Artificial Intelligence and Empirical Psychology

Aaron Sloman
Cognitive Studies Programme, University of Sussex, Brighton, UK
(Now at University of Birmingham, UK http://www.cs.bham.ac.uk/˜axs)

Commentary on Z. Pylyshyn:
Computational models and empirical constraints
Behavioral and Brain Sciences Vol 1 Issue 1 March 1978, pp 91 - 99
(This commentary: pp 115-6)

NOTE (4 May 2015):
I’ve noticed that somehow most of the periods (".") had got lost in the BBS version. They have been restored here.

If I conjecture that the sum of the first n odd numbers is always a perfect square, I can test this with n = 1, n = 2, n = 3, etc. Is this an empirical investigation? If I use a computer instead, is it being used for an "empirical exploration"? These would not normally be called empirical investigations, unlike running the same programs to test a computer.

Consider two definitions: An investigation is empirical_1 if it is based on examination of individual cases, but not if it uses a general proof. It is empirical_2 if (like physics and geology) it is concerned with objects in the world of experience, and not merely (like number theory and theory of computation) with formal abstract structures.

What Pylyshyn is really saying is that some work in Artificial Intelligence is empirical_1, like some mathematical explorations. In other words, AI often uses "formal," but not "substantive," empirical investigations.

Experiments with an AI program might be empirical in both senses. They could reveal a failure of the program to understand something it was intended to be able to cope with (a formal empirical discovery) or they could show that people sometimes use language in a fashion not previously noticed (a substantive empirical discovery). Similarly, a vision program may fail where it was intended to cope, or it may fail in tasks the programmer had not realised most people could cope with.

Pylyshyn suggests that empirical investigations can show "what kinds of relations must (sic) exist". Substantive empirical investigations might show what *can* exist, but "must" in this context presupposes a formal demonstration. The example mentioned, namely Waltz’s program, proves nothing about what must exist. Moreover the power of his label-set is a formal, not a substantive, empirical discovery, whose status as an explanation of human abilities depends on the unavailability of anything better.

AI versus computer simulation

A divergence between AI and simulation systems is predictable. Contrast (a) behaviour based on considerable expertise, built up over many years, like linguistic or perceptual skills, with (b) the floundering, exploratory, non-expert behaviour of beginners struggling with puzzles, like the novice logician, chess-player, or child seriator. Only the latter incompetent behaviour is easily amenable to observation. Deeply-compiled expert skills involve rapid and complex processes not available to introspection or laboratory observation. So the "simulators" will tend to concentrate on (b), unlike the AI fraternity.

But AI programs are still relevant to psychology, since they are testable by their generality, extendability and ability to account for the fine structure of phenomena. When adequate theories of human learning emerge, it may be possible to test some AI models by asking if they could be built up by processes typical of human learning. (Studying how infants learn is distinct from studying what they learn when, as in pre-computational psychology.) AI work tends to produce deeper insights into human processes than simulation studies, since expert behaviour, not fumbling problem-solving protocols, is most characteristically human.

Is parallelism relevant?

Admittedly, a serial computer may be no bar to studying brain processes since parallelism can be simulated as closely as required. But it is clear that many human abilities involve parallel processing at a cognitive level, e.g., a child producing number names, pointing at different objects, and monitoring the two processes to keep them in phase. But even if theories without such parallelism aren’t adequate explanations of how we do things, they are steps towards formalisations of the tasks we perform and the information required for this.

What is a "natural kind"?

There are simple algebraic tests for straightness of a line yet they probably have little to do with how people perceive straightness and other shape properties - an important unsolved AI problem. So the existence of a non-intelligent solution to a problem does not preclude the possibility of solutions using (human) intelligence.

Pylyshyn uses the notion of a pattern or problem being a "natural kind" for humans. Has he forgotten the variability of human beings, and the extent to which a "natural kind" may depend on a cultural context, like the symbols people can recognise easily? A particular class of patterns or problems which now does not form a "natural kind" for humans may one day form part of a widely practised skill. Consider the sight-reading of piano music.

Can we observe computational processes?

Pylyshyn assumes that we can observe intermediate states. But when people or programs produce protocols or answer questions about their strategies this may give misleading information about their normal functioning. Further, much information about what is going on, (e.g., about indexing strategies, matching procedures, rules for parsing and interpreting) may be quite inaccessible to processes concerned with external communication and global decision-making. Procedures may have been compiled into "unreadable" lower level languages, and sub-processes may use "private" work-spaces. Opening the machine to look at its innards would be like trying to understand a very high-level program by examining its machine-code compiled form.

Empirical constraints in common sense

As part of our ability to communicate and our self-knowledge about skills, beliefs, habits etc., we share encyclopaedic knowledge about what people can do. We use it when we gossip, read novels, judge others, or make plans. But psychologists often think that unless they do experiments they are not scientists, so they rarely attempt to analyse and codify this knowledge, as linguists and philosophers do.

Since people doing AI have fewer hang-ups about being scientists, they are more willing to start from common knowledge about what people can do (e.g., understand English, interpret drawings, plan actions, etc.). We can often test explanatory models by noting how their performance falls short of what we know people can do, without relying on new experimental results. Thus there are empirical constraints on AI theories embedded in common sense.

Until AI can account for most of what ordinary people know about people, there may be no urgent need for new psychological data, except in the rare cases where two different models appear to be equivalent in explanatory power, generality, extendability, etc. (Crucial psychological experiments may not always be feasible.)

Of course, common-sense is often mistaken. But, although it is not a good source of laws (indeed, it is doubtful whether there can be laws of psychology), it is a good source of information about possibilities - that is, things people *can* do. The discovery and explanation of possibilities is a major feature of the progress of science. And it is possibilities (abilities, capacities, skills) rather than laws that AI is mostly concerned with explaining.

This will be lost on most psychologists until their training problems are revised to "how is this possible?" instead of "why does this occur?" The latter encourages a search for correlated conditions instead of explanatory mechanisms.

How top-down is AI?

Pylyshyn suggests that AI workers try to devise complete systems. He should have said they try to devise working subsystems. A complete intelligent system would be a teachable robot, with moods emotions, etc. In relation to the task of designing a person, AI work is mainly bottom-up, not top-down, since most computer models deal with a small sub-component of some human ability (e.g. part of the ability to interpret pictures, understand stories, etc.). Pylyshyn meant that AI work first explains general features of an ability, and later adds refinements to explain details. But this is misleading, since relative to long-term goals, AI work is bottom-up, not top-down.

This has serious risks. By reflecting on processes typical of complete human beings, and on interactions between subsystems, we can formulate constraints which current computer models violate. These "interface" constraints may be more significant than constraints generated by the underlying computer - the brain.

For example, what you see can remind you of something, generate changes of mood, help you solve a problem, teach you a concept, help you understand a conversation, etc. What features of a vision system are necessary for this, and what features are required by the other systems? Further study may show that existing program structures are grossly inadequate, even if their factual content (e.g., about image- and scene-features) is correct. But until bottom-up explorations have generated much more technical know-how, it may be premature to switch to the top-down mode and try designing complete systems, except in occasional philosophical moments.

p 115-6