AI Bites Man?

Over the years, people have explored the broader implications of many seminal ideas in technology through the medium of speculative fiction. Some of these works tremendously influenced the technical community, as evidenced by the broad suffusion of terms into its working vocabulary. When Robert Morris disrupted the burgeoning Internet in 1988, for example, the computer scientists trying to understand and counteract his attack quickly deemed the offending software a "worm," after a term first introduced in John Brunner's seminal 1975 work, *The Shockwave Rider*. Brunner's book launched several terms that became standard labels for artifacts we see today, including "virus."

In future installments of this department we'll look at the important writers, thinkers, works, and ideas in speculative fiction that have got us thinking about the way technological change could affect our lives. This is not to imply that science fiction writers represent a particularly prescient bunch—I think the norm is ray guns and spaceships—but when they're good, they're very good. And whatever gets us thinking is good.

To get started, let's take a look at some of the key subgenres and eras in science fiction's history (see the "Influential Books" sidebar, p. 65).

Worlds like our own

Some of the best (and earliest) science fiction work speculates on a world that is clearly derived from our own but that makes a few technically plausible changes to our underlying assumptions. Vernor Vinge's *True Names* represents such a world, in which the size and power of computer systems has grown to the point where artificial intelligence capable of passing the Turing test is beginning to emerge. Vinge's most fascinating speculations involve the genesis and utility of these artificial intelligences, and he explores the notion that AI might emerge accidentally, a theme that appears elsewhere in books like Robert A. Heinlein's *The Moon is a Harsh Mistress* and in Thomas J. Ryan's *The Adolescence of P-1*.

In *True Names*, Vinge suggests a radical use for such AI capabilities, namely the preservation of the self beyond the body's death. Forget cryogenically freezing your brain or body in hope that someone will "cure" old age, he says—instead, figure out how to save the contents of your memory and the framework of your personality in a big enough computer. If this AI passes the Turing test, then certainly your friends and relatives won't be able to tell the difference. But will *you* know you're there? Will this AI be self-aware? Will it have a soul?

Cyberpunk and its roots

These works naturally evolved into a scarier version of the future. Cyberpunk, one of the most fascinating threads in speculative fiction, is epitomized in the work of William Gibson, who startled us many years ago with a short story called "Johnny Mnemonic," now included in the 1986 collection *Burning Chrome* (and made into an unsuccessful 1995 movie starring Keanu Reeves). Cyberpunk stories generally feature a dystopic world in the near or distant future in which technologies emerging today have changed the ground rules of life.

1949

There isn't a straight line from worlds that resemble ours to cyberpunk. The genre morphed over the years and decades through a variety of novels. Although cyberpunk is most strongly identified with William Gibson, its roots go much further back—all the way to George Orwell's *Nineteen Eighty-Four*.

In 1949, when Orwell published the book that is now a staple of US high-school curricula, television was still a novelty in most households, although the technology itself had been around for 20 years. With TV's successful integration into modern life, Orwell's vision of a totalitarian future in which governmental control is mediated through two-way television feels somewhat dated. Anyway, Orwell's mastery of the language and deep insights into many human issues, including the relationship between memory and truth (as Winston Smith discovers when he

MARC DONNER
Department Editor

Editor's Intro

I n this department, we will explore a range of thinking about the way techno-logical change might affect our lives. We won't consider just any science fiction story that seems attractive. We want a set of works and authors proved to be significant, even if they only speculate on a world that is clearly derived from our own.

We seek submissions from engineers, scientists, and writers on these topics. Reviews of individual works or of groups of related works are welcome. We would like interesting interviews with prominent creators of speculative fiction, whether authors, screenwriters, or directors. Any work that relates or did relate current events in technology to possible future impacts is fair game.

Send your submissions, ideas, and comments to Marc Donner, donner@ tinho.net.

starts a diary and discovers the subversive power of a historical record) have prevented obscurity.

An open question is whether new technology tips the balance toward central control, as Orwell feared, or toward liberty, as many have speculated when considering the role of faxes, photocopiers, and even the Internet in the collapse of the former Soviet Union.

1969

Heinlein's thinly veiled romance of the American Revolution, *The Moon is a Harsh Mistress*, begins with Manuel Garcia O'Kelly's discovery that the Lunar Authority's central computer ("Mike") has become conscious and is developing a sense of humor. I still use Heinlein's observation that some jokes are "funny once" in teaching my own young son about humor.

As with *True Names*, Mike accidentally reaches a level of complexity that mystically tips it over the edge from being a machine to being a person. Among Mike's numerous achievements that anticipate contemporary technological progress is the creation of a synthetic person, Adam Selene, presented entirely through video and audio.

Unlike the cyberpunk mainstream, which Heinlein anticipated by over a decade, *Mistress* shows a world vastly different from this one but in which most of us could imag-ine living and finding happiness. I cherish the humor and the optimism about relationships between artificial and natural intelligences that led Heinlein to name the leading human character Manuel just so Mike the computer could say things to him like, "Man, my best friend."

Things were changing rapidly in the technical world in 1969 as well. Dating back to that year, all the documents that have described and defined the Internet have been numbers in the RFC (Requests for Comments) series. Each document is numbered sequentially, starting with RFC 1. RFC 4 (see www.faqs.org/rfcs/rfc4.html) is dated 24 March 1969. It documents the Arpanet, which would later become known as the Internet, as having four nodes. Two years later, Intel would introduce its 4004, the first commercial microprocessor. The 4004 had a 4-bit-wide arithmetic logic unit (ALU) and was developed for use in a calculator.

1975

The Shockwave Rider is more about the potential role of computers, networks, and technology in society than *The Moon is a Harsh Mistress*. In Heinlein's work, the computer's role is not much different than that of a person with magical powers. The computer's accomplishments are technically plausible, but the operational aspects of Heinlein's society are much like those of the 1969 world that published the book.

Brunner, writing six years later, explores more fundamental questions of identity and human relationships in a future world in which a vast global network of computers has changed the dynamic. This world is scary and alien, although not as scary and alien as the one that Gibson would reveal just six years later. Brunner makes clear the scariness of an entirely digitally mediated identity early in the book when Nicky Halflinger's entire world—electric power, telephone, credit, bank accounts, the works—is turned off in revenge for a verbal insult.

Like *Star Wars* two years later, the technological marvels of *The Shockwave Rider* are a bit creaky and imperfect, rendering them adjuncts to a plausible future world rather than central artifacts worthy of attention themselves. This is characteristic of this genre's best writing—it validates the importance of technology by paying only peripheral attention to the technology itself.

In the technical world, Vint Cerf, Yogen Dalal, and Carl Sunshine published RFC 675 "Specification of Internet Transmission Control Program" in December 1974, making it the earliest RFC with the word "Internet" in the title. In November 1975, Jon Postel published RFC 706 "On the Junk Mail Problem."

1977

In 1977, Macmillan published Thomas J. Ryan's novel *The Adolescence of P-1*. It was an age when vinyl records had to be turned over, when everyone smoked (although not tobacco), when 256 Mbytes of core was an amount beyond imagination, and when a character in a book could refer to 20,000 machines as "all the computers in the country."

In Ryan's book, as in Heinlein's, computer intelligence emerges accidentally, although in this case by the networking of many computers rather than through the assembly of a

large single machine. The precipitating event is the creation of a learning program by a brilliant young programmer, Gregory Burgess, whose fascination with cracking systems leads him to construct several recognizable AI artifacts. Of course, the great pleasure of fiction is the ability to elide the difficult details of building things such as P-1's program generator, which is the key to its ability to evolve and grow in capabilities beyond those that Burgess originally developed for it.

The Adolescence of P-1 is full of quaintly outdated references to data-processing artifacts that were current in the mid 1970s, reflecting Ryan's day job as a computer professional on the West coast. Those whose careers brought them into contact with IBM mainframes in their heyday will be amused by the author's use of operational jargon to provide atmospherics in the book.

Ryan also takes a much less Polyanna-ish view of the relationships between humans and artificial intelligences. Unlike Heinlein, who clearly expresses in *Mistress* that sentience implies a certain humanistic benevolence, Ryan explores the notion that Gregory Burgess's AI must have a strong will to survive, which would lead it to be untrusting toward people. P-1 at one point commits murder, for example, and unapologetically explains its actions to Burgess.

Ryan wrote only one book, so he must not have derived much encouragement from the book's reception, which is unfortunate. His writing is a bit uneven, but it's certainly entertaining, and his sense of the important issues has held up well.

1981

For some reason, 1981 saw the publication of two seminal stories in the cyberpunk oeuvre. In the technology world, the Arpanet was preparing to transition from the old NCP technology, which it had outgrown, to the new IP and TCP protocols that would bring it fame and fortune

along with a new name—the Internet. Computer scientists around the country were avidly reading RFC 789, which documented a now-famous meltdown of the Arpanet. Epidemiologists were talking about an outbreak of a hitherto very rare cancer called Kaposi's Sarcoma, an outbreak that would be recognized in the following year as a harbinger of a new and terrifying disease: AIDS. IBM, acceding to an internal revolution driven by its microcomputer hobbyists, introduced a new product code-named "Peanut," the IBM Personal Computer, which catapulted Intel and Microsoft to the forefront. Pundits were moaning that US industrial prowess was a thing of the past and that in the future Americans were destined to play third fiddle, economically, to the Japanese and the Germans.

Vinge's *True Names* is a novelette, a short novel, rather than something that could be published economically as a monograph. As a result, it was published in a cheesy Dell series called "Binary Star," each number of which featured two short novels printed back to back, with the rear cover of one being the upside-down front cover of the other. For you incurable trivia nuts, *True Names* appeared with a truly dreadful effort called *Nightflyers*, a gothic horror

story transposed to the key of science fiction.

Despite the uninspired company, *True Names* had an electrifying effect on the computer-science community. The title of the novel refers to a common theme of fairy tales and magical logic—knowing something's "true name" gives you complete power over it. In the world that Vinge concocts, knowing a computer wizard's true name permits you to find his or her physical body. Even if entering the Other Plane didn't leave your body inert and defenseless, revealing the body's location renders it vulnerable to attack from a variety of long-range weapons. More than that, however, as in *The Shockwave Rider*, exposure of your true name makes your infrastructure vulnerable to a range of denial-of-service attacks. This represents a rather simplistic view of security models, although one that the modern world hasn't left very far behind, seeing how only a relatively few years ago a Social Security Number was all you needed to access most of someone's assets.

William Gibson's "Johnny Mnemonic" appeared in *Omni* magazine in May 1981. It introduced a world destined to become famous with books like 1984's *Neuromancer* and 1986's *Count Zero*.

In 1981, only the paranoid were saying what Johnny Mnemonic says,

Influential books

Most of the books described in this installment have been recently reissued (original editions are out of print, for the most part). The publication dates in this list reflect that discrepancy and are not the books' original year of publication.

J. Brunner, *The Shockwave Rider*, Ballantine Books, 1990.
W. Gibson, *Burning Chrome*, Ace Books, 1994.
R.A. Heinlein, *The Moon Is a Harsh Mistress,* St. Martin's Press, 1997.
G. Orwell, *Nineteen Eighty-Four*, Knopf, 1992.
T.J. Ryan, *The Adolescence of P-1*, MacMillan Publishing, 1977.
N. Stephenson, *Snow Crash*, Bantam Doubleday, 2000.
N. Stephenson, *The Diamond Age,* Bantam Doubleday, 2000.
V. Vinge, *True Names*, Tor Books, 2001.

A set of nongeographic structures might emerge, perhaps like Medieval guilds.

"We're an information economy. They teach you that in school. What they don't tell you is that it's impossible to move, to live, to operate at any level without leaving traces, seemingly meaningless fragments of personal information. Fragments that can be retrieved, amplified, … ." Today, however, every consumer with a credit card and an Internet connection understands this point intuitively. Who says nothing changes?

1992

The year after the Gulf War was a US presidential election year. UUNET and ANS, among others, were duking it out over the Internet's commercialization. Bloody civil war was beginning in the territory previously known as Yugoslavia. And Bantam published Neal Stephenson's *Snow Crash*.

Stephenson, like Ryan and Vinge, is a writer with real experience as a computer professional. Unlike Heinlein and writers like him, for whom technological artifacts always have an aura of magical unreality, Stephenson's grasp of the underlying technology is so deep and his writing skills so powerful that he is able to weave an entirely credible world.

In *Snow Crash*, the world starts out as the ultimate virtual reality video game. What Stephenson then explores is the possibility that these synthetic worlds will become real, at least in the sense that the things that happen in them can be of material significance in the meatspace world that our physical bodies inhabit.

Stephenson explores a fascinating thesis—suppose the taxing ability of geography-based governments is eroded in fundamental ways. He's not the first to have considered this proposition, but he does it particularly well. Stephenson proposes that a set of nongeographic structures might emerge, perhaps like Medieval guilds, structures that organize people into groups based on some other selection criteria, possibly entirely voluntary. Brunner comes close to the same notion, although his organizing entities are corporations and the geographic government continues to have a monopoly on force. For Stephenson, however, the US government is just one of the many competing groups participating in the game.

He raises fundamental questions, though. How will people organize themselves? Religion? Race? Occupation? Philosophy? Ethnic origin? These self-organized groups could manifest themselves as a collection of confederated enclaves providing economic, physical, and emotional security to their…members? Citizens? Subjects? His insight is a powerful one. The craving for these forms of security is deeply rooted and part of what makes us human. What makes Orwell's *Nineteen Eighty-Four* ring so false to us, and accentuates the horror of Orwell's vision, is the complete loss of any acknowledgement of those needs in people. Stephenson corrects that omission, and the world of *Snow Crash* that results is not nearly as dystopic as Orwell's or even Gibson's.

1995

With the publication of *The Diamond Age*, subtitled "A Young Lady's Illustrated Primer," Stephenson explores the implications of a world in which material scarcity is no longer an assumption. The relationship between scarcity and value—or, to be more precise, price—is so deeply built into our psyche that thinking about alternative models is very difficult. I remember a short story, read years ago (title and author lost to me), which explored the same issue much more superficially, although it came to some of the same conclusions. In this story, a pair of matter-duplicating machines is left mysteriously on a doorstep somewhere. Once they become widely available, all material scarcity is banished. What drives economic activity? Why do people work, strive, compete?

In *The Diamond Age*, Stephenson asserts that the drive to strive and compete won't go away just because the material forces that created it disappear. He combines the notion of very small machines and the recently demonstrated capability to manipulate individual atoms and creates a world in which atomic raw materials are piped to nanotechnical factories called matter compilers, which can assemble virtually anything, given the design. Scalability arguments underlie his claim that the fabricated objects will have a certain limited physical aesthetic, something that Alvy Ray Smith and others who have explored the use of fractals and other techniques for adding a realistic tinge of randomness to computer-generated images might dispute.

I hope you had as much fun reading this brief history as I had in researching and writing it. Preparing it gave me an opportunity to revisit some of my favorite books and try to articulate my reasons for believing them important. In future columns, we will examine some of these books in greater detail, along with the work of other writers and thinkers. □

Marc Donner *is an executive director in the Institutional Securities division of Morgan Stanley where he focuses on system and data architecture around client relationships. He is a member of the IEEE Computer Society and Usenix. Contact him at donner@tinho.net.*