The Simon Newcomb Awards

Pat Hayes and Ken Ford

Simon Newcomb (1835-1909) was a celebrated astronomer and thinker of his day who argued passionately, in articles published shortly before the Wright brothers took off, that manned flight was impossible. He combined a solid confidence in his own reasoning with a disdain for practical experiments. In many ways his arguments are similar to recent attacks on AI. They are short, elegant, convincing to his contemporaries, utterly wrong, and wonderfully silly, displaying an appealing mixture of partial insight with a failure to really comprehend what he was talking about. For example, there was the Stopping Problem argument. "Imagine the proud possessor of the aeroplane," suggested Newcomb sarcastically, "darting through the air at a speed of several hundred feet per second! It is the speed alone that sustains him. How is he ever going to stop?" (Newcomb, 1901). Newcomb intended his question rhetorically, but as everyone now knows, it has a perfectly good answer: "Very carefully."

The Simon Newcomb Award will be given in recognition of a similarly silly published argument against AI, especially when the writer's confidence in his views seems to arise from his ignorance of the subject. The ideal candidate is an eminent scientist or scholar in some other field — for example, a philosopher, sociologist or mathematician — who clearly fails to grok some basic idea of computer science. While any published argument may be nominated for the prize, the committee gives highest credit to arguments which are not just idiotic, but which use some technical issue in a way that displays some, but not enough, insight. Some argument forms are already judged unacceptable, including the following:
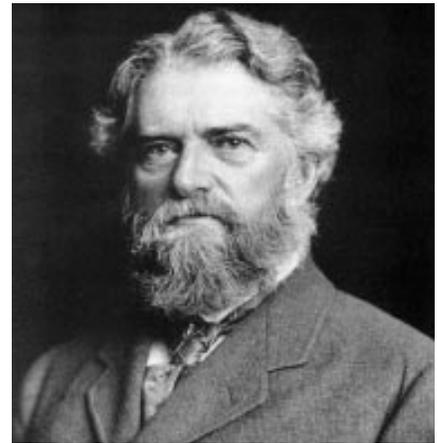
- Simple assertions that humans, but not computers, display some talent or property such as intuition, creativity, postmodern angst, kindness to animals, etc.
- Simple assertions that AI has run out of ideas, has been revealed as intellectually bankrupt, is a degenerating research paradigm, etc.
- Scornful observations that early (especially very early) optimistic predictions have not been borne out by practice, or that some problems are harder than some people once thought they were, or that some programs fail to exhibit genius-level competence, especially in performing a task they were not designed for.
- Claims that if AI were to succeed, things would be somehow worse than they are now, or that people would be somehow reduced in status.
- Any appeal to intellectual authority, especially to a continental philosopher.

We have decided to give an award for the silliest arguments against AI published each year. The Simon Newcomb Awards, as they are called, will be announced here in the AI Magazine. *Winners will be presented with a small statue (informally referred to as a 'Simon') in a short ceremony at a suitable national gathering. We invite nominations for future awards.*

The award is to be given for a specific argument, so that (just as with the Academy awards) a true star might receive a 'Simon' for each of several outstanding performances. We also expect to award the occasional 'Lifetime Achievement Award' in recognition of an entire career of silly attacks on the subject. Popular nominees (those supported by several submissions) will be announced at the same time as the Award winners. Those who are nominated but not selected for an Award may take solace in knowing that the nomination itself is a high honor. The nominees for the first Simon Newcomb Award were, Selmer Bringsjord, Harry Collins, Hubert Dreyfus, Gerald Edelman, Walter Freeman, Roger Penrose, Joseph Rychlak, John Searle, and Maurice Wilkes.

In the future, only one award will normally be made each year, but for this inaugural occasion, we are proud to announce four winners, in alphabetical order.

The Winners

The first Simon Newcomb Award Winners are:

- Hubert Dreyfus of University of California at Berkeley (*A Lifetime Achievement Award*)
- Roger Penrose of Oxford Universi-

ty *(for his revival of the classical Gödel argument)*

- Joseph Rychlak of Loyola University of Chicago *(for his exclusive-OR argument)*
- John Searle of University of California at Berkeley *(for his Chinese room and wall-is-wordprocessor arguments)*

Award Citations

Hubert Dreyfus

In many respects, Hubert Dreyfus' (1972) book *What Computers Can't Do,* reissued more recently with a similar title, was the pioneering work in this entire area. Hubert has been an inspiration and a guide to so many—including several of our other prizewinners—that it is impossible not to recognize his unique position in the field. Much of the content of these books consists of the kind of stuff we have excluded, however, and we couldn't find a really clear argument anywhere. A Lifetime Achievement award therefore seems just appropriate to recognize his special status and achievement, and we are pleased to award the first Simon Newcomb award to the distinguished Heideggerian scholar, Professor Hubert Dreyfus.

Roger Penrose

In a recent book (1989) Penrose has resurrected the classical Gödel argument, first presented by John Lucas many years ago. The reader is probably familiar with the outlines. Gödel showed that in almost any formal system there are sentences which, if the system is consistent, are true but unprovable. Penrose argues thus (in our words):

> I understand Gödel's argument, so I can *see* that if I had a Gödel sentence, it would be true. Thus, if I were merely an algorithm, I would have proven my Gödel sentence, which is impossible; so I am not an algorithm.

This argument has many silly aspects, but our favorite is that it depends on the speaker's assuming that he is both logically consistent and mathe-

matically omniscient, i.e., able to see the truth of any mathematically true sentence. Most versions of this argument miss these subtle aspects and are therefore simply invalid, but Penrose has both the intelligence to see the need for these claims and the honesty to make them both explicitly, thus raising himself from simple incompetence to a level of silliness which is clearly deserving of a Newcomb award. He claims omniscience merely by asserting it, a weak technique, but we were particularly impressed by his bold justification of his own consistency (p. 428):

> Mathematical ideas have an existence of their own, and inhabit an ideal Platonic world, which is accessible via the intellect only. When one "sees" a mathematical truth, one's consciousness breaks through into this world of ideas, and makes direct contact with it…mathematicians communicate…by each one having *a direct route to truth*. [Author's italics]

So much for Gödel's *second* theorem! Much of Penrose's book is thoughtful, coherent, and entertaining, and therefore would not qualify for an Award. Nevertheless, this argument is clearly one that Penrose wishes us to take seriously, as evidenced by the sober tone of Martin Gardner's foreword. Gardner is wrong, however: this is a laughing matter, and so we are honored to present a Simon Newcomb award to the famous mathematician and physicist, Professor Roger Penrose.

Joseph Rychlak

Rychlak, a well-known clinical psychologist, has published a book (1991a) on why AI won't work. While most of his prose is insufficiently coherent to really qualify for an award, he has earned one for the following silly argument. Our readers might think it was a mistake, but he repeats it clearly in a number of articles (1990, 1991b).

It has two parts. First, Rychlak argues that when a program runs on a machine it takes on the nature of the hardware (1991a, p. 164) :

> The two are effectively 'one'.

The software *is* the hardware at this point. This means, in effect, that the hardware must be accommodated as information is processed.

Others have made a similar claim, and this argument would hardly be deserving of an award by itself; but Rychlak achieves a unique status by the next part of his argument: since computer hardware relies on the exclusive-OR operation and cannot fully grasp the nature of inclusive-OR (yes, that's what he says), that it — and anything running on it — will therefore forever be unable to fully exhibit the uniquely human trait of "intrinsic oppositionality" (p. 335):

> None of (the forms of computer modeling) are amenable to intrinsic oppositionality in cognition. This is because they are founded on the logic of Boolean algebra, where disjunction is interpreted in binary fashion: "either *x* or *y,* but not both."

This argument combines ignorance and confusion in a dazzling display of silly virtuosity which is hardly equalled elsewhere. We are therefore proud to award an inaugural Simon Newcomb award to the distinguished social thinker, Professor Joseph Rychlak.

John Searle

Searle could have been given a lifetime achievement award for his proud career, now spanning some fifteen years, but his two most famous achievements are the Chinese room argument (1980) and the wall-is-wordprocessor argument (1992).

The Chinese Room — which we assume is familiar to our readers — is probably the most famous single argument ever produced attacking AI. Whether this deserves an award was, frankly, controversial. Some felt that an argument which can generate such an extraordinary amount of debate for more than a decade cannot really qualify as truly silly, no matter how inspired it may seem to many. Others said that the very familiarity of the argument may blunt one's appreciation for its true award-winning quality. Fortunately, John Searle guaranteed his place of

honor by his later performance, the wall-is-wordprocessor argument, which takes the same theme to greater heights. Searle, with characteristic élan, argues here that all AI claims are not false, but vacuous, because software is just an illusion (1992, p. 208):

1. For any object there is some description of that object such that under that description the object is a digital computer.

2. For any program and any sufficiently complex object, there is some description of the object under which it is implementing the program. Thus for example the wall behind my back is right now implementing the Wordstar program.

Searle's basic theme is that software doesn't really exist, and so the entire vocabulary of computer science is meaningless (p. 215):

…the 0's and 1's as such have no causal powers because they do not even exist except in the eyes of the beholder. The implemented program has no causal powers other than those of the implementing medium because the program has no real existence, no ontology, beyond that of the implementing medium. Physically speaking, there is no such thing as a separate "program level."

As with Newcomb on the difficulties of landing, one can see what he is getting at. Landing aircraft is a tricky business, and a programmed computer *is* merely hardware arranged in a suitable way, in a sense: but not a very interesting sense. Like Newcomb, Searle makes the mistake of taking a glimmer of insight and inflating it into a proof of impossibility, in the process declaring an entire science to be a hallucination.

Searle is in many ways our most deserving recipient of the Simon Newcomb Award. Like Newcomb, he is justly famous for his work in a different area. Like Newcomb, his arguments are clever. Like Newcomb, he conspicuously fails to understand a central idea of the new science and broadcasts his failure fervently and

eloquently. Like Newcomb, he responds to opposing arguments with wit and scorn while ignoring their content. Like Newcomb, he repeats himself again and again (Searle 1980b, 1980c, 1982, 1983, 1984a,1984b, …ah, the hell with it.) Like Newcomb, he advocates the utility of thought experiments while keeping himself aloof from technical details of actual research; and like Newcomb, he takes on an unwarranted air of authority when addressing a lay audience. The parallel is quite striking; and we are therefore very pleased to award an inaugural Simon Newcomb Award to the eminent philosopher, Professor John Searle.

Submission Requirements

Please send nominations for the 1995 Award by email to phayes@cs.uiuc. edu and kford@ai.uwf.edu. Nominations should give a brief description of the argument, a reference to its place of publication and the name and affiliation of the nominee. Permission of the nominee is not required.

An argument can win an Award only once, so repetitions of previous award-winning arguments are not acceptable unless they display some new significant variation on the original theme. If it is necessary to explain why the argument is silly, it may not be silly enough. The best arguments are those that a graduate

student in computer science might find hilarious. And finally, silly arguments within AI are not eligible for the award, only attacks on AI. Obviously, it would not be practical to give an award for every silly argument in AI.

Bibliography

Dreyfus, H. 1972. *What Computers Can't Do: A Critique of Artificial Reason.* New York: Harper & Row.

Newcomb, S. 1901. The Outlook for the Flying Machine. *The Independent,* 2508–2512.

Penrose, R. 1989. *The Emperor's New Mind.* New York: Oxford University Press.

Rychlak, J. F. 1990. George Kelly and the Concept of Construction. *International Journal of Personal Construct Psychology,* 3(1): 7–19.

Rychlak, J. F. 1991a. *Artificial Intelligence and Human Reason.* New York: Columbia University Press.

Rychlak, J. F. 1991b. The Missing Psychological Links of Artifical Intelligence. *International Journal of Personal Construct Psychology* 4(3): 241–249.

Searle, J. R. 1980. Minds, Brains and Programs. *Behavioral and Brain Sciences* 3: 417–457.

Searle, J. R. 1992. *The Rediscovery of the Mind.* Cambridge, Mass: The MIT Press.