Utilizing Child IQ Tests to Measure Robot Intelligence

Ben Goertzel

Feb 13, 2013

This document informally discusses some issues involved in using the WPPSI, the standard IQ testing instrument for preschool children, to assess the general intelligence of robots.

Basic Assumptions

I will assume, here, use of the exact same WPPSI questions administered to human children. Of course, there are obvious ways one could thoroughly adapt the WPPSI questions for robots, but that becomes a different sort of exercise than the one undertaken here. My concern here will be: How to make administration of the actual human-oriented WPPSI, a meaningful evaluation of robot intelligence.

The only exception I will suggest to the preceding paragraph, is to ignore questions relating specifically to the test subject's own personal body or experience, in ways that specifically pertain to the differences between human and robot. For instance, we shouldn't ask "touch your hair" to a robot without hair. "Touch your head" would be fine. Such questions play a very minor role in WPPSI, occurring occasionally and incidentally among other questions.

In terms of the physical test set-up, it seems fair to assume a robot sitting across a table from the human examiner. To emulate human testing conditions accurately, the robot will need ears to hear the examiner speak, eyes to see the pictures and objects displayed and the physical environment, a voice to answer questions, and hands or some other manipulators to move around blocks and puzzle pieces.

Optionally, it would seem unproblematic to assume that the robot has textual rather than voice communication with the examiner, as the test questions don't explicitly involve auditory pattern recognition, only visual. Given a robot that communicates via text and achieves a certain WPPSI score, one could automatically produce a robot communicating via voice and achieving the same score, by simply adding on a speech synthesizer and an effective speech-to-text engine.

One significant issue that becomes apparent when thinking about giving the WPPSI to robots, is that many WPPSI questions involve objects that are part of a typical human child's commonsense experience, and are not going to be part of a typical contemporary robot's experience. However, this aspect of the WPPSI could not be "fixed" without radically altering the nature of the test. The WPPSI is, fundamentally a mix of simple visual puzzle-solving type questions, and

simple commonsense knowledge type questions. To do well on the WPPSI, a robot will have to be able to identify pictures of objects common in a human child's life, and common in the storybooks commonly read to young children (e.g. most young children can recognize a picture of a pig, even if they've grown up in a city and never seen an actual pig). It will have to be able to identify the parts of these objects. It will also have to know basic everyday facts that a normal child knows – that airplanes fly, that cars drive faster than people walk, and so forth. As with a human child, some of these facts may be grounded in the robot's life experience, whereas other may be known to it only indirectly.

The Issues of Coaching and Specialized Engineering

It is important to note that, insofar as the WPPSI works for measuring the IQ of human children, it works ONLY if the children tested haven't been coached in the specific question types. This is the reason that WPPSI questions are not widely disseminated to parents of preschool age children (though they are available via various study guides, which the diligent parent can locate online). If a human child is coached in the particular sorts of questions given on the WPPSI, they can learn to do very well, without gaining the general intelligence that their test performance would normally be thought to indicate. This observation leads up to the main issue that must be confronted, when thinking about using the WPPSI for robots: The fact that a robot specifically and successfully trained to do well on WPPSI questions, would not necessarily have a level of general intelligence commensurate with an uncoached human child who did well on the same WPPSI questions.

There seems no completely airtight way to handle this problem. It would be very hard to design a rigorous competition between multiple robots of differing cognitive design, with the goal being WPPSI success, without giving a significant advantage to robots whose minds had been specifically engineered to do well on WPPSI-type questions. However, this doesn't necessarily stand in the way of using the WPPSI to assess the intelligence of a robot that has NOT been engineered or coached with the WPPSI specifically in mind. My tentative conclusion is that the WPPSI can be useful as

A way of assessing incremental progress toward general intelligence in a robotic system, whose design and education are not specifically WPPSI based

but NOT as

The assessment measure underlying a competition between robots of differing cognitive design

Educating a Robot in a WPPSI-Relevant Way

Suppose one wants to engineer and educate a robot in a WPPSI-relevant way, without engineering or teaching specifically toward the test. What kind of

education should a robot (and the underlying AI system) have, in order to do well on the WPPSI in a genuine way?

The following 7 capabilities would seem to be critical:

1. **Question answering**. Natural language question answering, about everyday objects and events that a young child would typically know about, including objects and events in the immediate physical environment of the robot; and including questions whose answers involve a few basic reasoning steps based on commonsense knowledge
2. **Object, event and part identification.** Identification of common (in an ordinary child's life) objects and events in pictures, including commonly recognized parts of objects
3. **Object manipulation.** Ability to manipulate objects on a table, such as blocks or puzzle pieces. 3D building doesn't seem critical (if it's an issue for specific robot actuators, it could be skimped on), but pushing things around on a table into different configurations seems critical.
4. **Visual pattern recognition.** Ability to recognize visual patterns regarding objects in the physical environment: objects with different shapes and textures, for example. Ability to recognize visual patterns when drawn on pieces of paper.
5. **Simple drawing.** Ability to draw on a piece of paper with a writing implement – not necessarily words or depictive pictures, but various sorts of marks. Ability to imitate marks that it's seen others write.
6. **Instruction following**. Ability to follow simple natural language instructions regarding simple verbal or physical activities, to be carried out in interaction with the requester
7. **Pragmatic interaction regarding task assignment.** Ability to understand when a task is being assigned, versus when an offhand comment is being made. Ability to understand when the task starts and when it's done, and what the subject is being asked to do. Ability to ask for clarification if the task is not clearly understood. Ability to understand verbal and physical corrections if the task is not being done properly.

WPPSI as a Tool for Gauging Robot Intelligence

In sum, suppose that the following two criteria were fulfilled:

1. a robot and its underlying AI engine were taught to carry out tasks embodying the above 7 capabilities in a reasonably robust way, so that each of these task types could be successfully executed in a variety of contexts besides the WPPSI specifically
2. this robot was not exposed to any WPPSI questions or very close analogues, during its training period (except those that are unavoidable during normal interaction, like basic question answering or picture naming)

In this case, qualitatively, it would seem that the robot was approaching the WPPSI in a genuine way, without specialized "coaching"….. If such a robot did well on the WPPSI, it would seem fair to provisionally conclude that it possessed general intelligence roughly comparable to that of a human preschooler. To validate such a conclusion, one could request a panel of child psychologists to design new questions measuring the same basic skills as the WPPSI, but differing in particulars, and within the physical capabilities of the robot in question. The performance of the robot on the new test questions would be highly informative.

The difficulty of using WPPSI as a challenge problem for a competition, lies in the difficulty of formalizing the above two criteria in a bulletproof way. There is significantly slipperiness in phrases like "a reasonably robust way", "a variety of contexts", and "or very close analogues." But this is not an issue if one's goal is merely to use WPPSI to evaluate the progress of an AGI project, for qualitative rather than comparative purposes.

Another way to look at the relation between WPPSI performance and robot general intelligence would be to create specialized test suites each of the 7 capabilities listed above. These test suites would involve a number of different problems regarding each capability; none closely resembling the WPPSI test questions. One could then study, across multiple instances of the same robot/Ai system with different levels of sophistication and/or different bodies of experience, how the robot's performance on the specialized test suites correlated with performance on the WPPSI. One would expect to find a positive correlation, of course. But if one finds greater correlation regarding overall performance than regarding capability-specific performance, this would provide evidence that the WPPSI is measuring some sort of general intelligence, rather than merely summing up performance on specific capabilities.