CAUTIONS ON THE SUPERHUMAN TRANSITION

BY WILLIAM H. CALVIN

I guess the critical parameter is how fast the Singularity takes place once the first-order workalikes come into existence.

If there is an intellectual runaway (ultra-intelligent machines quickly making even more intelligent ones), and progress speeds up by a ratio like that of human progress over natural evolution, there will be no possibility of human control. (Sometimes I feel as though I'm sitting on a beautiful beach, watching a tsunami sweeping in upon me.)

If the process takes five or ten decades, then planning such as William Calvin discusses here is critically important. And since no one knows the "transition speed," we'd better take the planning very seriously.
—Vernor Vinge

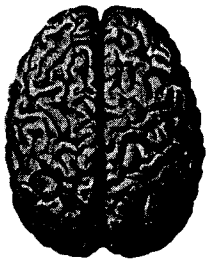
SOMEDAY soon, a computer will start to engage in conversation — to chat. Even as a first-order approximation, lacking most humanlike behaviors, this workalike will set in motion one of those historical transitions; afterward, nothing will be the same. Perhaps this change won't qualify as a singularity — an instant shift into totally unpredictable consequences — but we surely have a major transition coming up in the next several generations of humankind, and it needs discussion now.

As a neurophysiologist interested in how the human cerebral cortex's circuitry allows us to construct sentences and to speculate about tomorrow, I suspect that successful "downloading" of an individual's brain to a workalike computer is unlikely; dementia, psychosis, and seizures are all too likely. But on the basic question — of whether we can build a computer that talks like a human, is as endearing as our pets, thinks in metaphor and multiple levels of abstraction — there, I think that it will be relatively easy to construct a first-order workalike that reasons, categorizes, and understands speech. We'll even be able to make it run on principles closely analogous to those used in our brains.[1] I can already see one way of doing this, extrapolating from known wiring principles of the cerebral cortex, and there might be ad hoc ways of doing it — e.g., artificial intelligence.

Even the first-order workalike will be recognizably "conscious" — probably as self-centered as we are. And I don't mean the trivial aspects of consciousness (aware, awake, sensitive, arousable). This consciousness may comprise the capacities for focusing attention, mental rehearsal, abstraction, imagery, subconscious processing, "what-if" planning, decisionmaking — and especially the narratives we humans tell ourselves when awake or dreaming. To the extent that such functions can operate far faster than they do in our own millisecond-scale brains, we'll see an aspect of "superhuman" emerging from the "workalike."

But that's the easy part, just extrapolation of existing trends in computing technology, AI, and neurophysiological understanding. There are at least three hard parts.

ONE HARD PART will be to make sure the workalike fits into an ecology of animal species. Such as us.



Especially us. Competition is most intense between closely related species. That's why none of our Australopithecine cousins are still around, and why only two omnivorous ape species have survived (chimpanzee and bonobo). Our more immediate ancestors probably wiped out the others. "To keep every wheel and cog," wrote Aldo Leopold, "is the first precaution of intelligent tinkering."[2]

When automation rearrangements occur so gradually that no one starves, they are often beneficial. Everyone used to gather or hunt their own food, but agricultural technologies have gradually reduced the percentage of our population that farms to about 3 percent. That's freed up many people for other pursuits. The proportional mix of those "occupations" changes over time, as in the shift from manufacturing jobs to service jobs in recent decades.

Workalikes will further change the mix, displacing even some of the more educated workers — but not without significant benefits. Imagine a superhuman teaching machine as a teacher's assistant: one that could hold actual conver-