Letters to the Editor

The Mathematical Intelligencer *encourages comments about the material in this issue. Letters to the editor should be sent to the editor-in-chief, Chandler Davis.*

Prediction and the Spiteful Computer[1]

A deterministic system consists of two computers, Laplace and Baby Dostoevsky. Laplace is programmed to say at time T_1 what Baby Dostoevsky will do at some later time T_2. Baby Dostoevsky is programmed to do at time T_2 the opposite of what Laplace has said at time T_1. Baby Dostoevsky's method is the obvious one. Laplace's method is to calculate the state of the system at time T_2 given the initial state; this should be possible since the system is deterministic.

I suggest resolving this paradox as follows. Laplace's program includes a description of the initial state of the system. On the other hand, Laplace's program is part of the initial state of the system. Therefore, Laplace's program has to include a description of itself. There is no reason to suppose that the constraints this requirement imposes are consistent, and this resolves the paradox.

Going further, one can say that, because the supposed initial state (or program) leads to a contradiction, in fact there is no such initial state.

This is more or less the same as some of the solutions suggested in Akin's article, but it is perhaps expressed in a more mathematical and less physical way.

Richard Steiner
Department of Mathematics
University of Glasgow
University Gardens
Glasgow, G12 8QW
Scotland

This is not a physics paradox. The physical assumptions, Newtonian determinism, and uniform continuity of phase flow, as well as the requirement that predictions be secured through detailed microphysical computation, are all unnecessary scaffolding. The crux of the paradox is that the megacomputer L. is allegedly unable to make a certain prediction, which from other considerations, it obviously should be able to make. The draconian computational protocol, coupled with an assumption of determinism, is presumably intended to secure this predictability. This is computational overkill, as can be seen from the fact that the predictions L. has to make are trivially easy. L. needs to predict its own output and Baby D.'s kneejerk response.

The catch-up problem is irrelevant to the paradox's resolution. This problem arises from the profligate stipulation that predictions are to be secured through a detailed calculation based on an exact microphysical theory. We can communicate the full force of the paradox without this extra baggage, in fact, without significant physical assumptions: Call the realistic computer in this version, R. R. sticks to the essentials; it predicts only its own output and Baby D.'s inevitable negation. This is quite easy, so R. can be a small device that is not afflicted with a catch-up problem. Both are incapable of making the prediction in the required form, owing to Baby D.'s simple-minded spoiler tactics.

In short, R.'s version communicates the essential paradox, and is not tied to any particular physical theory; any possible world with enough stability for the construction of simple machines would suffice.

Not a physics paradox, this is a logical paradox, furthermore a *semantic* one because it springs from too permissive a stance on the issue of when to permit one thing to be "about" another thing. (The clearest example is Epimenides's paradox: "This sentence is false.") In Akin's paradox, L.'s output is a prediction *about* Baby D.'s response. Under the terms of the problem, specification of Baby D.'s response is tantamount to specification of L.'s prediction. Thus, L.'s prediction makes an indirect statement about itself. This dooms the prediction to be false under the specified semantic assignment. It is a lesson of modern logic that whereas rigorous use of self-application can be a wellspring, unbridled use is sure to generate paradox and self-contradiction.

L. can make the needed prediction and, say, store it in memory or relay it through a channel that Baby D. does not monitor. There is no failure of predictability, and, hence no conflict with determinism. But L. has a semantic difficulty; it cannot, without falling into error, have a certain one of its outputs *mean* (or represent or be about) its prediction.

L. is not precluded from making a valid prediction; it is precluded from expressing its prediction in a certain

[1] See *Mathematical Intelligencer*. vol. 14, no. 2, 45-47

manner. If L. were a human whose predictive utterances were monitored by Baby D., we might say that L. can know the truth (on the matter of Baby D.'s response), but cannot speak it.

William Eckhardt
250 South Wacker Drive #650
Chicago, IL 60606 USA

I think the self-reference in the input, in spite of being an obstacle, is not a valid objection, as J. von Neumann showed when he gave descriptions of self-reproductive automata. My opinion is that the solution is in the catch-up problem. For every $T \geq 0$, let $f(T)$ be the time needed by Laplace to predict the output of Baby Dostoevsky at time $T + 1$. The paradox of the spiteful computer is just a simple proof of $f(T) > T$ for every $T \geq 0$. It looks like a diagonalization argument against the existence of a "universal future predictor." This seems convincing to me.

Miguel A. Lerma
Facultad de Informática
Universidad Politécnica de Madrid
28660 Boadilla del Monte
Madrid
Spain

From my own "classical physicist's" point of view, *storing* numerical data incurs a cost which increases, logarithmically, with desired accuracy. (Storing D digits of information requires space, time, and expense of order D.) With chaotic dynamics, this cost increases further, logarithmically with time. A classical computer can neither contain an accurate description of its state nor predict its own future. This mechanistic picture avoids the paradox of Ethan Akin's *spiteful computer*.

Feedback differs from *prediction* in *influencing* the future rather than foretelling it. Thus, there is no predictive paradox in Lee Lorenz's wonderful portrayal of "Self-Awareness," from the 25 May 1992 *New Yorker*. Predicting the future is impossibly hard, while influencing it is easy. Useful to keep in mind in an election year!

William G. Hoover
Department of Applied Science
University of California at Davis
P.O. Box 808, L-794
Livermore, CA 94550 USA

Akin Replies

These letters confirm my experiences discussing this puzzle. Everyone says that the problem is simple, but the proposed answers display considerable variety.

Certainly, this is a conceptual puzzle and not a physics paradox. Contra Lerma, I have reluctantly concluded that the catch-up problem is not the answer. This does not mean that for a computer to predict its own output is the trivial task that Eckhardt suggests. Such self-prediction is the heart of the paradox. The separation of the system into Laplace and Dostoevsky is just a convenient portrayal.

My residual fondness for the catch-up problem comes from its suggestion that relative size is the binding constraint against successful prediction. Such a result would free me from the *Walden Two* nightmare: My fear that something of roughly my size and complexity, for example, B. F. Skinner, could predict, and so

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (Required by 39 U.S.C. 3685). (1) Title of publication: The Mathematical Intelligencer. A. Publication No.: 001-656. (2) Date of filing: 10/1/92. (3) Frequency of issue: Quarterly. A. No. of issues published annually, 4. B. Annual subscription price, $27.00. (4) Location of known office of publication: 175 Fifth Avenue, New York, NY 10010. (5) Location of the headquarters of general business offices of the publishers: 175 Fifth Avenue, New York, NY 10010. (6) Names and addresses of publisher, editor, and managing editor: Publisher: Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010. Editor: Chandler Davis, Department of Mathematics, University of Toronto, Toronto, Ontario, M5S 1A1 Canada. Managing Editor: Springer-Verlag New York Inc., 175 Fifth Avenue, New York, NY 10010. (7) Owner: Springer Export GmbH, Tiergartenstrasse 17, W-6900 Heidelberg, Federal Republic of Germany; and Springer-Verlag Berlin, Heidelberger Platz 3, W-1000 Berlin 33, Federal Republic of Germany. (8) Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total of bonds, mortgages or other securities: Dr. Konrad Springer, Heidelberger Platz 3, W-1000 Berlin 33, Federal Republic of Germany. (9) The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: has not changed during preceding 12 months. (10) Extent and nature of circulation. A. Total no. copies printed (net press run): Average no. copies each issue during the preceding 12 months, 6587; no. copies single issue nearest filing date, 6476. B. Paid circulation: 1. Sales through dealers and carriers, street vendors, and counter sales: Average no. copies each issue during preceding 12 months, 226; no. copies single issue nearest to filing date, 226. 2. Mail subscriptions: average no. copies each issue during preceding 12 months, 5276; no. copies single issue nearest to filing date, 5315. C. Total paid circulation: average no. copies each issue during preceding 12 months, 5502; no. copies single issue nearest to filing date, 5541. D. Free distribution by mail, carrier, or other means. Samples, complimentary, and other free copies: average no. copies each issue during preceding 12 months, 552; no. copies single issue nearest to filing date, 552. E. Total distribution: average no. copies each issue during the preceding 12 months, 6054; no. copies single issue nearest to filing date, 6093. F. Copies not distributed: 1. Office use, left-over, unaccounted, spoiled after printing: average no. copies each issue during the preceding 12 months, 512; no. copies single issue nearest to filing date, 383. 2. Return from news agents: average number of copies single issue during preceding 12 months, 21; no. of copies single issue nearest to filing date, 0. G. Total: average no. copies each issue during preceding 12 months, 6587; no. copies single issue nearest to filing date, 6476. I certify that the statements made by me above are correct and complete.

Craig Van Dyck

Craig Van Dyck
Vice-President, Production

SELF-ACTUALIZER

Drawing by Lorenz; © 1992
The New Yorker Magazine, Inc.

control, my behavior. I am less bothered by predictability by something vastly greater than myself, for example, an angel.

Notice that as long as Skinner does not inform me of his predictions, his control of me does not appear to raise any more logical contradiction than does his training of any other pigeon. He merely adjusts, unknown to me, parameters whose effects on my behavior he can, by assumption, predict. The catch-up result, or Hoover's storage-size variant, would suggest that the paradox reveals a limitation upon Skinner which would deny the possibility of even such nonparadoxical control.

The trouble is that the paradox can be reconstructed with gadgets which clearly do admit a kind of self-description. Although I originally described it using a finite array of particles, the puzzle remains in force even if the computers are infinite. For infinite computers certain kinds of self-description and even self-prediction are possible.

Let $\{C_i: i = 0, 1, \ldots\}$ be a sequence of finite computers increasing in size so that computer C_i can predict by time T (fixed throughout) the results of any $T + 1$ computation by any hookup of the earlier C_j's. The infinite computer is the union of the C_i's with each receiving inputs only from the programmer and the previous ones in line. Give C_0 a problem, C_1 the problem of predicting C_0, C_2 the problem of predicting C_1, etc. After completing its task, each component just keeps printing the same output. At time T the output is the sequence:

$$(\text{working}, C_0 \text{ says Ans}, C_1 \text{ says } C_0 \text{ says Ans}, \ldots),$$

whereas at time $T + 1$ the output is

$$(\text{Ans}, C_0 \text{ says Ans}, C_1 \text{ says } C_0 \text{ says Ans}, \ldots).$$

Thus, the subsystem, $\{C_1, C_2, \ldots\}$, does predict the outcome of the entire system but only provided the feedback necessary to exploit the prediction does not exist.

We are left with the issue of self-reference which, I think, holds the key. However, I disagree with Eckhardt's semantic analysis. Questions about the meaning and reference of such terms as "prediction" and "about" have to do with our interpretation, from the outside, of part of the wiring diagram of the system. Such metalanguage is not required by the computers themselves. I think Steiner has it right.

My Math Department colleague Stanley Ocken agrees, although phrasing it differently. He suggests that the problem is not well-posed in that my ideal-gas particles fog over the issue of setting the whole system up. He challenges me to state the paradox in terms of finite-state machines. I do not see how to do so but that may not be significant. My imaginative facility with computers collapses long before it is hamstrung by logic.

Ocken also suggests an alternative route of escape from my Skinnerian nightmare. Complexity theory implies that for many problems, like iteration of a function, methods which exploit size superiority, like parallel processing, cannot be used to compress the number of steps which must be performed in sequence to obtain a solution. (So to build my infinite computer above, we require a sequence of machines of increasing speed rather than size. No problem. Since we are using an infinite number of components anyway, there is no reason to feel bound by the speed of light, either. But back to our universe.) "That means," I told him, "that not only can Skinner not predict me but angels can't either. Of course, God still can because he is exempt from all these rules." "That's right," was his reply. When I looked startled at his certainty, Stanley, who is Orthodox, smiled and added, "I have other sources of information."

Postscript: The article exhibited a sample of unpredictability, human or computer. The cartoon illustrating the relationship between Laplace and Dostoevsky was misattributed. It is the work of Samuel Vaughan of Berkeley, California.

Ethan Akin
Department of Mathematics
The City College
137 Street and Convent Avenue
New York, NY 10031 USA