

CORRESPONDENCE ABOUT SELF-MOTIVATED AGENTS

William Bricken

March 1992

Memo 1

I have never been able to figure out how, at a given point in time, the current theorists believe that they (today!) have the answer. Cosmology is a superb example. But we don't need statistical mechanics, we need a theory of gestalt (as per Maturana's Tree of Knowledge).

Definition of life:

maintains its own boundaries and
is internally independent (autonomous)

So an artificial creature must function without regard to input/output and it must transmute perturbation for the outside to maintain its own internals. (your self-consistency, although I don't believe consistency has anything to do with it, more like resilience: maintenance of organization through variation.)

Here's my Maslow's hierarchy for autonomous programs:

1. Reactive:
 - sensors
 - effectors
 - map from sensors to effectorsthat is S->R behaviorism
2. Responsive: add
 - internal memory
 - maps from sensors to memory
 - maps from memory to effectorsthat is S->O->R behaviorism
3. Determined: add
 - maps from memory to memory
 - prioritized memory states (goals)that is O->S->R->O self-motivation
4. Coordinated: add
 - time-stampsthat is an anchor to external reality and other processes
(not necessarily other autonomies, could be Minsky-mind)

5. Autonomous: add
 -- meta-operators
 that is ability to alter all internal mappings

The implementation is a simple rulebase with pattern-matching and substitution as the only computational activities. (as you might expect, I still believe that matching and substitution are the fundamental operational semantics of all computation). Don't ask me at this point which rules to write.

Memo 2

A wonderful idea that we can model evolution. Sure to be publishable, sure to be total bullshit. Yep, I didn't really mean "Maslow-ian" cause I find those parallels repugnant. [Haven't railed in ages!] Even more so is the idea that we are God analogs for sentient computers. You see the messy epistemology (God, behavior, needs) such thoughts lead to.

I still believe that there is no application of biology to computation, save for areas that biology is itself not biological. You point out that survival itself is an inappropriate concept. Fundamentally, programs do not behave, they are syntactic jugglings. I vote for a completely new conceptual vocabulary. Programs are Aliens. Now, we are remarkably adept at partitioning our cognition into tiny domains that simulate computation, but please, such a partition is not a god, it is a degeneracy. Or: Alife might qualify for analogical consideration after it evolves *for a couple of million years*.

But I will not duck out on any mapping game:

- Survival -> runs without crashing
- Food -> has long-term access to processing resources
- Shelter -> free of i/o interrupts
- Companionship -> calls other programs (which also have 1-3 above)
- Fulfillment -> stores and accesses own output
- Transcendence -> some programmer runs it in a cognitive partition.