

Some Strangeness in the Proportion, or How to stop worrying and learn to love the Mechanistic Forces of Darkness

Eric Dietrich
Philosophy Department
Binghamton University
Binghamton, NY 13902-6000
dietrich@binghamton.edu

Abstract

Understanding humans requires viewing them as mechanisms of some sort, since understanding anything requires seeing it as a mechanism. It is science's job to reveal mechanisms. But science reveals much more than that: it also reveals enduring mystery – strangeness in the proportion. Concentrating just on the scientific side of Selinger's and Engström's call for a moratorium on cyborg discourse, I argue that this strangeness prevents cyborg discourse from diminishing us.

Unless you have spent the last several decades hibernating, under a rock, in a cave, on Mars, you are worried about the fate of the entire human species. This is a new worry. As the bubonic plague swept across Europe, it might well have been that various ancestors of some of us worried about the fate of their city or tribe, or even the fate of their people or nation. However, we are the first to systematically worry about the fate of our entire species. The potential disasters funding this worry are depressingly numerous and bewilderingly diverse: collision with an asteroid, runaway global warming, pandemics, and nuclear and biological warfare or terrorism, to list the current top four. Yet, as part of their job description, humanistic philosophers and social theorists are required to look beyond this list. Perhaps there are subtle dangers lurking in areas we mistakenly think are safe. Selinger and Engström are two such philosophers who claim to have located just such area: the one which includes cyborg discourse and the computational theory of mind (2007; see also, Madison, 1991). According to Selinger and Engström, this entire area is in fact so dangerous that they urge a moratorium on all cyborg discourse.

The area Selinger and Engström pinpoint lies at the intersection of various technologies and sciences, including artificial intelligence, cognitive and computer science, communications and media, commerce, and medical replacement technology, including nanotechnology. The danger is due the definition of "human" arising at this

intersection. The definition of humans Selinger and Engström see emerging is one that embraces cyborgs and robots, and lauds the engineering of better partial-humans or even replacing us outright with robots or artificial agents of one sort or another. At the heart of this new definition of "human" is the computational theory of mind, the brave, new theory that rescued psychology from behaviorism and made artificial intelligence possible. Selinger and Engström believe that this computational definition of what it means to be human will lead us to a mechanistic, profoundly limiting view of ourselves that will likely have disastrous political, social, and environmental consequences. Since they quite rationally would like humankind to avoid adding yet another way to ruin itself to an already burgeoning list, they want to block this definition. Hence, they want computationalists to stop promoting cyborgs, robots, and to limit the use of the computational theory of mind.

One can be forgiven for thinking that Selinger and Engström are over-reacting, for, a sentiment similar to theirs plagues (and has plagued since 1859) the theory of evolution. If we are nothing but African apes, then we are likely to recast ourselves as *merely* animals, and thereby give up (among other things) our duties to ourselves and the rest of the planet. But, we are African apes. Yet we still have duties to ourselves and the rest of the planet, and we act on said duties. So, though we are African apes, nothing dire follows from that.

But Selinger and Engström are, in fact, on to something, and the evolutionary worry is deeper than it appears, as we will soon see. Mechanistic views of living things really do diminish them. Medical students in Victorian England used to dissect living, fully awake dogs in order to learn anatomy. The desperately screaming dogs were held in place by nailing their paws to wooden boards. None of this bothered the students. They believed that the dogs' "pain-response" was really just a *mechanical* reaction, for after all, dogs couldn't feel pain, they weren't conscious; they couldn't be conscious, they were *mechanisms*. History is rife with examples of this sort of thing. Anyone the least bit familiar with animal rights activism in the twenty-first century knows that the fight is about getting experimenters to view their subjects as more than mere mechanisms, akin to lab equipment. Almost every single animal experimented on, whether it is for groundbreaking cancer research, or for better cosmetics, is, in the mind of the experimenter, categorized as a tool, which is a kind of machine. And we can do anything we want to machines. In fact, what gives the evolutionary worry above its teeth is that by reducing humans to *mere* animals, evolution reduces humans to machines because in fact many

people think of animals as machines.

So if the computational theory of mind is fully embraced, if humans deeply and completely think of themselves as machines, we will be diminished, it seems.

But is this really true? The key, I think, is what Selinger and Engström say in their conclusion: "Our concern is that CTM is an effort to convert something we do not know and can not manipulate terribly successfully—individual minds and consciousness—into something we do and can manipulate pretty well—numbers, codes, interactive algorithms, etc." There are two ways to interpret this claim, depending on how one interprets the verb "convert." I call these the *substitution* way, and the *scientific* way. These two interpretations pervade the paper. On the substitution interpretation, "convert" means to change the topic, to swap out minds and consciousness as the focus of our attention and swap in numbers, symbols, and algorithms. We substitute something we can understand for something we can't. On the scientific interpretation, we come to understand minds and consciousness *by discovering that they are symbolic processes*, i.e., *algorithms*. (Actually, understanding minds qua thinking things and understanding consciousness are two entirely different things. Understanding consciousness is right out (see Dietrich and Hardcastle, 2004). But understanding minds as thinking things is definitely possible. The computational theory of mind is only useful for understanding thinking, but restricted to thinking, it is probably correct (Dietrich, 2006, 2001, 1990; Dietrich and Markman, 2003). So from here on out, I shall restrict my discussion to thinking, to cognition.)

Selinger and Engström sometimes talk as if they are worried that we will substitute what we understand for what we don't. But we are unlikely to do this for any length of time. Never mind the social consequences Selinger and Engström are worried about, the raw substitution will have *scientific* consequences, and if those don't pan out, we will abandon the substitution (adverse social consequences would also be a strong incentive). Humans don't often engage in massive self-deception when it doesn't work (even religion is sidelined when there is real work to do).

At other times, Selinger and Engström seem to oppose the scientific quest to understand the mind. There is a deep relation between understanding anything and reducing that thing to some sort of mechanism (broadly construed): understanding a thing seems to *entail* reducing it to a mechanism of one kind or another. So, Selinger and Engström *are* opposed to scientifically understanding the mind. I call the background

fear supporting this opposition the *fear of the mechanistic forces of darkness* (Dietrich, 2000, 1995). The fear is unwarranted, but for an unexpected reason. Contra Clark (2007), Selinger and Engström have not misunderstood the nature of computation, they are not operating with an outdated notion, rather they, like many, fail to appreciate the true nature of science.

Science creates mysteries. Yes, it solves them, too, but that is only half the story. I mean this in three ways. First, science creates new mysteries out of its current solutions: solving one problem creates another one. For example, we are an African ape, but in fairly obvious ways, we constitute our own phylum. Though the theory of evolution is one of the most powerful and well-confirmed theories ever, we are now confronted with the problem of explaining how, while sharing around 95 percent of our DNA with chimpanzees, we are nevertheless orders of magnitude more intelligent, and hence, orders of magnitude different. In this first mode, science produces a string of problems, one after another, but none are truly enduring.

Second, science creates enduring mysteries. In this mode, science reveals mysteries that cannot be solved (Dietrich, 2007). Consciousness is the most important example. It is *the* central property of the mind, yet is forever beyond reductive explanation (one has to be quite careful what one infers from this, see the previous Dietrich and Hardcastle reference). This fact alone is enough to guarantee that the computational theory of mind is not to be feared: it *can't* achieve hegemony, dark or otherwise, over explaining humankind. Mathematics, too, reveals many of these enduring mysteries (Chaitin, 2005). But of course, the discipline that shines here is philosophy. Philosophy shows us that the most important questions about human existence are intractable conundrums that can never be solved (Dietrich and Hardcastle, 2004; Nagel, 1986, 1979). Of course, many philosophers will disagree with this, thereby proving my point.

Third, because of the first two ways of producing, or, better, revealing mysteries, doing science literally shows us, not just an understandable world of intricate mechanisms, but also a world more mysterious and therefore more open-ended and oddly beautiful than is commonly noticed. *How is it possible* that solving one problem produces another (mere complexity doesn't seem sufficient)? *How is it possible* that some central problems are intractable? I doubt these two questions have answers. What kind of a universe do we live in and what kind of creatures are we that all these mysteries

exist and are graspable as such by us? It is unlikely the computational theory of mind, or any other theory, can answer these questions.

So what does this tell us about Selinger's and Engström's thesis? The computational theory of mind can tell us only a part of what we want to know. We won't be inclined to use it to diminish humans since such diminution will obviously be a lie. There is clearly more to us, to everything, than theory can reveal. And the computational theory of mind shares this property with all other theories no matter what their subject. I don't know why science and the universe work this way, but they do. So, the mechanistic forces of darkness are not dark at all. In fact, they are forces of light, for it is via mechanism (again, broadly construed) that we understand the universe. Without mechanism, all would be chaos. Nevertheless, there remains some strangeness in the proportion. And this ought to reassure Selinger and Engström and all who worry about the onslaught of science.

References

- Chaitin, G. (2005). *Meta Math!: The Quest for Omega*. Pantheon Books, New York, 2005.
- Clark, A. (2007). The Frozen Cyborg: A Reply to Selinger and Engström. *Phenomenology and the Cognitive Sciences*, this issue.
- Dietrich, E. (2007). Science and Enduring Mystery. Paper presented to the *Science and Enduring Mystery Symposium*, October, 2007, Binghamton University.
- Dietrich, E. (2006). Representation. In P. Thagard (ed.), *Handbook of Philosophy of Psychology and Cognitive Science*. Elsevier Science.
- Dietrich, E. (2001). It Does So! Review of Jerry Fodor's *The Mind Doesn't Work That Way: The scope and limits of computational psychology*. *AI Magazine* v.22, n. 4, pp. 141-144.
- Dietrich, E. (2000). Cognitive Science and the Mechanistic Forces of Darkness, or Why

- the Computational Science of Mind Suffers the Slings and Arrows of Outrageous Fortune. *Techné: eJournal of the Society for Philosophy and Technology*, Winter, <http://scholar.lib.vt.edu/ejournals/SPT/v5n2/dietrich.html>
- Dietrich, E. (1995). AI and the mechanistic forces of darkness. *J. of Experi. and Theor. AI*, 7 (2), 155-161.
- Dietrich, E. (1990). Computationalism, *Social Epistemology*. 4 (2), pp. 135-154. (with commentary).
- Dietrich, E. and Hardcastle, V. (2004). *Sisyphus's Boulder: Consciousness and the Limits of the Knowable*. John Benjamins.
- Dietrich, E. and A. B. Markman (2003). Discrete Thoughts: Why cognition must use discrete representations. *Mind and Language*. v. 18, n. 1, pp. 95-119.
- Madison, G. (1991) Merleau-Ponty's Deconstruction of Logocentrism, in M. Dillon ed. *Merleau-Ponty Vivant*. SUNY Press: Albany.
- Nagel, T. (1979). Subjective and Objective, in *Mortal Questions*. Cambridge, UK: Cambridge University Press, pp. 196-213.
- Nagel, T. (1986). *The View From Nowhere*. New York: Oxford University Press.
- Selinger, E. and Engström, T. (2007). A Moratorium on Cyborgs: Computation, Cognition, and Commerce. *Phenomenology and the Cognitive Sciences*, this issue.