Naive Physics and Alien Minds:

Speculations about Intelligences Adapted to Alien Environments

Allan Combs, California Institute of Integral Studies
Ben Goertzel, Novamente LLC

April 3, 2009

Introduction

What properties might characterize an advanced nonhuman mind that was adapted to an environment dramatically different from our own?

We lack a sufficiently advanced and abstract theory of mind to address this question rigorously; but even from the little we do understand about intelligence, it is possible to draw some nontrivial conclusions.

Here we present some speculations about how differences in the "naive physics" presented to minds by different environments, might be reflected in differences in these minds' cognitive and social structures. We illustrate our general ideas with some specific speculations about minds in complex fluid environments.

Minsky's "Universal" Ingredients of Intelligence

Minsky (1985) assures us that any truly intelligent being, at least any one that we are likely to try to communicate with, must think and act in a fashion not unlike ourselves. He posits that any intelligent alien would share with ourselves certain ingredients including.

SUBGOALS ---------- to break hard problems into simpler ones.
SUB-OBJECTS ------ to make descriptions based on parts and relations.
CAUSE-SYMBOLS--- to explain and understand how things change.
MEMORIES ---------- to accumulate experience about similar problems.
ECONOMICS -------- to efficiently allocate scarce resources.
PLANNING ---------- to organize work, before filling in details.
SELF-AWARENESS-- to provide for the problem-solver's own welfare.

We will take issue with his hypothesis below, in our discussion of minds in complex fluid environments.

But even if Minsky's observation were correct, it's not clear how useful it would be. His ingredients are so abstract that they might apply to any of a wide variety of environments ("worlds"), so that the actual form of a communication -- the physical form of information exchange, metaphors utilized, etc. – might vary extremely widely.

Naive Physics

An important notion to introduce here is that of “naive physics” – that is, the physics of an environment as it is presented to an embodied mind in its everyday interactions with the world (a common notion in AI theory; see e.g., Hayes, Patrick (1995). The Second Naive Physics Manifesto. Computation and Intelligence: Collected Readings. AAAI Press).

Specific aspects of naive physics related to objects in our own everyday environment, for example, include (but are not limited to):

- Recognition of objects amidst noisy perceptual data
- Recognition of surfaces and interiors of objects
- Recognition of objects as manipulable units
- Recognition of objects as potential subjects of fragmentation (splitting, cutting) and of unification (gluing, bonding)
- Recognition of the agent’s body as an object, and as parts of the agent’s body as objects
- Division of universe of perceived objects into "natural kinds", each containing typical and atypical instances

Specific aspects of naive physics related to temporality and causality, in our own everyday environment, include:

- Distinguishing roughly-subjectively-instantaneous events from extended processes
- Identifying beginnings, endings and crossings of processes.
- Identifying and distinguishing internal and external changes
- Identifying and distinguishing internal and external changes relative to one's own body
- Interrelating body-changes with changes in external entities

Our own minds – and the minds of other animals sharing our everyday environment – are strongly tied to these “naive physics” patterns, which have to do with the ways our brains and bodies are structured, the way our environment is structured, and the emergent patterns spanning our brains, bodies and environment.

Minds in Complex Fluid Environments

As an intriguing thought-experiment regarding intelligences adapted to fundamentally alien environments, it's interesting to ask: What differences might characterize the mind of an intelligent agent evolved in a “complex fluid

environment" -- meaning, a portion of the universe dominated by liquids or plasmas widely varying in viscosity, density, Reynolds Number and so forth?

To jog your mind regarding the complexity of fluid behavior you may wish to look at the following videos and others like them:

- turbulent water flow: http://www.naturefootage.com/video_clips/NZ17_024
- a fluid vortex: http://www.youtube.com/watch?v=j2dNHQtcmSM&feature=related
- patterns vaguely reminiscent of lifeforms: http://video.google.com/videosearch?q=turbulent+flow&hl=en&emb=0&aq=-1&oq=#q=turbulent+flow&hl=en&emb=0&start=30
- http://www.youtube.com/watch?v=_AJgEa2dbJU&feature=related

In a complex, fluid environment environment, one may suspect that – on average, compared to the situation in worlds dominated by solid objects -- each event would be strongly correlated with multiple, relatively easily detectable future and past events.

Hence, one may suspect that minds adapted to this sort of environment might not develop a "folk psychology / folk physics" notion of causality in the same way that humans have. (Note that causality is not a fundamental concept in any human scientific theory. Rather, it's an intuitive concept we have created to help conceptualize our everyday world. It is a "naive physics" notion not a "scientific physics" notion.)

Causality, however, is arguably one of the roots of the notion of "will" as humans experience it. The concept that "agent A wills X" is tied in the the notion that "some internal event E1 within A's mind, causes some internal event E2 in A's mind, body or surroundings to happen." Of course there is much more to will than causality – but without something similar to our notion of causality, nothing very similar to our notion of will is likely to exist.

Without causality and will, what kind of psychology might exist? Brainstorming freely, a possible alternative could be a psychology built around the notion of "flow," so that for instance

- Instead of willing X, an agent might think of itself as flowing in direction X.
- Instead of X causing Y, an agent might conceive of situations using patterns such as "there is a flow leading from X to Y."

The naïve physics of agents in fluid environments would involve many different kinds of flow, including various sorts of turbulent and laminar flow, potentially different sorts of plasmatic state, etc.

If these agents often encountered turbulent regimes, then just as we have a well-developed naïve physics for relations between objects, they might have a well-developed naïve physics for the transition from orderly to chaotic/turbulent flow. “Routes to chaos” and “routes to complexity” as studied in dynamical systems theory might be intuitive and immediate to these agents. (A vaguely “fluid-like” visualization of period-doubling bifurcation route to chaos may be found at http://physics.mercer.edu/hpage/pcompr.gif).

Rather than focusing on building items from components, these agents would likely focus on creating self-organized patterns via exerting pressures of various sorts from various directions. Understanding of routes to chaos and complexity would be important here. Shaping would likely be more important than building. Physical tools might be less important than patterned sequences of contextually appropriate shaping activity.

Our language, itself, is based on decomposition of meaning into combinations of discrete objects; but this is not the only possible way to communicate. One can imagine a language (an organized system of patterns intended to convey meanings) composed of continuous flows and changes in fluid state. In this sort of language, there might be no atomic component remaining persistent across different instances of communication – that is, nothing similar to a word, character or phoneme. Rather, the persistent patterns across different instances of communication with related meaning might be higher-level – say, statistical structural patterns in a strange attractor emergent from a borderline-turbulent fluid state.

The rigid distinctions between individuals that we take for granted might not be so relevant in this context either. Consider for instance the case of intelligent agents comprising viscous fluid or plasma regions, floating around in some less viscous substrate. These agents might then be able to merge with each other, or temporarily open their boundaries to each other to share internals to a limited extent. Also, in this state, flow patterns in one agent might be able to “seed” similar flow patterns in another, allowing transmittal of “thought patterns” in a direct way not easily possible for agents with solid boundaries around their brains. If this sort of communication were possible, the notion of an individual mind as an atomic entity might be much less significant, and the “phenomenal self” on which human psychology is based might take a very different form, or not exist in any recognizable way.

Overall, what this thought experiment suggests is that, in the case of intelligent agents evolved in a complex fluid environment, very familiar qualities such as cause, will, self, and compositional language might not be present.

Returning to Minsky's enumeration above, we find that many of his posited universals could plausibly be quite different for minds evolved for complex fluid environments (the most suspect ones are underlined):

SUBGOALS ---------- to break hard problems into simpler ones.
SUB-OBJECTS ------ to make descriptions based on parts and relations.
CAUSE-SYMBOLS--- to explain and understand how things change.
MEMORIES ---------- to accumulate experience about similar problems.
ECONOMICS -------- to efficiently allocate scarce resources.
PLANNING ---------- to organize work, before filling in details.
SELF-AWARENESS-- to provide for the problem-solver's own welfare.

The notions of subgoal and subobject seem closely tied to the solid-object environment and might not play any significant role in the architecture or psychology of intelligences evolved for complex fluid environments. The notion of cause-symbols might be replaced with something quite different based on notions of flow rather than causation. And the rigid notion of an individual self, strictly separate from other selves, might be replaced by something quite different in a mind adapted to a complex fluid environment.

Communicating with Intelligences in Profoundly Alien Environments

How then might we communicate with intelligences existing in a complex fluid environment, or another environment very different from our own?

Applying a general pattern recognition algorithm to detect the patterns constituting their minds and communications would likely be infeasible.

One potentially valuable conceptual strategy would be to:

1. Use a pattern recognition algorithm to recognize simple regularities in the environment

2. Project identical copies, and novel variations, of these regularities into the environment and observe what other regularities occur as a response

3. Bias one's pattern recognition algorithm to find new (maybe more complex) regularities, that are similar to the regularities that have already obtained responses

4. Then return to step 2 above with these new patterns, and also:

    a. If certain responses seemed particularly interesting for some reason, then try to project patterns that one infers will lead to these responses, based on inference from the response patterns one has observed.

    b. If one of the responses seems surprisingly similar to some pattern P one has projected, then respond to it with the response that P obtained, or some variant judged appropriate

Note that the “return to step 2” portion of Step 4 is essentially monologue, whereas the other portions of Step 4 are attempts at dialogue. Of course this sort of generic process is quite crude and is merely intended as a suggestion for

how to get started attempting to probe for intelligent agents in a very unfamiliar environment. If interesting patterns are discovered, then after that the interaction and discovery process is likely to go in unexpected directions.

For example, if one were interacting with a complex fluid environment that one suspected to contain intelligent agents, one might proceed as follows:

1. Use advanced AI pattern recognition software to recognize patterns of fluid flow. For instance, this might involve methods used to recognize statistical patterns in complex systems, such as Jim Crutchfield's "epsilon machines" approach which is able to recognize the statistical structure of complex, multi-lobed strange attractors (which often emerge in fluids that are in states lying near the boundary between laminar and turbulent flow). Create a library of the most significant patterns found. Then, generalize from these patterns to create a theory of what patterns are "fundamentally similar" to these significant patterns found.

2. Use jets projecting various forms of fluid, or other similar technology to create fluid flow patterns that, according to one's theory are fundamentally similar to the patterns recognized in the environment. Use one's pattern recognition software to identify how the environment is habitually responding to the stimuli one is giving it. Create a new library, consisting of those patterns that have obtained habitual responses from the environment, and also information on what those habitual responses are.

3. Use the pattern recognition software to search for additional patterns, focusing especially on searching for more complex patterns with similarity to the patterns one has found to generate responses, and to the patterns constituting these responses.

4. Proceed as in Step 4 above, with these new patterns, attempting to conduct an incrementally improving "fluidic dialogue" with the environment, and hoping that one can focus in on the more intelligent portions of the environment.

If there are intelligent curious agents in the environment and they observe you interacting with their environment in complex ways as the above process suggests, perhaps they will attempt to communicate in a manner related to your "fluid pattern projection" activities, which will then make the interaction process more interesting and complex than before the intelligent agents arrived!