On the Transcendence of the Martian Penguin
(originally published on TQ, 7/12/08)

What follows has undoubtedly been written before, probably word for word, but I don't think I am going to sleep until *I've* written it. Yeah. So:

There is a fairly strong parallel between the intuitive appeal of the Turing Test for artificial intelligence and the debate over the constituents of life. In particular, several of the usual definitional elements of "living organisms" are clearly extraneous to what we recognize as life. If I run across something that walks and talks like a penguin, let alone like a blue-green algae, I am *certainly* going to consider it to be alive.

But let's suppose that this is a miraculous Martian penguin, of unknown origins. It is unrelated to the terran family of organisms, and in fact is unrelated to any other organism: it's unique. It is incapable of reproduction, apparently immortal, not carbon-based, not cellular, and runs such a perfect homeostatic loop that it does not need to metabolize outside energy. In short, it breaks pretty damn near every one of the usual "rules" for what is alive. But it does act like a penguin, waddling about, barking at the astronauts, sliding down the talus slopes. Got it? It is a discrete material field that maintains homeostasis and responds to stimuli.

Without any question, the exobiologists would see this creature as a living thing, without even realizing that it lacks qualities shared by most other living things on earth. (I say *most*. We are accustomed to shoehorn into the definition the large number of earth organisms that do not reproduce (as individuals), or are not part of a living species (during extinctions, or by mutation), or do not metabolize outside matter (seeds, eggs, etc.) or do not have cells (as with viruses, although these at least get debated). Note that our excuses for these inclusions radically weaken the definitional claims. Rather than say "an organism must have quality X to be a living thing," we say: "an organism must have quality X, or be descended from an organism that has quality X during at least some part of its life cycle.")

Again, we cannot dodge the semantic bullet by asking where the penguin came from. A fundamental tenet of biology in a creationist-ridden age is that it is not interested in abiogenesis. It is all the same thing if the penguin was descended from Martian dinosaurs or baked from scratch by a whimsical space god.

Furthermore, as our exobiologists began to investigate the Martian penguin, and realized what an odd duck it was, they would not come to think of it as *less* alive. Rather, they would come to think of it as a *superior form* of life. Most especially, if the penguin had a frictionless metabolism and was immortal, we would immediately identify these traits as a sort of perfection of life processes. In retrospect, the idea that ingestion, excretion, and death are constituent elements of life seems like defining a truck as something that has rust, bad brakes, and at least nine more payments.

Here, too, we see the parrelel with AI. The quickest way for an A.I. to bust a Turing Test is to answer a question too quickly and with too much precision and accuracy. But among humans, the accepted group of intelligent organisms, speedy, precise, and accurate answers are considered a sign of intelligence, and there is no upper boundary on this identity.

What I want to propose, then, is a general pattern for these "walks like a duck" tests:

• *In tests where an entity E of unknown status S_1 is being compared to a population P whose status is defined* a priori*, then:*

• *Any continuous variable whose value is considered a monotonically increasing (decreasing) indicator of S_1 within group P cannot have a higher (lower) value at which it is considered an indicator of -S_1.*

Or in other words:

• *If you would say that a human being was a genius for doing something, you can't turn around and say that an algorithm is unintelligent just because it did the same thing (...better).*

This provides what I would call a non-transcendable definition of S. The alternative, a transcendable definition, inherently posits that there is a category S_2 "above" S_1; that is, equivalent to S_1 in every way except for certain variables in which it performs *so well* that we place it outside of S_1. Examples of transcendable definitions include teenagers, middleweights, or—more subtly—bacteria (qua unicellularity). But is an essential part of human pride that being alive and intelligent are *not* transcendable categories: we never speak of any range below which *and beyond which* something is not intelligent.

We are accustomed to note that when artificial systems outperform biological systems in one arena, they will fall massively short in another arena. ELIZA responds to you in milliseconds, but makes no sense. A bottle of a frigorific solution has homeostasis (for awhile), but it doesn't *do* anything. These trade-offs, however empirical, are not embedded in our definitions. If a glass of ice water suddenly started tap-dancing and laying eggs, or if ELIZA wrote the great American novel, we would be in difficulties to explain why they were not, respectively, alive and intelligent.

We haven't had this problem yet. There is no Martian penguin. And thus we have designed tests (like the Turing Test) that are germane on the downslope, but produce false negatives on the upslope. Any conceivable version of strong AI would fail the Turing Test, just as our penguin fails the usual test of alive-ness, because both tests imply transcendable definitions.

Perhaps it would be more accurate, in many ways, to retract the scope of our terms somewhat. When we speak of life processes, we are usually looking only at particular scales of time and space. We do not ask whether a proton is alive, and we speak of ecosystems, stars, and galaxies as alive only by analogy. What we *really* mean by living organisms is matter fields that exhibit homeostasis and response to stimuli across about 10^{-8} to 10^{3} meters, and 10^{3} to 10^{12} seconds. Below and above those chalk lines, we do not look for "organisms." Clearly this is a transcendable definition, and yet it is an appropriate one. A star exhibits many of the same life processes as a starfish, though it is roughly a billion times larger and lives ten billion times as long. But we do not expect biologists to concern themselves with stars, or astronomers to concern themselves with marine biology. There is an overarching field of study, perhaps thermodynamics or complexity theory, but they will have to come up with some term that transcends "life." And as for intelligence....