government strong enough to suppress internal anarchy, but weak enough to be always obligated to foreign pressure…. Possession, which is one of the three things that Lao-Tze wishes us to forego, is certainly dear to the heart of the average Chinaman. As a race, they are tenacious of money—not perhaps more so than the French, but certainly more than the English or the Americans. Their politics are corrupt, and their powerful men make money in disgraceful ways…. Nevertheless, as regards the other two evils, self-assertion and domination, I notice a definite superiority to ourselves in Chinese practice. There is much less desire than among the white races to tyrannize over other people. The weakness of China internationally is quite as much due to this virtue as to the vices of corruption ….[51]

[51] Russell [1967], 552.

Wang himself noted the historical, if not conceptually necessary, collusion between Great Britain's imperialism and the aesthetic and autonomous morality embraced by its intellectuals:

> In the heyday of British imperialism, a rather influential ethical doctrine came out of Cambridge which signaled out esthetic enjoyments and personal affection as good in themselves. In theory, this doctrine does not reject the suggestion that for a long time to come eliminating miseries will be of more moral value than pursuing personal enjoyments. In practice, it has had the effect of enhancing the complacency of intellectual aristocrats.[52]

[52] Wang [1974], 329–30.

In 1985 Wang received honorary doctorates from Peking (China's 'Harvard') and Tsinghau (China's 'MIT') Universities, and was again seriously considering returning to China. Although China was reversing the repressions of the Cultural Revolution under Deng Xiaoping, Wang "sorrowfully torn up his invitation" to teach and do research in Beijing after the 1989 Tiananmen Square massacre.[53]

[53] Köhler [2011], 58.

Racial intermarriage was of special interest to the Chicago School of Sociology because in their minds it was ultimate proof of assimilation and hence, in their view, of becoming American. However, it should be kept in mind that anti-miscegenation laws were in effect since before the United States was a republic and remained in effect until ruled unconstitutional by the U.S. Supreme Court in 1967. For example, prior to the Cable Act of 1922, women, but not men, lost their U.S. citizenship if they married a foreign spouse. The Cable Act guaranteed such women could retain their citizenship but only if married to an "alien eligible to naturalization." Since at the time of the act's passage, Asian aliens were not considered to be racially eligible for U.S. citizenship, any woman who married an Asian alien lost her U.S. citizenship.

Around 1988 Wang met Hanne Tierney[54] at an invitation-only gathering of artists, academics and other accomplished persons called the "Reality Club." Hanne Tierney, who would become Hao's third wife, was a German immigrant, who lived in Prague and journeyed to America by way of England at the age of nineteen. She authored a successful children's book about her experience in America and then created an *avante-garde* style of puppetry that offered her a three-dimensional language of sculpture. According to Hanne, Hao characterized the difference between herself and himself: "You are an *immigrant*, I am an *exile*." Hao once explained to her that "China was God", that is, China served much the same purpose for exiled Chinese as God did in Western religion: what overseas Chinese desired from China was *forgiveness*, and the *good deeds* they performed in a hostile environment were done for China. The "Death of God" explains the depth of suffering caused by Wang's loss of his faith in Marxism and Maoism in 1979. Hao and Hanne visited China together in 1992.

[54] After reading Wang's obituary in the *New York Times*, I contacted his surviving wife Hanne Tierney, who provided these recollections.

4.4 Wang, Wittgenstein, and Gödel.

Wang approvingly quotes a mimeographed manuscript of Yueh-Lin Chin [1981]: "Chinese philosophers were all of them different grades of Socrateses."[55]

[55] Wang [1986], 194.

Towards the end of his life, Wang continued to work on a manuscript for *Gödel, Wittgenstein and Purity of Mind: Logic as the Heart of Philosophy* but had great difficulty in completing his manuscript and set the project aside. Despite his reservations about Gödel's overly optimistic faith in the "possibility of philosophy as a rigorous science," Wang unreservedly praised Gödel's way of life:

> *Gödel exemplified, I think, a way of life and work that inspires a greater faith in reason, questions the 'prejudices of the time' and stirs our imagination to strive for more autonomy by examining our largely derivative sense of what is important in life.*[56]

[56] Wang [1987], ix.

While Wang never became a *Wittgensteinian*, he seriously engaged Wittgenstein's *philosophical* insights. Wang's second period of engagement with Wittgenstein's philosophy spanned the years 1981 – 1995. Wang admired Wittgenstein's life-long struggle to achieve *perspicuity*: "*One main difference between Wittgenstein and most contemporary Anglo-American academic philosophers would seem to be the indulgence of the latter in clever, small arguments clouded by all sorts of extraneous detail.*" [57]

[57] Wang [1976], 348.

Let's summarize some of our conclusions thus far. As a *Chinese/American* philosopher, Wang was able to pursue philosophy in ways that embodied the virtues of both the Analytic and Chinese Traditions—e.g., combining the technical precision of Gödel's meta-mathematical approach to logic as well as the artistic and aesthetic perspicuity of Wittgenstein and Chinese philosophy. As an *analytically trained philosopher,* Wang was able to correct misconceptions of logic held by many Anglo-American philosophers due to the "guilt by association" in their minds of logic with positivism and the Vienna Circle, a view about the nature of logic that was not held by Gödel, who briefly attended the Circle. As a *mathematical logician*, Wang was able to provide a meta-mathematical perspective on logic to Chinese logicians. Wang quotes a mimeographed manuscript of Yueh-Lin Chin [1981] written in Kunming in 1943: "One of the features characteristic of Chinese philosophy is the underdevelopment of what might be called logico-epistemological consciousness."[58] As an *active researcher* in *mathematical logic* and *computation*, Wang was ahead of his time in proposing new graphical representations of computability, pioneering collaborations between academia and the emerging computer industry, and proposing new distinctions (e.g., feasible as opposed to theoretical computability). As an *Asian/American* academic, the trajectory, and limitations, of Wang's professional career sheds light on the unexamined practices of academic philosophy in the

[58] Wang [1986], 193.

Anglo-American tradition, and the complex sociological dynamics that entered into the formation of the professional and personal lives of the "Thinking Orientals" of a previous generation and still enters into the formation of Asian American philosophers.

Figure 4.1: Puppet of Hanne Tierney "Hanne Tierney in a Nutshell"

Figure 4.2: Hao Wang at Knoedler Gallery, 1993

4.5 Wang's Empty Boat.

From 1991–1994, Juliet Floyd spent time in the company of Hao Wang and his wife Hanne Tierney and she notes Wang's return the conception of philosophy that had attracted him as a youth in China:

> Wang was inclined to consider artists with as much respect as he did scientists.... Wang's literary ambitions became stronger over time, and importantly shape his final writings. These bear an important relation to his ideas about "intuition" and his interest in Wittgenstein, whom he came to regard as 'art centered' rather than 'science centered' in his conception of philosophy.... Wang felt the literary effects of Wittgenstein's writing were not irrelevant to the content of his philosophy. I think we can assume that Wang felt the same way about his own books.... In the manner of Walter Benjamin, or perhaps better, of his Chinese forebears, Wang often proceeded by arranging quotations, without interpretation, in an effort to draw out the reader's reflection and response, thereby showing, but not himself directly stating.

One of Wittgenstein's most memorable sayings in the *Tractatus* is his famous ladder metaphor:

> 6.54 My propositions are elucidatory in this way: he who understands me finally recognizes them as senseless, when he has climbed out through them, on them, over them. (He must so to speak throw away the ladder, after he has climbed up on it.)

Wittgenstein's ladder, Wang noted, evokes the raft of the *Diamond Sūtra*: [59]

> The dharma I am preaching is analogous to a raft (which is to be discarded after use); even the dharma can be discarded, *a fortiori* the non-dharma.

Wang completed his last book *A Logical Journey: From Gödel to Philosophy* just months before his death s week before his 74th birthday, which was published posthumously in 1996.

In 2001 Hanne Tierney in collaboration with Wang's daughter and jazz musician Jane Wang created a performance piece "How Wang-Fo Was Saved"[60] at the *FiveMyles Gallery* in Crown Heights, Brooklyn.

The script for "How Wang-Fo Was Saved" was adapted from on a children's book by the Belgian-born French novelist and essayist Marguerite Yourcenar (1903–1987) that retells an ancient Chinese legend: Captured as he is walking with a disciple, the elderly painter Wang-Fo is brought before the Emperor. Spoiled by the beauty of Wang-Fo's paintings, the Emperor complains that nothing in reality is as beautiful as the artist's depictions. Wang-Fo is therefore to be executed for his lies. The Emperor commands Wang-Fo to finish painting one last canvas. Dutifully serving his sentence by completing his last canvas, Wang-Fo first paints a lake, then he draws a rowboat. As the waters rise and fill the Emperor's throne room, old Wang-Fo climbs into the boat and rows away. What is left behind is an empty boat.[61]

Who can free himself from achievement

And from fame, descend and be lost

Amid the masses of men?

He will flow like Tao, unseen,

He will go about like Life itself

With no name and no home.

Simple is he, without distinction

To all appearances he is a fool.

His steps leave no trace. He has no power.

He achieves nothings, has no reputation.

Since he judges no one,

No one judges him

Such is the perfect man:

His boat is empty.

We shall end our reflections on Hao Wang's logical journey with a meditation on Wittgenstein's famous advice:

Wovon man nicht sprechen kann, darüber muss man schweigen.

Whereof one cannot speak, thereof one must pass over in silence.

Even Wittgenstein could not follow his own austere advice, so perhaps it is more fitting to conclude with the paradoxical humor and literary elegance of Zhūangzi:

> 荃者所以在魚，得魚而忘荃；蹄者所以在兔，得兔而忘蹄；言者所以在意，得意而忘言。
>
> 吾安得夫忘言之人而與之言哉！

The bamboo fish net exists for catching fish. Once the fish is caught, forget the net!

The rabbit snare exists for trapping rabbits. Once the rabbit is trapped, forget the snare!

Words exist because they are used for expressing meaning. Once the meaning is grasped, forget the words!

Where can I meet those who have forgotten words and have a word with them?[62]

Wang's many unsung contributions to the analytic tradition in philosophy and his life "chock full of contradictions" shed light not only on his own struggles but also on "*the difficult course of an entire generation and people.*" The goal of this article has been to explore how Wang's life could add to the conversation about the Analytic Tradition and Chinese Philosophy and to serve as a catalyst for diversifying what Wang once praised as "the elusive concept of an American spirit."[63]

4.6 TODO

cleaning bib

Figure 4.3: "How Wang-Fo Was Saved" FiveMyles Gallery, 2001

[59] Wang [1986], 100.

[60] See Alexis Sottile, *"Hanne Tierney's String Theory," The Village Voice,* 10/30/2001.

[61] Merton [1965], 115. 'Chuang Tzu' is another name for Zhūangzi.

[62] I am indebted to Professor T. K. Chu for calling my attention to these poignant and paradoxical words of Zhūangzi.

[63] Wang [1986], 195, remarking on how despite the fact that Quine was influenced by mathematical logic and Dewey by Hegelian philosophy, there is nonetheless a "certain partial convergence of views." I wish to thank my colleague Ed Casey for giving me the opportunity to participate in the APA panel on Chinese Philosophy and the Analytic Tradition, organized by Professors Chung-ying Cheng and Linyu Gu. I also wish to thank Hanne Tierney for talking with me and for photographs. When I asked why she had agreed to meet sight unseen, Hanne with a sparkle in her eyes, replied: *"You had me at 'Hao'!"* In an email following up our conversation, Hanne wrote: *"It was such a pleasure to me to meet you and talk about Hao. He would have enjoyed it all himself very much, but he also would have so completely known what Gary was talking about...."*

References