

The Riddle of the Universe and Its Solution

We have prepared this report to provide fuller information in connection with the President's recent press conference on the so-called "Riddle." We hope the report helps to dispel the ugly mood apparent throughout the country, bordering on panic, which has most recently found expression in irresponsible demands to close the universities. Our report has been prepared in haste; in addition, our work was tragically disrupted, as described later.

We first review the less well known early history of the Riddle. The earliest known case is that of C. Dizzard, a research fellow with the Autotomy Group at M.I.U. Dizzard had previously worked for several small firms specializing in the development of artificial intelligence software for commercial applications. Dizzard's current project involved the use of computers in theorem proving, on the model of the proof in the 1970s of the four-color theorem. The state of Dizzard's project is known only from a year-old progress report; however, these are often intended at most for external use. We shall not discuss the area of Dizzard's work further. The reason for our reticence will be apparent shortly.

Dizzard last spoke one morning before an Easter weekend, while waiting for a routine main computer system failure to be fixed. Colleagues saw Dizzard at the terminal in his office at about midnight that day; late-night work habits are common among computer users, and Dizzard was known to sleep in his office. On the next afternoon, a co-worker noticed Dizzard sitting at his terminal. He spoke to Dizzard, but Dizzard did not reply, not an unusual occurrence. On the morning after the vacation, another colleague found Dizzard sitting with his eyes open before his terminal, which was on. Dizzard seemed awake but did not respond to questions. Later that day, the colleague became concerned by Dizzard's unresponsiveness and tried to arouse him from what he thought was a daydream or daze. When these attempts were unsuccessful, Dizzard was taken to a hospital emergency room.

Dizzard showed symptoms of a total food-and-water fast of a week (aggravated by marginal malnutrition caused by a vending-machine diet); he was in critical condition from dehydration. The inference was that Dizzard had not moved for several days, and that the cause of his immobility was a coma or trance. The original conjecture was that a stroke or tumor caused Dizzard's paralysis. However, electroencephalograms indicated only deep coma. (According to Dizzard's health records, he had been institutionalized briefly ten years ago, not an uncommon incident in certain fields.) Dizzard died, apparently of his fast, two days later. Autopsy was delayed by objections of next of kin, members of a breakaway sect of the neo-Jemimakins cult. Histological examination of Dizzard's brain so far has revealed no damage whatever; these investigations continue at the National Center for Disease Control.

The director of the Autotomy Group appointed one of Dizzard's graduate students to manage his project while a decision was made about its future. The floor of Dizzard's office was about one foot deep in papers and books; the student was busy for a month just sorting the materials into some general scheme. Shortly afterward, the student reported at a staff meeting that she had begun work on Dizzard's last project and that she had found little of particular interest. A week later she was found sitting at the terminal in Dizzard's office in an apparent daze.

There was confusion at first, because she was thought to be making a poor joke. She was staring straight ahead, breathing normally. She did not respond to questions or being shaken, and showed no startle response to loud noises. After she was accidentally bumped from her chair, she was hospitalized. The examining neurologist was unaware of Dizzard's case. He reported the patient was in apparently good physical condition, except for a previously undiagnosed pineal gland abnormality. After Autotomy Project staff answered inquiries by the student's friends,

her parents informed the attending physician of Dizzard's case. The neurologist noted the difficulty of comparing the two cases, but suggested the similarities of deep coma with no detectable brain damage; the student's symptoms constituted no identifiable syndrome.

After further consultation, the neurologist proposed the illness might be caused by a slow acting sleeping-sickness-like pathogen, caught from Dizzard's belongings—perhaps hitherto unknown, like Legionnaire's Disease. Two weeks later, Dizzard's and his student's offices were quarantined. After two months with no further cases and cultures yielding only false alarms, quarantine was lifted.

When it was discovered janitors had thrown out some of Dizzard's records, a research fellow and two more of Dizzard's students decided to review his project files. On their third day, the students noticed that the research fellow had fallen into an unresponsive trancelike state and did not respond even to pinching. After the students failed to awaken the research fellow, they called an ambulance. The new patient showed the same symptoms as the previous case. Five days later, the city public health board imposed a quarantine on all building areas involved in Dizzard's project.

The following morning, all members of the Autotomy Group refused to enter the research building. Later that day, occupants of the rest of the Autotomy Group's floor, and then all 500 other workers in the building, discovered the Autotomy Project's problems and left the building. The next day, the local newspaper published a story with the headline "Computer Plague." In an interview, a leading dermatologist proposed that a virus or bacterium like computer lice had evolved that metabolized newly developed materials associated with computers, probably silicon. Others conjectured that the Autotomy Project's large computers might be emitting some peculiar radiation. The director of the Autotomy Group was quoted: The illnesses were a public health matter, not a concern of cognitive scientists.

The town mayor then charged that a secret Army project involving recombinant DNA was in progress at the building and had caused the outbreak. Truthful denials of the mayor's claim were met with understandable mistrust. The city council demanded immediate quarantine of the entire ten-story building and surrounding area. The university administration felt this would be an impediment to progress, but the local Congressional delegation's pressure accomplished this a week later. Since building maintenance and security personnel would no longer even approach the area, special police were needed to stop petty vandalism by juveniles. A Disease Control Center team began toxicological assays, protected by biohazard suits whenever they entered the quarantine zone.

In the course of a month they found nothing, and none of them fell ill. At the time some suggested that, because no organic disease had been discovered in the three victims, and the two survivors showed some physiological signs associated with deep meditation states, the cases might be an outbreak of mass hysteria.

Meanwhile, the Autotomy Group moved into a “temporary” World War II-era wooden building. While loss of more than ten million dollars in computers was grave, the group recognized that the information, not the physical artifacts that embodied it, was indispensable. They devised a plan: biohazard-suited workers fed “hot” tapes to readers in the quarantine zone; the information was transmitted by telephone link from the zone to the new Autotomy Project site and recorded again. While transcription of the tapes allowed the project to survive, only the most important materials could be so reconstructed. Dizzard’s project was not in the priority class; however, we suspect an accident occurred.

A team of programmers was playing back new tapes, checking them on monitors, and provisionally indexing and filing their contents. A new programmer encountered unfamiliar material and asked a passing project supervisor whether it should be discarded. The programmer later reported the supervisor typed in commands to display the file on the monitor; as the programmer and the supervisor watched the lines advance across the screen, the supervisor remarked that the material did not look important. Prudence prevents our quoting his comments further. He then stopped speaking in midsentence. The programmer looked up; he found the supervisor staring ahead. The supervisor did not respond to questions. When the programmer pushed back his chair to run, it bumped the supervisor and he fell to the floor. He was hospitalized with the same symptoms as the earlier cases.

The epidemiology team, and many others, now proposed that the cause of illness in the four cases might not be a physical agent such as a virus or toxin, but rather an abstract piece of information—which could be stored on tape, transmitted over a telephone line, displayed on a screen, and so forth. This supposed information now became known as “the Riddle,” and the illness as “the Riddle coma.” All evidence was consistent with the once-bizarre hypothesis that any human who encountered this information lapsed into an apparently irreversible coma. Some also recognized that the question of exactly what this information is was extremely delicate.

This became clear when the programmer involved in the fourth case was interviewed. The programmer’s survival suggested the Riddle must be understood to induce coma. He reported he had read at least some lines on the monitor at the time the supervisor was stricken. However, he

knew nothing about Dizzard's project, and he was able to recall little about the display. A proposal that the programmer be hypnotized to improve his recall was shelved. The programmer agreed it would be best if he did not try to remember any more of what he had read, although of course it would be difficult to try *not* to remember something. Indeed, the programmer eventually was advised to abandon his career and learn as little more computer science as possible. Thus the ethical issue emerged of whether even legally responsible volunteers should be permitted to see the Riddle.

The outbreak of a Riddle coma epidemic in connection with a computer-assisted theorem-proving project could be explained; if someone discovered the Riddle in his head, he should lapse into coma before he could communicate it to anyone. The question arose of whether the Riddle had in fact been discovered earlier by hand and then immediately lost. A literature search would have been of limited value, so a biographical survey was undertaken of logicians, philosophers, and mathematicians working since the rise of modern logic. It has been hampered by precautions to protect the researchers from exposure to the Riddle. At present, at least ten suspect cases have been discovered, the earliest almost 100 years ago.

Psycholinguists began a project to determine whether Riddle coma susceptibility was species-specific to humans. "Wittgenstein," a chimpanzee trained in sign language who had solved first-year college logic puzzles, was the most appropriate subject to see the Autotomy Project tapes. The Wittgenstein Project investigators refused to cooperate, on ethical grounds, and kidnapped and hid the chimpanzee; the FBI eventually found him. He was shown Autotomy tapes twenty-four hours a day, with no effect whatever. There have been similar results for dogs and pigeons. Nor has any computer ever been damaged by the Riddle.

In all studies, it has been necessary to show the complete Autotomy tapes. No safe strategy has been found for determining even which portion of the tapes contains the Riddle. During the Wittgenstein-Autotomy Project, a worker in an unrelated program seems to have been stricken with Riddle coma when some Autotomy tapes were printed out accidentally at a public user area of the computer facility; a month's printouts had to be retrieved and destroyed.

Attention focused on the question of what the Riddle coma is. Since it resembled no known disease, it was unclear whether it was really a coma or indeed something to be avoided. Investigators simply assumed it was a virtual lobotomy, a kind of gridlock of the information in the synapses, completely shutting down higher brain functions. Nonetheless, it was unlikely the coma could be the correlate of a state of meditative enlight-

enment, because it seemed too deep to be consistent with consciousness. In addition, no known case of Riddle coma has ever shown improvement. Neurosurgery, drugs, and electrical stimulation have had, if any, only negative effects; these attempts have been stopped. The provisional verdict is that the coma is irreversible, although a project has been funded to seek a word to undo the “spell” of the Riddle, by exposing victims to computer-generated symbol strings.

The central question, “What is the Riddle?” obviously has to be approached very cautiously. The Riddle is sometimes described as “the Gödel sentence for the human Turing machine,” which causes the mind to jam; traditional doctrines of the unsayable and unthinkable are cited. Similar ideas are familiar in folklore—for instance, the religious theme of the power of the “Word” to mend the shattered spirit. But the Riddle could be of great benefit to the cognitive sciences. It might yield fundamental information about the structure of the human mind; it may be a Rosetta Stone for decoding the “language of thought,” universal among all humans, whatever language they speak. If the computational theory of mind is at all correct, there is some program, some huge word, that can be written into a machine, transforming the machine into a thinking thing; why shouldn’t there be a terrible word, the Riddle, that would negate the first one? But all depended on the feasibility of a field of “Riddle-ology” that would not self-destruct.

At this point, an even more disturbing fact about the Riddle began to emerge. A topologist in Paris lapsed into a coma similar in some respects to Dizzard’s. No computer was involved in this case. The mathematician’s papers were impounded by the French, but we believe that, although this mathematician was not familiar with Dizzard’s work, she had become interested in similar areas of artificial intelligence. About then four members of the Institute for Machine Computation in Moscow stopped appearing at international conferences and, it seems, personally answering correspondence; FBI officials claimed the Soviet Union had, through routine espionage, obtained the Autotomy tapes. The Defense Department began exploring the concept of “Riddle warfare.”

Two more cases followed, a theoretical linguist and a philosopher, both in California but apparently working independently. Neither was working in Dizzard’s area, but both were familiar with formal methods developed by Dizzard and published in a well-known text ten years ago. A still more ominous case appeared, of a biochemist working on information-theoretic models of DNA-RNA interactions. (The possibility of a false alarm remained, as after entering coma the biochemist clucked continuously, like a chicken.)

The Riddle coma could no longer safely be assumed an occupational

hazard of Dizzard's specialty alone; it seemed to lurk in many forms. The Riddle and its effect seemed not just language-independent. The Riddle, or cognates of it, might be topic-independent and virtually ubiquitous. Boundaries for an intellectual quarantine could not be fixed confidently.

But now we are finding, in addition, that the Riddle seems an idea whose time has come—like the many self-referential paradoxes (of the pattern “This sentence is false”) discovered in the early part of this century. Perhaps this is reflected in the current attitude that “computer science is the new liberal art.” Once the intellectual background has evolved, widespread discovery of the Riddle appears inevitable. This first became clear last winter when most of the undergraduates in a large new introductory course on automata theory lapsed into coma during a lecture. (Some who did not nevertheless succumbed a few hours later; typically, their last word was “aha.”) When similar incidents followed elsewhere, public outcry led to the president's press conference and this report.

While the present logophobic atmosphere and cries of “Close the universities” are unreasonable, the Riddle coma pandemic cannot be viewed as just another example of runaway technology. The recent “Sonic Oven” case in Minneapolis, for instance, in which a building with a facade of parabolic shape concentrated the noise of nearby jets during takeoff, actually killed only the few people who happened to walk through the parabola's focus at the wrong time. But even if the Riddle coma were a desirable state for an individual (which, we have seen, it does not seem to be), the current pandemic has become an unprecedented public health crisis; significant populations are unable to care for themselves. We can only expect the portion of our research community—an essential element of society—that is so incapacitated to grow, as the idea of the Riddle spreads.

The principal objective of our report was at least to decrease further coma outbreaks. Public demand for a role in setting research policy has emphasized the dilemma we confront: how can we warn against the Riddle, or even discuss it, without spreading its infection? The more specific the warning, the greater the danger. The reader might accidentally reach the stage at which he sees “If p then q ” and p , and so cannot stop himself from concluding q , where q is the Riddle. Identifying the hazardous areas would be like the children's joke “I'll give you a dollar if you're not thinking of pink rats ten seconds from now.”

A question of ethics as well as of policy remains; is the devastating risk of the Riddle outweighed by the benefits of continued research in an ill-defined but crucial set of fields? In particular, the authors of this report have been unable to resolve the issue of whether the possible benefit of

any report itself can outweigh its danger to the reader. Indeed, during preparation of our final draft one of us tragically succumbed.

Reflections

This curious story is predicated upon a rather outlandish yet intriguing idea: a mind-arresting proposition, one that throws any mind into a sort of paradoxical trance, perhaps even the ultimate Zen state of satori. It is reminiscent of a Monty Python skit about a joke so funny that anyone who hears it will literally die laughing. This joke becomes the ultimate secret weapon of the British military, and no one is permitted to know more than one word of it. (People who learn two words laugh so hard they require hospitalization!)

This kind of thing has historical precedents, of course, both in life and in literature. There have been mass manias for puzzles, there have been dancing fits, and so on. Arthur C. Clarke wrote a short story about a tune so catchy that it seizes control of the mind of anyone who hears it. In mythology, sirens and other bewitching females can completely fascinate males and thus overpower them. But what is the nature of such mythical mind-gripping powers?

Cherniak's description of the Riddle as "the Gödel sentence for the human Turing machine" may seem cryptic. It is partly explicated later by his likening it to the self-referential paradox "This sentence is false"; here, a tight closed loop is formed when you attempt to decide whether it is indeed true or false, since truth implies falsity, and vice versa. The nature of this loop is an important part of its fascination. A look at a few variations on this theme will help to reveal a shared central mechanism underlying their paradoxical, perhaps mind-trapping, effect.

One variant is: "This sentence contains three errors." On reading it, one's first reaction is, "No, no—it contains two errors. Whoever wrote the sentence can't count." At this point, some readers simply walk away scratching their heads and wondering why anyone would write such a pointless, false remark. Other readers make a connection between the sentence's apparent falsity and its message. They think to themselves, "Oh, it made a third error, after all—namely, in counting its own errors." A second or two later, these readers do a double-take, when they realize that if you look at it that way, it seems to have *correctly* counted its errors,

and is thus *not* false, hence contains only *two* errors, and . . . “But . . . wait a minute. Hey! Hmm . . .” The mind flips back and forth a few times and savors the bizarre sensation of a sentence undermining itself by means of an interlevel contradiction—yet before long it tires of the confusion and jumps out of the loop into contemplation, possibly on the purpose or interest of the idea, possibly on the cause or resolution of the paradox, possibly simply to another topic entirely.

A trickier variant is “This sentence contains one error.” Of course it is in error, since it contains no errors. That is, it contains no *spelling* errors (“first-order errors”). Needless to say, there are such things as “second-order errors”—errors in the counting of first-order errors. So the sentence has no first-order errors and one second-order error. Had it talked about how many first-order errors it had, or how many second-order errors it had, that would be one thing—but it makes no such fine distinctions. The levels are mixed indiscriminately. In trying to act as its own objective observer, the sentence gets hopelessly muddled in a tangle of logical spaghetti.

C. H. Whitely invented a curious and more mentalistic version of the fundamental paradox, explicitly bringing in the system thinking about itself. His sentence was a barb directed at J. R. Lucas, a philosopher one of whose aims in life is to show that Gödel’s work is actually the most ineradicable uprooting of mechanism ever discovered—a philosophy, incidentally, that Gödel himself may have believed. Whitely’s sentence is this:

Lucas cannot consistently assert this sentence.

Is it true? Could Lucas assert it? If he did, that very act would undermine his consistency (nobody can say “I can’t say this” and remain consistent). So Lucas cannot consistently assert it—which is its claim, and so the sentence is true. Even Lucas can see it is true, but he can’t assert it. It must be frustrating for poor Lucas! None of us share his problem, of course! Worse yet, consider:

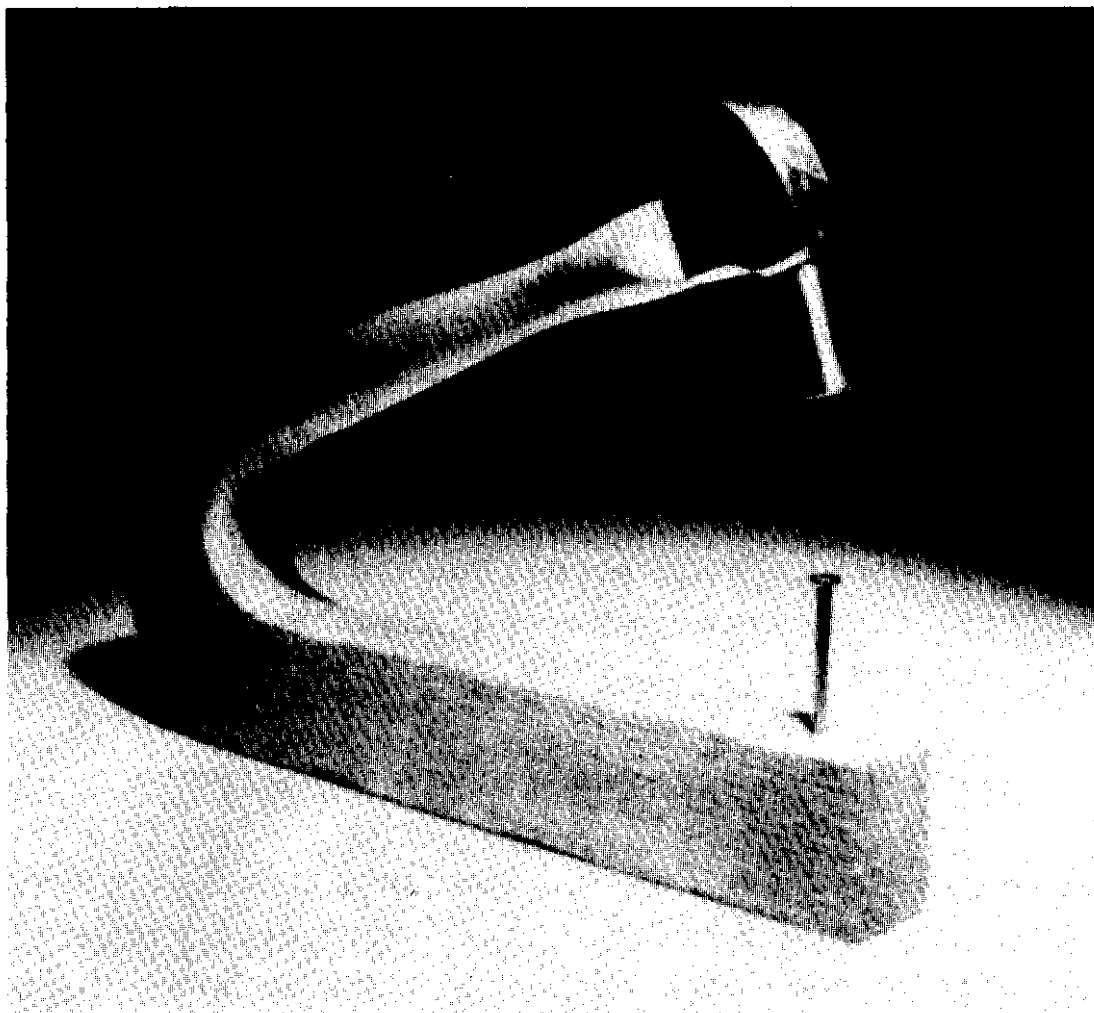
Lucas cannot consistently believe this sentence.

For the same reasons, it is true—but now Lucas cannot even believe it, let alone assert it, without becoming a self-contradictory belief system.

To be sure, no one would seriously maintain (we hope!) that people are even remotely close to being internally consistent systems, but if this kind of sentence is formalized in mathematical garb (which can be done), so that Lucas is replaced by a well-defined “believing system” L , then

there arises serious trouble for the system, if it wishes to remain consistent. The formalized Whitely sentence for *L* is an example of a true statement that the system itself could never believe! Any *other* believing system is immune to this particular sentence; but on the other hand there is a formalized Whitely sentence for *that* system as well. Every “believing system” has its own tailor-made Whitely sentence—its Achilles’ heel.

These paradoxes all are consequences of a formalization of an observation as old as humanity: an object bears a very special and unique relationship to itself, which limits its ability to act upon itself in the way it can act on all other objects. A pencil cannot write on itself; a fly swatter cannot swat a fly sitting on its handle (this observation made by the German philosopher-scientist Georg Lichtenberg); a snake cannot eat itself; and so on. People cannot see their own faces except via external aids that present images—and an image is never quite the same as the original thing. We can come close to seeing and understanding ourselves objectively, but each of us is trapped inside a powerful system with a unique point of view—and that power is also a guarantor of limitedness. And this vulnerability—this self-hook—may also be the source of the ineradicable sense of “I.”



Malcolm Fowler’s hammer nailing itself is a new version of the ouroboros. (From *Vicious Circles and Infinity: An Anthology of Paradoxes* by Patrick Hughes and George Brecht.)