Algorithm and Antinomy

Jerry L. Sherman

In "The Ways of Paradox" Quine writes, "One man's antinomy is another man's falsidical paradox, give or take a couple of thousand years." That is, what is to one era a collision of cherished principles may later become an explainable mental illusion. In calling the antinomy a "crisis in the evolution of thought," Quine suggests that it forces fundamental change: it "packs a surprise that can be accommodated by nothing less than a repudiation of part of our conceptual heritage."[1] I agree, except that most attempted change has been toward a tighter and more complex structuring of thought, at the cost of intuitive meanings, while what is really required of us is to rethink the task of structuring itself. That is, these antinomies are a challenge to formalization and a mark of its limits. What gets repudiated, I believe, is not the intuitions that formalized thought fails to successfully contain, but the container itself, the algorithmic model.

As difficult as the liar paradox and other semantic paradoxes have been to formally solve, they have not been very difficult to diagnose. Diagnosis shows that thought itself is not undone by the paradoxes, but is their master. Yet diagnostic theory does not yield rules by which paradox can be avoided. I will argue that in principle such rules are impossible, because formal rules of logic depend upon the separation of form and content, and it is the essence of semantic paradoxes to disturb that separation.

Now to many people my claim that these antinomies will become less serious if we abandon the need to produce an algorithmic model of thought will have an absurd quality. "*Of course* antinomy results from the need for algorithm," one might say, "but that is like saying that antinomy results from the need to make sense!" Such a response equates algorithm with making sense. As an example, in an earlier paper on this subject that I wrote as a student of Bill Hart, I used the phrase "an algorithmic solution," and he wrote in the margin, "that is, a coherent, systematic understanding of truth." He did not see "algorithmic" as an optional quality in rationality but as essential to it. But I will argue that

even though the finding of rules is an important part of being rational, it is not rationality itself. It is better understood as an "agenda," a task attached to certain kinds of projects. As such, it can be set aside, in a sense, not because no one will hope to do it, but because its provisional nature is recognized. *If you* want to build Leibniz's mental calculator, you will have to deal with this problem. But if you only want to make sense, i.e., be rational, these paradoxes do not threaten that goal. Thus they cease to be antinomies reflecting a bewildering collision of the basic principles of thought. They are instead a collision between rather crude principles by which we have been attempting to mechanize thought.

To support this assertion I will first look briefly at the history of our need for algorithmic constructions. I believe it has peaked and is being replaced by the need to understand thought and consciousness. Then I will look at the paradoxes diagnostically, emphasizing a feature of these paradoxes that reveals why understanding them does not produce a set of rules by which they can be avoided. My final point will be that the phenomenon of antinomy, by showing the weakness of the model of thought that we held, calls us to a different view, one not limited by what we could build, but mirroring what we are.

I. Conjoined with Universal Idiocy?

J.S. Mill made an effort to systematize inductive reasoning, but he added the proviso that no general method will avoid bad results "if conjoined with universal idiocy." In *The Many Faces of Realism*, Putnam quotes him at that point while showing the extreme difficulty of giving formal criteria for a strong argument from analogy. He then adds: "In Mill's day, this would not have seemed a noteworthy observation, but today, when we understand that a properly formalized algorithm *must* give the results it is supposed to if slavishly followed "even by a moron," as one says, we can see that Mill was admitting that he had not succeeded in *formalizing* inductive logic."[2] Since Mill's time, then, there has been a tightening of expectations about what a formalization of thought might do, we do expect a system to work mechanically, without any "judgment calls."

I do not suggest that this aspiration is new. Rationality has always been a search for the principles of things, for that which does not

change, and formalization is the spelling out of those principles. The categorical syllogism is a mechanism, and Leibniz could never have said, "let us calculate," without envisioning an elaborate mechanization of thought. Mill, then, was not reflecting an earlier position as much as he was making excuses for the much more difficult realm of inductive logic. But Putnam's words can help us to be aware of what we aspire to: "a properly formalized algorithm *must* give the results it is supposed to if slavishly followed "even by a moron." For "slavishly" and "moron" we could substitute "mechanically" and ""machine"; for that is our aspiration, to produce a machine that can think.

I do not mean that analytic philosophers are wanting to be programmers, but only that the technological agenda of modern science has profoundly affected our view of what is rational. That which cannot be reduced to rules is not rational. When the rules we have discerned produce the odd results seen in these antinomies, we are sure that we have not found the proper rules, or have not found the complex arrangement of them that will work. But these persistent problems are more likely the proof that thought cannot be contained in rules.

This century is a witness to failed dichotomies and evaporated foundations, and formalization does not fare well in such a climate.[3] Thus the literature on this problem changes. In Martin's first book on the Liar Paradox, the papers reflect technical industry, as if we had the tools and were just working out the details. But fourteen years later, in *Recent Essays on Truth and the Liar Paradox*,[4] the attitude seems different, at least in some of the papers.

Neither Saul Kripke nor Tyler Burge promises a rule-governed description of the conditions in which paradox will appear. Kripke's "Outline of a Theory of Truth"[5] continues with one of the established methods of solving the Liar Paradox, the "truth-value gap" approach.[6] But he makes it plain that "it would be fruitless to look for an *intrinsic criterion* that will enable us to sieve out—as meaningless, or ill-formed—those sentences that lead to paradox." And, "No syntactic or semantic feature of (1) guarantees that it is unparadoxical." It depends upon the empirical facts behind those particular sentences.[7] As he presents his won version of the theory, he writes, "The feature I have stressed about ordinary statements, that there is no intrinsic guarantee of

their safety (groundedness) and that their 'level' depends upon empirical facts, comes out clearly in this model."[8]

Tyler Burge's article, "Semantical Paradox," finds all the variants of the "truth-value gaps" approach futile in the end due to the "Strengthened Liar" paradox, which "is really the original Liar reiterated for the sake of those who seek to undercut paradox primarily by appeal to a distinction between falsehood and some other kind of truth failure."[9] He continues instead with a form of Tarski's hierarchical approach, ending up with subscripts on the word "true," which is used indexically. But his motivation is rather different from much of what has been done in this area. He outlines the intuitive steps when the paradox occurs, and then he writes, "The first task of an account of semantical paradox is to explicate the moves from (a) to (b) and from (b) to (c). Most recent accounts have either ignored such reasoning as the above or sought simply to block it by formal means. I think a more satisfying approach is to interpret the reasoning so as to *justify* it." In the end he writes, "Our aim is to dissolve the paradoxes by accounting for specific judgments by means of a theory of language that does not require us to make implausible claims about the linguistic or pragmatic properties of the discourse. . . ."[10]

Burge looks very carefully at how the healthy operation of reason leads to paradox in certain cases. For instance, he writes, "This account explains why there *seems* to be a change of truth value. . ."[11] and he uses his formal terms and subscripts to describe the process. But, again, this diagnostic theory does not supply a set of rules by which to anticipate and eliminate paradox. The effect of the formalization is to clarify the diagnosis, and the effect of the diagnosis is to increase our confidence in the healthy functioning of reason. No promises are made that a robot will be able to benefit from all of this, and Burge agrees with Kripke that the health of a set of statements must be empirically determined: "*Pathologicality is not an intrinsic condition but a disposition to produce disease for certain semantical evaluations*—evaluations that in a context may or may not be implicated as appropriate."[12]

Both these writers use formalization as a way to describe what has already occurred, but not to build a system that will automatically prevent such occurrences. Thus both follow Mill's observation, that no

system can be built that will not produce bad results if conjoined with universal idiocy.

When Mill used that expression, he was dealing with the difficulty of formalizing inductive reasoning. For example, the argument from analogy has a form that is meaningful and useful as a guide to one's deliberations, but that form alone never indicates whether the argument is strong or weak. That depends upon the content.[13] This contrasts with deductive logic, in which we believe we can judge on the form without knowing the content. Inductive reasoning cannot be formalized because form and content cannot be strictly separated.

It turns out, though, that form and content cannot be strictly separated in deduction, either. That is, one way to understand the paradoxes is to see them as special cases in which the content of a locution interferes with its form. In those cases the mechanism we call formal logic malfunctions.

II. Essential Reflexiveness

The diagnosis of the semantic paradoxes was nearly complete as soon as Russell observed that the problem was self-reference. These expressions are generally also negative—an *untrue* statement, a *non*-self-describing adjective, the class of classes *not* members of themselves—but this turns out not so important, since positive expressions of parallel form are also "pathological," to use Burge's term.[14] Self-reference was "treated" through the hierarchial structure of formal systems that would not allow it to occur, but this was too broad a solution to apply to natural language.

The phrase I have here, "essential reflexiveness," is a fine-tuning of the self-reference diagnosis. There are many innocent ways that a word or sentence can refer to itself, but the problem occurs when an operation turns upon itself its own essence. An expression brings to bear upon itself exactly the function that an expression of that kind has. That is, propositions claim to be true, so the problematic one is the one that says it is not true (or that it is true). An adjective describes, so Grelling's paradox occurs when we ask if an adjective describes itself.[15] A class includes members, and Russell's paradox occurs when the defining property of a class in non-self-inclusion (or self-inclusion).

These observations are not really new, but the consequence of this diagnosis is significant. It amounts to proof that no formal solution to the paradoxes is possible. Paradox occurs when a certain kind of operation is turned upon itself. But the syntax of a language system will never reveal when that is likely to occur, because it is the meanings of the words that determine when essential reflexiveness might occur. So no rules for its elimination could be devised. This is a fine-tuning of the self-reference diagnosis, but it cannot be put into a structural preventive measure. We will always need to know what the words mean before deciding if a problem exists. Thus we have again the point made by Kripke and Burge that it is an empirical matter whether an expression is paradoxical.

This diagnosis suggests that one could produce paradox at will by finding, for an expression of any form or function, a content that interferes with that form, as "self-descriptive" interferes with the function of an adjective. I believe this is possible in formal treatments, or at least the possibility can be expressed generally through a formalization. But in the real world of natural meanings it seems that there are a few main functions, like affirmation/negation in propositions or description in adjectives, that can be turned inward, and thus it is not the case that there are as many possibilities of paradox as there are meanings of words. On the other hand, there is always the possibility that the meaning of a word will have an effect on a structural form. The contents are supposed to just "ride along in their forms, but if the meanings connect with what the form is doing, the form and content distinction breaks down.

For example, the formal statement, $(x)(Rx > Ex)$, does not have a problem of syntax; but if R means "is a rule" and E means "has an exception," we end up saying "every rule has an exception." That does not put us into a revolving door of truth states as does the liar sentence, but it is necessarily false, because if every rule does have an exception, then this one does, and there is at least one rule that does not have an exception. The liar paradox does not appear because the self-negation of the universal statement yields a particular statement, and the matter rests there: the statement is necessarily false. Nevertheless, something odd has happened here, and it is because the universal statement, which is a structural form, has been affected by its content. R and E are predicates

about rules which interfere with the functioning of the structure that was built to express rules.[16]

The general point is that no form can be protected from its content, because the meanings that are built into the form will sometimes appear within the content. But if the form-content distinction cannot be preserved always, neither can the flawless functioning of logic be guaranteed.

III. A Model?

Diagnostic treatment of the paradoxes does not yield rules by which to prevent their occurrences, but shows why no such rules are possible. We can formalize our after-the-fact diagnostic analysis of a pathological expression, but the theory of semantic paradox (e.g. Burge's) only explains; it does not predict. This finding weakens our confidence in the construction of a rule-governed artificial language that can do much of what real language does. At the same time, the fact that we can understand what is happening increases our appreciation of reason itself.[17] We do not have nervous breakdowns over these paradoxes; our reason does not "lock up" over them. And since the self-referential nature of these expressions generally robs them of any actual content other than what they say about themselves, we never come near to a practical problem. Thus it seems fair to claim that the antinomies are not a crisis for thought itself, but for a certain project, the algorithmic modelling of thought.

Yet this has been a crisis of thought from within the perspective in which rationality and algorithm are equated. It seems to be moving us out of that perspective, because it is one of many indications that we cannot *construct* intelligence. One is inclined to look for a better model, but the very idea of a "model" may be prejudicial, favoring the view of intelligence as something we can build. The shift occurring now seems to be toward assuring ourselves that this reason of ours is quite healthy, too healthy to be contained. It is the *power* of reason that makes paradoxes unavoidable; to deprive it of that power would be like removing weapons from soldiers because of an occasional suicide. Sentences can negate other sentences, and thus they can negate themselves, just as classes can exclude themselves. Seeing the inevitability of this increases our respect for the phenomenon of language.

Though wary of a model for thought, I do think the great thought-experiment of the antinomies can point us toward a theory of consciousness, not a theory by which to predict and control anything, but an explanation. How does language operate in a brain to produce this amazing thing called consciousness, which is so remarkable that many philosophers and scientists do not even admit it really exists? Perhaps the features of thought that frustrate our attempts to formalize it are exactly what we need to focus on to get insight into its essential nature. After all, to understand consciousness we want to know how it differs from other natural processes, which we can at least remotely conceive of modelling mechanically. Or we want to know how it differs from a computer program, which we can build. If so, then the aspect of thought that eludes our constructive skills ought to be particularly significant.

Let me add just a shred of flesh to this speculative suggestion. A diagnostic tool that would fit well with my thoughts here is the use-mention distinction.[18] It was said, I believe, that Russell slipped on that distinction, and that suggests it is a rather slippery thing. When I applied the use-mention distinction in my own ponderings on this, I found that it had real explanatory power but an amazing elusiveness. I would examine and rewrite various expressions and find myself unable to say whether or not I had achieved what I sought, the elimination of paradox. Yet if all I had to do was manipulate a chain of truth-value reversals, as suggested by the structure of truth-functional logic, even if that were quite tedious, I would not be left unsure whether the alteration has solved the problem.

I concluded then—and I will conclude here—that conscious thought cannot be modelled in a linear manner, but involves us in a rich and complex layering of use and mention, with symbols interacting with each other and themselves, being used and mentioned in a facile process that defies all system. The historical role of the antinomies may be to turn us away from the mechanization of thought and toward the respectful exploration of its power.

NOTES

1. W.V. Quine, "The Ways of Paradox," in *The Ways of Paradox and Other Essays* (New York: Random House, 1966). Reprinted in Perry and Bratman, *Introduction to Philosophy*, 807, 811.

2. Hilary Putnam, *The Many Faces of Realism (La Salle, Illinois: Open Court, 1987), 73-74.*

3. One result of the trend is irrationalism, even in philosophy. But that is not a necessary direction; there are still many ways to make good sense without believing that thought can be reduced to rules.

4. Robert L. Martin, ed., *The Paradox of the Liar* (New Haven, London: Yale University Press, 1970) and *Recent Essays on Truth and the Liar Paradox* (New York: Oxford University Press, 1984).

5. Saul Kripke, "Outline of a Theory of Truth," in Martin, *Recent Essays on Truth and the Liar Paradox*, first published in *The Journal of Philosophy* 72, 1975.

6. Some sentences in certain contexts lack a truth value, that is, are meaningless. This solution amounts to a rejection of bivalency in truth states. The problem is to say exactly when a sentence lacks truth value.

7. Ibid., 55. His "(1)" was "Most of Nixon's assertions about Watergate are false," which by itself seems innocent. But it is easy to invent another sentence and some circumstances that create the Liar Paradox.

8. Ibid., 71-72.

9. Tyler Burge, "Semantical Paradox" in Robert Martin, ed., *Recent Essays on Truth and the Liar Paradox*, 87. This was first published in *Journal of Philosophy* 76, 1979. The original strengthened liar was simply, "1. 1 is not true" (instead of "false"). Burge's point is that much more complex formulations end up in the same situation; it is the "Persistent Liar."

10. Ibid., 93, 114.

11. Ibid., 99.

12. Ibid., 101.

13. x is like y, x is A, therefore y is A. But then must we judge the relevant similarities, etc., of x and y, and this depends upon what x and y and A mean.

14. "This statement is true" is not paradoxical, but empty; it has no content other than what it says about itself. It could be called parasitic upon itself, even while not paradoxical. The class of classes that are members of themselves is problematic, too, and yet this is very difficult to accept intuitively, since we know of classes that are members of themselves. The emptiness of the positive form of Russell's paradox made sense to me when I realized that self-membership could be an accidental property of a class but not the defining property of one. Just as non-self-membership is a property by which a class excludes itself from itself, self-membership is a property that defines itself circularly, and thus cannot be itself define anything at all. Like the truth-teller sentence,, it is parasitic upon itself.

15. "Heterological" does not describe itself, but—oops—now it does! The positive form is a problem, too, just like the class of self-membership; "autological" is, by itself, meaningless. It can be an accidental property of an adjective, like "short," but as a defining property it sucks up its own meaning in circularity.

16. See Benson Mates, *Skeptical Essays* (Chicago: University of Chicago Press, 19810 17. He points out the New Testament reference (Titus 1.12,13) to the Cretan prophet who said, truly, that Cretans are always liars, produces what is known as "the Epimenides," but that is not in fact an example of the liar paradox; it is necessarily false, but not paradoxical. He adds, "We do not have an antinomy here, for no contradiction has been derived, but we do have a proof of the curious and implausible result that, if Epimenides happened to make the false statement attributed to him, then some Cretan must have said something true." This parallels the case above: saying "every rule has an exception" necessitates that there is at least one rule without an exception. No paradox occurs because the negated universal statement yields a particular one; but to have produced a necessarily true *essential* statement by just saying something is perhaps more interesting than paradox itself.

17. There has not been space here for examples, but in thinking through the way these paradoxes actually occur, one can experience an intuitive release from their paradoxical sound; they can be disarmed. Sections II through IV in Burge's article have that effect.

18. In Tarski's truth rule, "'snow is white' is true if snow is white," the words are first mentioned and then used, and it is clear which is which, but in "this sentence is true" use and mention occur together.