THE WELL-DESIGNED CHILD

John McCarthy
Stanford University
jmc@cs.stanford.edu
http://www-formal.stanford.edu/jmc/
September 28, 2007

- The world into which a human is born is complex. We'll discuss some of the complexities, concentrating on one— 3-d semi-permanent, movable objects.

- Evolution has equipped a baby with some innate knowledge of the world. That works better than starting as a blank slate.

- Consider a well-designed logical robot child (WDC). Mostly it will have the innate abilities we conjecture real children have, but sometimes we can do better.

THE LOCKEAN BABY

• John Locke 1693: The baby starts out as a blank slate. It builds its knowledge by inferring associations among stimuli.

• Since 1950 people proposed to start with a Lockean blank slate baby machine and have it learn from its experience.

• Starting from the blank slate, I'll bet it's a lengthy process to infer the existence of 3-d objects. I don't think anyone has made an AI system that can do it.

• A billion years of evolution has provided us with pretty good prejudices about the world far better than a blank slate. Our robot child should also have them

THE WORLD IS COMPLICATED

• The world's structure is not directly describable in terms of the input-output relations of a person. The basic structure of the world involves elementary particles on time scales of 10^{-25} seconds, but intelligence only evolved in structures of more than 10^{23} elementary particles.

• Even at the level at which a small child can perceive it, the world is extremely complicated. Here are some of the complications.

• Reality and appearance Animals and humans don't perceive the structure of our environment directly. Senses have evolved to give *partial* information about objects and their relations.

• The world is 3-d, but our senses react to surfaces.

MORE COMPLICATIONS

- semi-permanent objects Much of the world consists of three-dimensional objects that have masses, moments, compliances, hardnesses, chemical composition, shapes, outer surfaces with textures and colors, are often made of identifiable parts which sometimes move relative to each other. A particular object can disappear from perception and reappear again.

Note that the structure and location of an object in the world are more persistent than its appearance and location in the visual field or relative to the hands.

- temporal structure The environment of a child has a complex temporal structure. Some items change in seconds, others last for hours, days, or years. Babies live in the present. The concepts of *tomorrow* and *yesterday* are not learned right away.

- causality Events cause changes in objects and their relations and cause other events.

STILL MORE COMPLICATIONS

- solidity Objects are solid and do not ordinarily penetrate one another. Some are rigid and some are flexible.

- gravity Unsupported objects fall to a lower surface.

- kinds of objects Objects have kinds, and objects of the same kind have properties associated with the kind. Babies are ready very early to learn what kinds there are.

- natural kinds Many of the objects a child encounters, e.g. lemons, belong to *natural kinds*. The objects of a natural kind have yet undiscovered properties in common. Therefore, a natural kind is not definable by an *if-and-only-if* sentence formulated in terms of observables.

- relations Objects not only have individual properties and belong to kinds, but objects and kinds have relations with one another. At least some ternary relations such as betweenness are basic. Also "A is to B as C is to D" seems to be basic. In its numerical use, it reduces to the equality of two fractions, but the quaternary relation seems to be basic in common sense usage.

WHAT DO BABIES KNOW AND WHEN?

• Q: If the world is so complicated, how can babies do anything purposeful?

• A: They know simple cases of phenomena.

• There's good psychological evidence that they have an innate knowledge of solid objects that continue existing even when out of sight. How do they remember an object that has gone out of sight so as to recognize it when it reappears?

• They are ready to ascribe purposeful action to humans and animals and to try to influence them.

- the principle of mediocrity (from the astronomers) Our child is like other children. This lets it reason in both directions.

- persons Some objects are animate and have purposes analogous to those of the child. They can be influenced but are sometimes to be feared. Defects in innate understanding of persons, e.g. autism, harm the child.

WHAT DO WE WANT IN A WELL-DESIGNED ROBOT CHILD?

- Distinguish appearance from reality

- natural kinds Natural kinds don't have if-and-only-if definitions. There may always be more properties to be learned. To a small child, all kinds are natural. The robot child should think in terms of natural kinds.

- three-dimensional objects These are more stable than the perception of them by any sense.

- perceive motion as continuous

- actions and their effects

- recognize parts Recognize parts of an object and their relations to the others. It would be interesting to have a grammar of 3-d physical structure analogous to that of sentences.

- focussed curiosity

- grammar of goal regression To do A, I need to do B first, and to do B, I need to do C first.

- introspection Children begin to do this by age 3, do it well by age 5. The WDC needs it.

THE SPELKE EXPERIMENT

This psychological experiment exhibits abilities of human babies we want in the WDC.

Elizabeth Spelke described a number of experiments that she and others did to discover and verify innate mental abilities. The technique uses the fact that a baby will look longer at something surprising than at something that seems familiar.

Here's one that was first done in 1973 and was repeated by Spelke in 1993 with two months old babies. There are experimental babies and control babies and the experiment has two phases. In the first phase the control babies are shown nothing. The experimental babies see an object go behind a screen and shortly another object emerges on the other side of the screen. The timing is such as would be appropriate if the first object struck the second object and knocked it from behind the screen. The babies are shown the phenomenon enough times to get bored with it and stop paying attention.

In the second phase of the experiment the screen is removed. There are two variants. In the first variant, the first object strikes the second and knocks it onward. In the second variant the first object stops short of the second, but the second object takes off as though it had been struck. The control babies look at both variants for the same amount of time. The experimental babies look longer at the second variant.

The conclusion is that the experimental babies inferred that the first object had struck the second when the event occurred behind the screen. When the screen was removed, they were not surprised when the expected event was shown to occur but were surprised and looked longer when this expectation was not met.

The conclusion is that babies have innate expectations about dynamics. For details see Spelke's 1994 article in *Cognition*, *Initial knowlege: six suggestions*.

THE WELL-DESIGNED CHILD AS A LOGICAL ROBOT

- Not even a sketch of a design—just some ideas.

- Appearance and reality

$Appears(appearance, object)$ is too simple except in a limited context, but children think in limited contexts. $Holds(Appears(person, appearance, object), s)$ says more and is suitable for referring to the child from the outside.

CONTEXTS

• The correctness of a child's beliefs and references to objects depends on context. Children change context frequently.

• When looking at a child's ideas from the outside, we can use the theory of contexts as objects introduced in my *Notes on formalizing context*. In that theory $Ist(c, p)$ is true if the proposition p holds in the context c. $Value(c, exp)$ is the value of the expression exp, e.g. $Color(Block1)$, in context c.

• $Value(Beliefs(Child1, S0), Color(Block1)) = Red$ asserts that in the context of Child1's beliefs in situation s, Block1 is red. The context theory lets us *enter* the context $Beliefs(Child1, S0)$. Then we have $Color(Block1) = Red$.

THREE CHALLENGES AND A CONCLUSION

- Make a robot baby that can do the Spelke experiment honestly. But what's honestly? It may start knowing about collisions and knowing about occluded objects, but at least it shouldn't start with knowledge combining the two.

- Its contexts should change like those of a baby.

- It should properly relate 2-d appearances to 3-d reality and also relate tactile appearances to reality. Here's a puzzle.

- What babies know presents challenges to AI—to AI of all schools.

• This talk is partly based on the manuscript http://www-formal.stanford.edu/jmc/child.html.