Journal of Artificial General Intelligence 0 (2009) Submitted 2009-01-10

What Must a World Be That a Humanlike Intelligence May Develop In It?

Ben Goertzel BEN@GOERTZEL.ORG
1405 Bernerd Place
Rockville MD 20851, USA

Abstract

Among those who believe that richly embodied AGI is a promising path to creating AGI systems displaying human-level general intelligence, the possibility of virtual-world embodiment, as opposed to real-world robotic embodiment holds considerable appeal. Here we consider the question of what properties a virtual world should have in order to constitute an adequate environment for the cognitive development of a human-like, human-level general intelligence. We ask what properties a virtual world must have so that an AGI embodied in that world could viably infer humanlike theories of naive physics and folk psychology, and carry out tasks typically required in cognitive development tasks and preschool play centers. Based on these considerations we suggest a "minimal adequate environment" we call "BlocksNBeadsWorld," in which agents can construct objects from blocks using adhesives, and can also fill containers, coat objects and create fabrics and substances with various sorts of differentially adhesive beads.

Keywords: Artificial general intelligence, virtual worlds, cognitive development, physics simulation

1. Introduction

Many contemporary AI theorists believe that humanlike artificial general intelligence (AGI) will be most easily achieved via creating AGI learning systems, embodying them in roughly humanlike bodies, and interacting with these bodies in roughly humanlike environments. However, the quantification of "how roughly" is the subject of much debate even among those who agree on the general value of rich, realistic embodiment for AGI. Some feel that vaguely humanoid robots or even mobile wheeled robots are adequate to get one a long way toward AGI (Brooks, 2002); others suggest that a more humanlike sense of touch (Yohanan and MacLean, 2008) or kinesthetics is key; and others, such as myself, suspect that a less precisely faithful approach will suffice, such as embodiment of AGI systems in virtual characters in virtual worlds similar to multiplayer game worlds. My goal in this paper is not to rehash these familiar arguments; I will assume here, at least for sake of discussion, that rich embodiment is a valuable approach to use in creating and teaching AGI, and that virtual world technology is a worthwhile avenue to explore for the implementation of rich embodiment.

Even if one accepts the "AGI in virtual worlds" approach, however, there remains a large open question regarding the fidelity of the virtual world required. In other words, what does it really take to make a virtual world adequate as a "CogDevWorld" suitable for cognitive development of AGI systems? A precise physics simulation of the everyday human world is

beyond the scope of current science; but it seems unlikely (as I will argue below) that this is really required for AGI purposes. Yet current game worlds and virtual worlds like Second Life are clearly much too crude to suffice, as there are multiple critical human cognitive phenomena they don't easily support. What a CogDevWorld really requires is some middle ground between these two extremes, but it's not clear on the face of it exactly what this means. My goal here is to clarify this issue, by articulating a set of requirements that a virtual world must fulfill in order to serve the requirements of a developing AGI, and then describing a specific sort of world called BlocksNBeadsWorld that seems the minimum framework capable of satisfying the requirements.

2. The Value of Embodiment: A Learning Theory Perspective

The concept of intelligence is multifaceted (see Legg and Hutter (2006) for an inventory of numerous prior definitions), but one formulation that I have found useful is the one I articulated in *The Structure of Intelligence* (Goertzel, 1993): "the ability to achieve complex goals in complex environments," This formulation implies that pattern recognition is the key to achieving intelligence, based on a high-level algorithm such as

- Recognize patterns regarding which actions will achieve which goals in which situations
- Choose a goal that is expected to be good at goal achievement in the current situation

A subtle point is that this formulation implies some kind of averaging over the (potentially infinite) class of (goal, environment) pairs. If one is assessing the intelligence of a system as some sort of average of "the ability of system S to achieve goal G in environment E" over pairs (G, E), then weighting implicit in the average cannot be ignored – and turns out to be a conceptually critical entity.

A fully formalized definition of intelligence that is generally consistent with (though not precisely identical to) this basic idea is presented in (Legg and Hutter, 2006), and also lies at the core of Marcus Hutter's (2005) AIXI/AIXItl design and Juergen Schmidhuber's Godel machine (2006), all of which are essentially modern improvisations on the core idea of Solomonoff (1964, 1964a) induction.

How do you average over the space of goals and environments? If you average over all possible goals and environments, weighting each pair by its mathematically assessed complexity perhaps (so that success with simple goals/environments is rated higher, perhaps using Solomonoff induction related algorithmic information theory formulations to measure simplicity), then you have a definition of "how generally intelligent a system is," where general intelligence is defined in an extremely mathematically inclusive way. But it's not clear that this pure-mathematics type of approach is really all that relevant to the creation of humanlike AGI, or indeed the creation of useful AGIs in general. This question of how to define the average leads us to the topic of "everyday world AGI." Let's define the "everyday world" as the portion of the physical world that humans can directly perceive and interact with -- this is meant to exclude things like quantum tunneling, plasma dynamics in the centers of stars, etc.

My strong suspicion is that everyday-world general intelligence is not mainly about being able to recognize totally general patterns in totally general datasets (for instance, patterns among totally general goals and environments). I suspect that the best approach to this sort of totally general pattern recognition problem is ultimately going to be some variant of "exhaustive search through the space of all possible patterns" (which is what AIXItl does, for example) ... meaning

that approaching this sort of "truly general intelligence" is not really going to be a useful way to design an everyday-world AGI or a significant components of one.

Put differently, I suspect that all the AGI systems and subcomponents one can actually build in reality are so poor at solving the problem of being generally intelligent as implied by a simple pure-mathematics averaging, that it's better to characterize AGI systems, *not* in terms of how well they do at this general problem, but rather in terms of what classes of goals/ environments they are *really good* at recognizing patterns in.

It is key to recognize that the environments existing in the everyday physical and social world that humans inhabit are drawn from a pretty specific probability distribution (compared to say, the "universal prior," a standard probability distribution that assigns higher probability to entities describable using shorter programs; see e.g. Hutter (2005) for its use), and that for this reason, looking at problems of compression or pattern recognition across general goal/environment spaces without everyday-world-oriented biases, is not going to lead to everyday-world AGI.

The important parts of everyday-world AGI design are the ones that (directly or indirectly) reflect the specific distribution of problems that the everyday world presents an AGI system. And this distribution is really hard to encapsulate in a set of mathematical test functions. Because, we don't know what this distribution is. And this is a strong argument why we should be working on AGI systems that interact with the real everyday physical and social world, or the most accurate simulations of it we can build.

One could formulate this "everyday world" distribution, in principle, by taking the universal prior and conditioning it on a huge amount of real-world data. However, I suspect that simple, artificial exercises like conditioning distributions on text or photo databases don't come close to capturing the richness of statistical structure in the everyday world.

So, my contention is that

- the everyday world possesses a lot of special structure
- the human mind is structured to preferentially recognize pattern related to this special structure
- AGIs, to be successful in the everyday world, should be specially structured in this sort of way too

To encompass this everyday-world bias (or other similar biases) into the abstract mathematical theory of intelligence, we might say that intelligence relative to goal/environment class C is "the ability to achieve complex goals (in C) in complex environments (in C)". And we could formalize this by weighting each goal or environment by a product of

- its simplicity (e.g. measured by program length)
- its membership in C, considering C as a fuzzy etc

We could then then characterize a system's intelligence in terms of which goal/environment sets C it is reasonably intelligent for. In fact, this comes vaguely close to Pei Wang's (2008) definition of intelligence as "adaptation to the environment."

But, a key point to be noted is how much of human intelligence has to do, not with this general definition of intelligence, but with the subtle abstract particulars of the C that real human intelligences deal with (which equals the everyday world).

2.1 Some Properties of the Everyday World That Help Structure Intelligence

The properties of the everyday world that help structure intelligence are diverse and span multiple levels of abstraction. Most of this paper will focus on fairly concrete patterns of this nature, such as are involved in naive physics and folk psychology. However, it's also worth noting the potential importance of more abstract patterns distinguishing the everyday world from arbitrary mathematical environments.

The propensity to search for hierarchical patterns is one huge potential example of an abstract everyday-world property. I strongly suspect the fact that searching for hierarchical patterns works so well, in so many everyday-world contexts, is most likely because of the particular structure of the everyday world -- it's not something that would be true across all possible environments (even if one weights the space of possible environments using program-length according to some standard computational model). However, this sort of assertion is of course highly "philosophical," and difficult to defend convincingly given the current state of science and mathematics.

Going one step further, in my 1993 book *The Evolving Mind* (Goertzel, 1993) I identified a structure called the "dual network", which consists of superposed hierarchical and heterarchical networks: basically a hierarchy in which the distance between two nodes in the hierarchy is correlated with the distance between the nodes in some metric space. Another high level property of the everyday world may be that dual network structures are prevalent. This would imply that minds biased to represent the world in terms of dual network structure are likely to be intelligent with respect to the everyday world.

The extreme commonality of symmetry groups in the (everyday and otherwise) physical world is another example: they occur so often that minds oriented toward recognizing patterns involving symmetry groups are likely to be intelligent with respect to the real world.

I suggest that the number of properties of the everyday world of this nature is huge ... and that the essence of everyday-world intelligence lies in the list of varyingly abstract and concrete properties, which must be embedded implicitly or explicitly in the structure of a natural or artificial intelligence for that system to have everyday-world intelligence.

Apart from these particular yet abstract properties of the everyday world, intelligence is just about "finding patterns in which actions tend to achieve which goals in which situations" ... but, this simple meta-algorithm is, I conjecture, well less than 1% of what it takes to make a mind.

You might say that a sufficiently generally intelligent system should be able to infer these general properties from looking at data about the everyday world. Sure. But I suggest that would require a massively greater amount of processing power than an AGI that embodies and hence automatically utilizes these principles? It may be that the problem of inferring these properties is so hard as to require a wildly infeasible AIXItl / Godel Machine type system.

2.2 Important Open Questions

A few important questions raised by the above are as follows:

1. What is a reasonably complete inventory of the highly-intelligence-relevant subtle patterns/biases in the everyday world?
2. How different are the intelligence-relevant subtle patterns in the everyday world, versus the broader physical world (the quantum microworld, for example)?

3. How accurate a simulation of the everyday world do we need to have, to embody most of the subtle patterns that lie at the core of everyday-world intelligence?
4. Can we create practical progressions of simulations of the everyday world, such that the first (and more crude) simulations are very useful to early attempts at teaching proto-AGIs, and the development of progressively more sophisticated simulations roughly tracks the development of progress in AGI design and development.

Here I will not essay to explore all these questions, but will rather focus on the third one: how to make a simulation that encapsulates the most relevant everyday-world patterns? That is: how to make an adequate CogDevWorld? Addressing this issue requires some subtlety, because we don't really know what those patterns are. The approach I suggest is to attempt to simulate the everyday-world building blocks from which these patterns are made.

3. Naive Physics and Folk Psychology

In order to determine an adequate "requirements specification" for a CogDevWorld that gives rise to the key cognition-supportive patterns of the everyday world, I've turned to two relevant ideas from the AI literature, which have already been studied and discussed by many others: naive physics and folk psychology.

Naive physics (Hayes, 1985) refers to the theories about the physical world that human beings implicitly develop and utilize during their lives. For instance, when you figure out that you need to pressure the knife slightly harder when spreading peanut butter rather than jelly, you're not making this judgment using Newtonian physics or the Navier-Stokes equations of fluid dynamics; you're using heuristic patterns that you figured out through experience. Maybe you figured out these patterns through experience spreading peanut butter and jelly in particular. Or maybe you figured them out before you ever tried to spread peanut butter or jelly specifically, via just touching peanut butter and jelly to see what they feel like, and then carrying out inference based on your experience manipulating similar tools in the context of similar substances.

Other examples of similar "naive physics" patterns are easy to come by, e.g.

- What goes up must come down.
- A dropped object falls straight down.
- A vacuum sucks things towards it.
- Centrifugal force throws rotating things outwards.
- An object is either at rest or moving, in an absolute sense.
- Two events are simultaneous or they are not.
- When running downhill, one must lift one's knees up high
- When looking at something that you just barely can't discern accurately, squint

These sorts of heuristic patterns constitute "naive physics."

Attempts to axiomatically formulate naive physics have historically come up short, and I doubt this is a promising direction for AGI. However, I do think the naive physics literature does a good job of identifying the various phenomena that the human mind's naive physics deals with. So, from the point of view of CogDevWorld design, naive physics is a useful source of requirements. Ideally, we we would like CogDevWorld to support all the fundamental phenomena that naive physics deals with.

One important question is how close this support needs to stick to the particulars of real-world naive physics. Is it important that an AI in CogDevWorld can play with the specific differences between spreading peanut butter versus jelly? Or is it enough that it can play with spreading and smearing various substances of different consistencies? How close does the analogy between CogDevWorld naive physics and real-world naive physics need to be? This is a question to which we have no scientific answer at present; but, in order to design a particular CogDevWorld, some answer must be posited.

My own working assumption is that the analogy does not need to be extremely close, so in the following section I will propose a CogDevWorld (BlocksNBeadsWorld) that encompasses all the basic conceptual phenomena of real-world naive physics, but does not attempt to emulate their details. Part of my motivation for taking this direction is that it's a much more feasible path in the near term. It's not yet clear whether there's any way to make an extremely accurate simulation of real-world naive physics without first creating an underlying extremely accurate simulation of Newtonian physics, fluid dynamics, and so forth. And the latter sort of simulation is still at the research stage, and is the sort of problem whose subproblems occupy world-class supercomputers. So at present our only practical hope for a CogDevWorld is to make one whose naive physics corresponds roughly and conceptually to real-world naive physics.

Related, and somewhat coupled, to naive physics is naive psychology, which is more typically called "folk psychology." Folk psychology is the set of informal theories people have about other peoples' minds. There is a strong intersection between folk psychology and naive physics, because people often reason about inanimate objects via anthropomorphizing them and then applying folk psychology. An important requirement on any CogDevWorld is that its representation of intelligent agents must be rich enough to support the full spectrum of folk psychology.

My suggestion is that, if we create a simulation world capable of roughly supporting naive physics and folk psychology, then we are likely to have a simulation world that gives rise to the key inductive biases provided by the everyday world for the guidance of humanlike intelligence.

3.1 Requirements for CogDevWorld From Naive Physics

Naive physics has many different formulations; in this section I draw heavily on Smith and Casati (1994), which explicitly provides an ontology of naive physics ideas, from which it is relatively straightforward to limn a list of required naive physics phenomena that any CogDevWorld should support if it is to effectively foster closely humanlike cognitive development. Smith and Casati divide naive physics phenomena into 5 categories; I now review these categories and identify a number of important things that intelligent agents must be able to do relative to each of them.

3.1.1 Objects, Natural Units and Natural Kinds

One key aspect of naive physics involves recognition of various aspects of objects. This is an area where current virtual world technology is relatively strong, yet not quite strong enough, e.g. it doesn't handle breaking and fusing of objects well. Specific aspects of naive physics related to objects 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

3.1.2 Events, Processes and Causality

Recognizing properties of events in time is an aspect of naive physics that doesn't impose too many special requirements on a virtual world; events in a virtual world are immediately time-stamped. Specific aspects of naive physics related to temporality and causality are:

- 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

Mainly, what is required of a virtual world in order to allow these sorts of naive physics is a variety of different processes occurring on a variety of different time scales, intersecting in complex patterns, and involving processes inside the agent's body, outside the agent's body, and crossing the boundary of the agent's body.

3.1.3 Stuffs, States of Matter, Qualities

An area where current virtual world technology falls far short is the presentation of a diversity of states of matter. Virtual worlds today are basically about rigid objects, whereas objects in the real world stretch, fold, have bumps and sticky surfaces, etc. These various properties of objects commonly appear as the foundation of linguistic metaphors ("a sticky situation", "a bit of a stretch for him", etc.) and cognitive metaphors as well. There are also various phenomena like rainbows and mirages that have powerful analogical utilizations (for instance, to an AGI that's never seen a mirage or anything like it, the notion that "the world is an illusion" will never have the same depth as it does to a human). Along these lines, some important aspects of naive physics are:

- Perceiving gaps between objects: holes, media, illusions like rainbows, mirages and holograms
- Distinguishing the manners in which different sorts of entities (e.g. smells, sounds, light) fill space
- Distinguishing properties such as smoothness, roughness, graininess, stickiness, runniness, etc.
- Distinguishing degrees of elasticity and fragility

- Assessing separability of aggregates

3.1.4 Surfaces, Limits, Boundaries, Media

Gibson (1977, 1979, 1982) has argued that naive physics is not mainly about objects but rather mainly about surfaces. Surfaces have a variety of aspects and relationships that are important for naive physics, such as:

- Perceiving and reasoning about surfaces as two-sided or one-sided interfaces
- Inference of the various ecological laws of surfaces
- Perception of various media in the world as separated by surfaces
- Recognition of the textures of surfaces
- Recognition of medium/surface layout relationships such as: ground, open environment, enclosure, detached object, attached object, hollow object, place, sheet, fissure, stick, fibre, dihedral, etc.

Figure 1. One of Sloman's example test domains for real-world inference. Left: a number of pins and a rubber band to be stretched around them. Right: use of the pins and rubber band to make a letter T.

As a concrete, evocative "toy" example of naive everyday knowledge about surfaces and boundaries, consider Sloman's (2008) example scenario, depicted in Figure 1 and drawn largely from (Sauvy and Sauvy, 1974) (see also related discussion in Sloman, 2008a), in which "A child can be given one or more rubber bands and a pile of pins, and asked to use the pins to hold the band in place to form a particular shape.... For example, things to be learnt could include:

- There is an area inside the band and an area outside the band
- The possible effects of moving a pin that is inside the band towards or further away from other pins inside the band. (The effects can depend on whether the band is already stretched.)
- The possible effects of moving a pin that is outside the band towards or further away from other pins inside the band.
- The possible effects of adding a new pin, inside or outside the band, with or without pushing the band sideways with the pin first.
- The possible effects of removing a pin, from a position inside or outside the band.
- Patterns of motion/change that can occur and how they affect local and global shape (e.g. introducing a concavity or convexity, introducing or removing symmetry, increasing or decreasing the area enclosed).

- The possibility of causing the band to cross over itself. (NB: Is an odd number of crosses possible?)
- How adding a second, or third band can enrich the space of structures, processes and effects of processes.

3.1.5 Motivation, Requiredness, Value

Gestalt (Kohler, 1938) and ecological (Gibson, 1977, 1979, 1982) psychology suggest that humans perceive the world substantially in terms of the affordances it provides them for goal-directed action. This means that a CogDevWorld should provide:

- Perception of entities in the world as differentially associated with goal-relevant value
- Perception of entities in the world in terms of the potential actions they afford the agent, or other agents

The key point is that entities in the world need to provide a wide variety of ways for agents to interact with them, enabling richly complex perception of affordances.

3.2 Requirements for CogDevWorld From Folk Psychology

Finally, the following are aspects of folk psychology that should be enabled within any CogDevWorld:

- Mental simulation of other agents
- Mental theory regarding other agents
- Attribution of beliefs, desires and intentions (BDI) to other agents via theory or simulation
- Recognition of emotions in other agents via their physical embodiment
- Recognition of desires and intentions in other agents via their physical embodiment
- Analogical and contextual inferences between self and other, regarding BDI and other aspects
- Attribute causes and meanings to other agents behaviors
- Anthropomorphize non-human, including inanimate objects

The main special requirement placed on a CogDevWorld by the above aspects pertains to the ability of agents to express their emotions and intentions to each other. Humans do this via facial expressions and gestures, both of which are typically impoverished in contemporary games and virtual worlds.

3.3 Requirements for Bodies in CogDevWorld

The above points have focused on the world external to the body of the AGI agent embodied and embedded in the world, but the issue of the AGIs body also merits consideration. Here the requirements seem fairly simple: while not strictly necessary, it would seem strongly preferable to provide the AGI with fairly rich analogues of the human senses of touch, sight,

sound, kinesthesia, taste and smell. Each of these senses provides different sorts of cognitive stimulation to the human mind; and while similar cognitive stimulation could doubtless be achieved without analogous senses, the provision of such seems the most straightforward approach.

As vision already is accorded such a prominent role in the AI and cognitive science literature, I won't take time elaborating on the importance of vision processing for humanlike cognition. The key point for CogDevWorld is the support of a sufficiently robust collection of materials that object recognition and identification become interesting problems. A virtual world in which there is only a small fixed fund of object types or shapes will not likely do, nor will a world in which objects can't stick together and then separate depending on context.

Audition is valuable for many reasons, one of which is that it gives a very rich and precise method of sensing the world that is different from vision. The fact that humans can display normal intelligence while totally blind or totally deaf is an indication that, in a sense, vision and audition are redundant for understanding the everyday world. However, it may be important that the brain has evolved to account for both of these senses, because this forced it to account for the presence of two very rich and precise methods of sensing the world – which may have forced it to develop more abstract representation mechanisms than would have been necessary with only one such method. At any rate, exact simulation of complex real-world acoustics seems unnecessary for a CogDevWorld, but a crude approximation would seem valuable, including aspects such as sound intensity decaying with distance, individual sounds being difficult to distinguish amidst a general clamor, etc.

Touch is a sense that is, in my view, generally badly underappreciated within the AI community. In particular the cognitive robotics community seems to worry too little about the terribly impoverished sense of touch possessed by most current robots (though fortunately there are recent technologies that may help improve robots in this regard; see Nanowerk (2008)). Touch is how the human infant learns to distinguish self from other, and in this way it is the most essential sense for the establishment of an internal self-model. Touching others' bodies is a key method for developing a sense of the emotional reality and responsiveness of others, and is hence key to the development of theory of mind and social understanding in humans. For this reason, among others, human children lacking sufficient tactile stimulation will generally wind up badly impaired in multiple ways. A CogDev world should supply an AI agent with a body that possesses skin, which has varying levels of sensitivity on different parts of the skin (so that it can effectively distinguish between reality and its perception thereof in a tactile context); and also varying types of touch sensors (e.g. temperature versus friction), so that it experiences textures as multidimensional entities.

Related to touch, kinesthesia refers to direct sensation of phenomena happening inside the body. Rarely mentioned in AI, this sense seems quite critical to cognition, as it underpins many of the analogies between self and other that guide cognition. Again, it's not important that an AGI's virtual body have the same internal body parts as a human body. But it seems valuable to have the AGI's virtual body display some vaguely human-body-like properties, such as feeling internal strain of various sorts after getting exercise, feeling discomfort in certain places when running out of energy, feeling internally different when satisfied versus unsatisfied, etc.

Taste is a cognitively interesting sense in that it involves the interplay between the internal and external world; it involves the evaluation of which entities from the external world are worthy of placing inside the body. And smell is cognitively interesting in large part because of its relationship with taste. A smell is, among other things, a long-distance indicator of what a certain entity might taste like. So, the combination of taste and smell provides means for conceptualizing relationships between self, world and distance. What seems to be valuable for a CogDevWorld is that different entities have multidimensional tastes and smells, and that there be

correlations between these. Simulation of the precise details of human taste and smell is almost surely cognitively irrelevant.

3.4 Are These Requirements Adequate?

It is difficult to know if any such list of requirements is sufficient. There are always more and more phenomena one could cite. However, my qualitative argument for the sufficiency of the requirements list is simple: in a CogDevWorld satisfying the above requirements,

- one could carry out all the standard cognitive development experiments described in developmental psychology books (Piaget, 1955; Shultz, 2003)
- one could implement intuitively reasonable versions of all the standard activities in all the standard learning stations in a contemporary preschool (see (Goertzel and Bugaj, 2008) for a review of preschool design from an AI/virtual-worlds perspective)

Typical preschool activities include for instance building with blocks, playing with clay, looking in a group at a picture book and hearing it read aloud, mixing ingredients together, rolling/throwing/catching balls, playing games like tag, hide-and-seek, Simon Says or Follow the Leader, measuring objects, cutting paper into different shapes, drawing and coloring, etc.

As typical, not necessarily representative examples of tasks psychologists use to measure cognitive development (drawn mainly from the Piagetan tradition, without implying any assertion that this is the only tradition worth pursuing), consider the following:

- Which row has more circles- A or B? A: O O O O O, B: OOOOO
- If Mike is taller than Jim, and Jim is shorter than Dan, then who is the shortest? Who is the tallest?
- Which is heavier- a pound of feathers or a pound of rocks?
- Eight ounces of water is poured into a glass that looks like the fat glass in Figure 2 and then the same amount is poured into a glass that looks like the tall glass in Figure 2 (below). Which glass has more water?
- A lump of clay is rolled into a snake. All the clay is used to make the snake. Which has more clay in it -- the lump or the snake?
- There are two dolls in a room, Sally and Ann, each of which has her own box, with a marble hidden inside. Sally goes out for a minute, leaving her box behind; and Ann decides to play a trick on Sally: she opens Sally's box, removes the marble, hiding it in her own box. Sally returns, unaware of what happened. Where will Sally would look for her marble?
- Consider this rule about a set of cards that have letters on one side and numbers on the other: "If a card has a vowel on one side, then it has an even number on the other side." If you have 4 cards labeled "E K 4 7", which cards do you need to turn over to tell if this rule is actually true?
- Design an experiment to figure out how to make a pendulum that swings more slowly versus less slowly

Of course, this "argument via preschools and cognitive development tests" doesn't prove anything definitively, but it does seem highly suggestive. It is indeed possible that the standard psych experiments don't dig deep enough, and that some of the "intuitively reasonable versions"

of preschool activities satisfying the above requirements might unintentionally miss the really cognitively critical aspects of the corresponding real-world preschool activities. But, I consider these possibilities fairly unlikely; and in carrying out the sort of design process we are involved in here, one must inevitably rely on intuition to a certain extent.

Figure 2. Example of Piagetan "conservation of volume" task used to assess child cognitive development. In the BlocksNBeadsWorld context, the cups of milk would be replace by cups of beads. See video at http://www.youtube.com/watch?v=qYtNhNP69lk&feature=related

4. BlocksNBeadsWorld

In this section I will briefly describe a simple virtual world approach that appears to fulfill the above requirements, without requiring anywhere near a complete simulation of realistic physics.

The class of worlds I propose is called BlocksNBeadsWorld, and consists of the following aspects:

- 3D blocks of various shapes and sizes and frictional coefficients, that can be stacked
- Adhesive that can be used to stick blocks together, and that comes in two types, one of which can be removed by an adhesive-removing substance, one of which cannot (though its bonds can be broken via sufficient application of force)
- Spherical beads, each of which has intrinsic unchangeable adhesion properties defined according to a particular, simple "adhesion logic"
- Each block, and each bead, may be associated with multidimensional quantities representing its taste and smell; and may be associated with a set of sounds that are made when it is impacted with various forces at various positions on its surface

Interaction betwen blocks and beads would be calculated according to standard Newtonian physics, which would be compute-intensive in the case of a large number of beads, but tractable using distributed processing. For instance if 10K beads were used to cover a humanoid agent's face, this would provide a fairly wide diversity of facial expressions; and if 10K beads were used to form a blanket laid on a bed, this would provide a significant amount of flexibility in terms of rippling, folding and so forth. Yet, this order of magnitude of interactions is very small compared to what is done in contemporary simulations of fluid dynamics or, say, quantum chromodynamics.

One key aspect of the spherical beads is that they can be used to create a variety of rigid or flexible surfaces, which may exist on their own or be attached to blocks-based constructs. The specific inter-bead adhesion properties of the beads could be defined in various ways, and will surely need to be refined via experimentation, but a simple scheme that seems to make sense is as follows.

Each bead can have its surface tesselated into hexagons (the number of these can be tuned), and within each hexagon it can have two different adhesion coefficients: one for adhesion to other beads, and one for adhesion to blocks. The adhesion between two beads along a certain hexagon is then determined by their two adhesion coefficients; and the adhesion between a bead and a block is determined by the adhesion coefficient of the bead, and the adhesion coefficient of the adhesive applied to the block. A distinction must be drawn between rigid and flexible adhesion: rigid adhesion sticks a bead to something in a way that can't be removed except via breaking it off; whereas flexible adhesion just keeps a bead very close to the thing it's stuck onto. Any two entities may be stuck together either rigidly or flexibly. Sets of beads with flexible adhesion to each other can be used to make entities like strings, blankets or clothes.

Using the above adhesion logic, it seems one could build a wide variety of flexible structures using beads, such as (to give a very partial list):

- fabrics with various textures, that can be draped over blocks structures,

- multilayered coatings to be attached to blocks structures, serving (among many other examples) as facial expressions
- liquid-type substances with varying viscosities, that can be poured between different containers, spilled, spread, etc.
- strings tyable in knots; rubber bands that can be stretched; etc.

Of course there are various additional features one could add. For instance one could add a special set of rules for vibrating strings, allowing BlocksNBeadsWorld to incorporate the creation of primitive musical instruments. Variations like this could be helpful but aren't necessary for the world to serve its essential purpose.

Note that one does not have true fluid dynamics in BlocksNBeadsWorld, but, it seems that the latter is not necessary to encompass the phenomena covered in cognitive developmental tests or preschool tasks. The tests and tasks that are done with fluids can instead be done with masses of beads. For example, consider the conservation of volume task shown in Figure 2 below: it's easy enough to envision this being done with beads rather than milk. Even a few hundred beads is enough to be psychologically perceived as a mass rather than a set of discrete units, and to be manipulated and analyzed as such. And the simplification of not requiring fluid mechanics in one's virtual world is immense.

Next, one can implement equations via which the adhesion coefficients of a sphere are determined in part by the adhesion coefficients of nearby spheres, or spheres that are nearby in certain directions (with direction calculated in local spherical coordinates). This will allow for complex cracking and bending behaviors – not identical to those in the real world, but with similar qualitative characteristics. For example, without this feature one could create paperlike substances that could be cut with scissors – but *with* this feature, one could go further and create woodlike substances that would crack when nails were hammered into them in certain ways, and so forth.

And finally, the combination of blocks and beads seems ideal for implementing a more flexible and AGI-friendly type of virtual body than is currently used in games and virtual worlds. One can easily envision implementing a body with

- a skeleton whose bones consist of appropriately shaped blocks
- joints consisting of spheres, flexibly adhered to the bones
- flesh consisting of spheres, flexibly adhered to each other
- internal "plumbing" consisting of tubes whose walls are spheres rigidly adhered to each other, and flexibly adhered to the surrounding flesh (the plumbing could then serve to pass spheres through, where slow passage would be ensured by weak adhesion between the walls of the tubes and the spheres passing through the tubes)

This sort of body would support rich kinesthesia; and rich, broad analogy-drawing between the internally-experienced body and the externally-experienced world. It would also afford many interesting opportunities for flexible movement control.

The philosophy underlying these suggested bead dynamics is somewhat comparable to that outlined in Wolfram's (2002) book *A New Kind of Science*. There he proposes cellular automata models that emulate the qualitative characteristics of various real-world phenomena, without trying to match real-world data precisely. For instance, some of his cellular automata demonstrate phenomena very similar to turbulent fluid flow, without implementing the Navier-Stokes equations of fluid dynamics or trying to precisely match data from real-world turbulence. Similarly, the beads in BlocksNBeadsWorld are intended to qualitatively demonstrate the real-

world phenomena most useful for the development of humanlike embodied intelligence, without trying to precisely emulate the real-world versions of these phenomena.

Without the beads, BlocksNBeadsWorld would appear purely as a "Blocks World with Glue" – essentially a substantially upgraded version of the Blocks Worlds frequently used in AI, since first introduced in (Winograd, 1972). Certainly a pure "Blocks World with Glue" would have greater simplicity than BlocksNBeadsWorld, and greater richness than standard Blocks World; but this simplicity comes with too many limitations, as shown by consideration of the various naive physics requirements inventoried above. One simply cannot run the full spectrum of humanlike cognitive development experiments, or preschool educational tasks, using blocks and glue alone. One can try to create analogous tasks using only blocks and glue, but this quickly becomes extremely awkward. Whereas in the BlocksNBeadsWorld the capability for this full spectrum of experiments and tasks seems to fall out quite naturally.

5. Conclusion

I have argued that for the proper development of humanlike intelligence it is important to provide an environment containing the various subtle patterns in the everyday human world that provide the human mind with its inductive biases. However, since we don't know exactly what these patterns and biases are, the best approach seems to be to turn to naive physics and folk psychology to gain a qualitative understanding of the sorts of phenomena in which they lie. Based on this motivation, I have articulated a set of requirements that any CogDevWorld must fulfill, in order to provide an educational environment for AIs that roughly emulates the primary naive physics and folk psychology phenomena that humans encounter in the real world. I hasten to add that I am not claiming these requirements are *necessary* in order for a CogDevWorld to support the development of a human-level, roughly human-like AGI system. In fact, I suspect they constitute overkill. However, at this stage it is difficult to be confident exactly *which aspects* are really necessary.

One obvious question is whether something like BlocksNBeadsWorld is really close enough to real-world naive physics and folk psychology to have significant advantages over a simpler virtual world closer to current worlds like Second Life or OpenSim. However, I think the answer is relatively clear, and my argument was made above already: in BlocksNBeadsWorld, unlike current virtual worlds, one can do nearly every sort of task done in a human preschool, and one can run nearly every sort of psychological test done by cognitive developmental psychologists. I think this provides a strong qualitative argument that there is some sort of fundamental adequacy in the BlocksNBeadsWorld approach.

Consider, for instance, three scenarios:

1. A CogDevWorld containing realistic fluid dynamics, where a child can pour water back and forth between two cups of different shapes and sizes, to understand issues such as conservation of volume
2. A CogDevWorld more like today's Second Life, where fluids don't really exist, and things like lakes are simulated via very simple rules, and pouring stuff back and forth between cups doesn't happen unless it's programmed into the cups in a very specialized way
3. A BlocksNBeadsWorld type CogDevWorld, where a child can pour masses of beads back and forth between cups, but not masses of liquid

My qualitative judgment is that Scenario 3 is going to allow a young AI to gain the same essential insights as Scenario 1, whereas Scenario 2 is just too impoverished. I have explored dozens of similar scenarios regarding different preschool tasks or cognitive development experiments, and come to similar conclusions across the board. Thus, my current view is that something like BlocksNBeadsWorld can serve as an adequate CogDevWorld, supporting the development of human-level, roughly human-like AGI.

And, if this view turns out to be incorrect, and BlocksNBeadsWorld is revealed as inadequate, then I will very likely still advocate the conceptual approach enunciated above as a guide for designing CogDevWorlds. That is, I would suggest to explore the hypothetical failure of BlocksNBeadsWorld via asking two questions:

- Are there basic naive physics or folk psychology requirements that were missed in creating the specifications, based on which the adequacy of BlocksNBeadsWorld was assessed?
- Does BlocksNBeadsWorld fail to sufficiently emulate the real world in respect to some of the articulated naive physics or folk psychology requirements?

The answers to these questions would guide the improvement of the world or the design of a better one.

Regarding the practical implementation of BlocksNBeadsWorld, it seems clear that this is within the scope of modern game engine technology, however, it is not something that could be encompassed within an existing game engine without significant additions; it would require substantial custom game engine engineering. There exist commodity and open-source physics engines that efficiently carry out Newtonian mechanics calculations; while they might require some tuning and extension to handle BlocksBeadWorld, the main issue would be achieving adequate speed of physics calculation, which given current technology would need to be done via modifying existing engines to appropriately distribute processing among multiple GPUs.

BlocksNBeadsWorld could be used to build many different sorts of environments for developmental AI systems, but the avenue that interests me the most is that of an "AGI Preschool" as described in (Goertzel and Bugaj, 2008). It seems to me that a BlocksNBeadsWorld foundation would be the easiest clearly-adequate approach to developing a virtual-world preschool for young AGI systems, and this is an avenue that interests me in the extreme.

References

Brooks, R. A., Flesh and Machines, Pantheon Books, New York, NY, 2002

Gibson, J.J. 1977. The theory of affordances. In R. Shaw & J. Bransford (eds.), Perceiving, Acting and Knowing. Hillsdale, NJ: Erlbaum.

Gibson, J.J. 1979. The Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

Goertzel, Ben. 1993. The Evolving Mind. Gordon and Breach.

Goertzel, Ben and Stephan Bugaj. 2008. AGI Preschool. Proceedings of the Second Conference on Artificial General Intelligence, Atlantis Press.

Hayes, Patrick. 1985. "The Second Naive Physics Manifesto," in Jerry Hobbs and Robert Moore (Eds): Formal Theories of the Commonsense World, Ablex

Hutter, Marcus (2005). Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability. Springer, 2005

Kohler, Wolfgang. 1938. The Place of Value in a World of Facts. Liveright Press, New York.

Legg, S. and Marcus Hutter. (2006). A Formal Measure of Machine Intelligence. In Proc. Annual machine learning conference of Belgium and The Netherlands (Benelearn-2006). Ghent, 2006.

Legg, S. and Marcus Hutter. A Collection of Definitions of Intelligence. In B. Goertzel, editor, Advances in Artificial General Intelligence, IOS Press, 2006.

Nanowerk. 2008. Carbon nanotube rubber could provide e-skin for robots. http://www.nanowerk.com/news/newsid=6717.php, downloaded 1-10-09

Piaget, Jean. The Construction of Reality in the Child. Routledge and Kegan Paul.

Reed, Edward and Rebecca Jones. 1982. "Reasons for realism: selected essays of James J. Gibson." Hillsdale, NJ: Erlbaum, 1982. (pp.431-437)

Sattler, J.M. 2001 . Assessment of Children: Cognitive Applications. Jerome M. Sattler Publisher, San Diego.

Sauvy, Jean and Simonne Suavy. 1974. The Child's Discovery of Space: From hopscotch to mazes – an introduction to intuitive topology. Penguin

Schmidhuber, J. (2006). Gödel machines: Fully Self-Referential Optimal Universal Self-Improvers. In B. Goertzel and C. Pennachin, eds.: Artificial General Intelligence, p. 119-226, 2006.

Shultz, Thomas R. 2003. Computational Developmental Psychology. MIT Press.

Sloman, Aaron. 2008. A New Approach to Philosophy of Mathematics: Design a young explorer, able to discover "toddler theorems." Invited talk at University of Sussex, Tuesday 9th December 2008

Sloman, Aaron. 2008a. The Well-Designed Young Mathematician. Artificial Intelligence, December 2008

Smith, Barry and Roberto Casati. 1994. Naive Physics: An Essay in Ontology. Philosophical Psychology, 7/2, 225-244.

Solomonoff, Ray. 1964. "A Formal Theory of Inductive Inference, Part I" Information and Control, Vol 7, No. 1 pp 1-22, March 1964.

Solomonoff, Ray. 1964. "A Formal Theory of Inductive Inference, Part II" Information and Control, Vol 7, No. 2 pp 224-254, June 1964

Understanding Natural Language by T. Winograd, Academic Press, 1972

Wang, Pei. 2008. What Do You Mean by "AI"? Proceedings of The First Conference on Artificial General Intelligence, IOS Press

Winograd, Terry. 1972. Understanding Natural Language. Edinburgh University Press.

Wolfram, Stephen. 2002. A New Kind of Science. Wolfram Media, Inc.

Yohanan, S. and Karon E. MacLean. The Haptic Creature Project: Social Human-Robot Interaction through Affective Touch. In Proceedings of the AISB 2008 Symposium on the Reign of Catz & Dogs: The Second AISB Symposium on the Role of Virtual Creatures in a Computerised Society, volume 1, pages 7-11, Aberdeen, Scotland, UK, April, 2008.