

Physical symbol system

A **physical symbol system** (also called a formal system) takes physical patterns (symbols), combining them into structures (expressions) and manipulating them (using processes) to produce new expressions.

The **physical symbol system hypothesis** (**PSSH**) is a position in the philosophy of artificial intelligence formulated by Allen Newell and Herbert A. Simon. They wrote:

"A physical symbol system has the necessary and sufficient means for general intelligent action."^[1]

— Allen Newell and Herbert A. Simon

This claim implies both that human thinking is a kind of symbol manipulation (because a symbol system is necessary for intelligence) and that machines can be intelligent (because a symbol system is sufficient for intelligence).^[2]

The idea has philosophical roots in Hobbes (who claimed reasoning was "nothing more than reckoning"), Leibniz (who attempted to create a logical calculus of all human ideas), Hume (who thought perception could be reduced to "atomic impressions") and even Kant (who analyzed all experience as controlled by formal rules).^[3] The latest version is called the computational theory of mind, associated with philosophers Hilary Putnam and Jerry Fodor.^[4]

The hypothesis has been criticized strongly by various parties, but is a core part of AI research. A common critical view is that the hypothesis seems appropriate for higher-level intelligence such as playing chess, but less appropriate for commonplace intelligence such as vision. A distinction is usually made between the kind of high level symbols that directly correspond with objects in the world, such as <dog> and <tail> and the more complex "symbols" that are present in a machine like a neural network.

Contents

Examples

Arguments in favor of the physical symbol system hypothesis

Newell and Simon

Symbols vs. signals

Criticism

Dreyfus and the primacy of unconscious skills

Searle and his Chinese room

Brooks and the roboticists

Connectionism

Embodied philosophy

See also

Notes

References

Examples

Examples of physical symbol systems include:

- Formal logic: the symbols are words like "and", "or", "not", "for all x" and so on. The expressions are statements in formal logic which can be true or false. The processes are the rules of logical deduction.
- Algebra: the symbols are "+", "x", "y", "1", "2", "3", etc. The expressions are equations. The processes are the rules of algebra, that allow one to manipulate a mathematical expression and retain its truth.
- A digital computer: the symbols are zeros and ones of computer memory, the processes are the operations of the CPU that change memory.
- Chess: the symbols are the pieces, the processes are the legal chess moves, the expressions are the positions of all the pieces on the board.

The physical symbol system hypothesis claims that both of these are also examples of physical symbol systems:

- Intelligent human thought: the symbols are encoded in our brains. The expressions are thoughts. The processes are the mental operations of thinking.
- A running artificial intelligence program: the symbols are data. The expressions are more data. The processes are programs that manipulate the data.

Arguments in favor of the physical symbol system hypothesis

Newell and Simon

Two lines of evidence suggested to Allen Newell and Herbert A. Simon that "symbol manipulation" was the essence of both human and machine intelligence: the development of artificial intelligence programs and psychological experiments on human beings.

First, in the early decades of AI research there were a number of very successful programs that used high level symbol processing, such as Newell and Herbert A. Simon's General Problem Solver or Terry Winograd's SHRDLU.^[5] John Haugeland named this kind of AI research "Good Old Fashioned AI" or GOFAI.^[6] Expert systems and logic programming are descendants of this tradition. The success of these programs suggested that symbol processing systems could simulate any intelligent action.

And second, psychological experiments carried out at the same time found that, for difficult problems in logic, planning or any kind of "puzzle solving", people used this kind of symbol processing as well. AI researchers were able to simulate the step by step problem solving skills of people with computer programs. This collaboration and the issues it raised eventually would lead to the creation of the field of cognitive science.^[7] (This type of research was called "cognitive simulation".) This line of research suggested that human problem solving consisted primarily of the manipulation of high level symbols.

Symbols vs. signals

In Newell and Simon's arguments, the "symbols" that the hypothesis is referring to are physical objects that represent things in the world, symbols such as <dog> that have a recognizable meaning or denotation and can be composed with other symbols to create more complex symbols.

However, it is also possible to interpret the hypothesis as referring to the simple abstract 0s and 1s in the memory of a digital computer or the stream of 0s and 1s passing through the perceptual apparatus of a robot. These are, in some sense, symbols as well, although it is not always possible to determine exactly what the symbols are standing for. In this version of the hypothesis, no distinction is being made between "symbols" and "signals", as David Touretzky and Dean Pomerleau explain.^[8]

Under this interpretation, the physical symbol system hypothesis asserts merely that intelligence can be *digitized*. This is a weaker claim. Indeed, Touretzky and Pomerleau write that if symbols and signals are the same thing, then "[s]ufficiency is a given, unless one is a dualist or some other sort of mystic,

because physical symbol systems are Turing-universal.^[8] The widely accepted Church-Turing thesis holds that any Turing-universal system can simulate any conceivable process that can be digitized, given enough time and memory. Since any digital computer is Turing-universal, any digital computer can, in theory, simulate anything that can be digitized to a sufficient level of precision, including the behavior of intelligent organisms. The necessary condition of the physical symbol systems hypothesis can likewise be finessed, since we are willing to accept almost any signal as a form of "symbol" and all intelligent biological systems have signal pathways.

Criticism

Nils Nilsson has identified four main "themes" or grounds in which the physical symbol system hypothesis has been attacked.^[2]

1. The "erroneous claim that the [physical symbol system hypothesis] lacks symbol grounding" which is presumed to be a requirement for general intelligent action.
2. The common belief that AI requires non-symbolic processing (that which can be supplied by a connectionist architecture for instance).
3. The common statement that the brain is simply not a computer and that "computation as it is currently understood, does not provide an appropriate model for intelligence".
4. And last of all that it is also believed in by some that the brain is essentially mindless, most of what takes place are chemical reactions and that human intelligent behaviour is analogous to the intelligent behaviour displayed for example by ant colonies.

Dreyfus and the primacy of unconscious skills

Hubert Dreyfus attacked the necessary condition of the physical symbol system hypothesis, calling it "the psychological assumption" and defining it thus:

- *The mind can be viewed as a device operating on bits of information according to formal rules.*^[9]

Dreyfus refuted this by showing that human intelligence and expertise depended primarily on unconscious instincts rather than conscious symbolic manipulation. Experts solve problems quickly by using their intuitions, rather than step-by-step trial and error searches. Dreyfus argued that these unconscious skills would never be captured in formal rules.^[10]

Searle and his Chinese room

John Searle's Chinese room argument, presented in 1980, attempted to show that a program (or any physical symbol system) could not be said to "understand" the symbols that it uses; that the symbols themselves have no meaning or semantic content, and so the machine can never be truly intelligent from symbol manipulation alone. ^[11]

Brooks and the roboticists

In the sixties and seventies, several laboratories attempted to build robots that used symbols to represent the world and plan actions (such as the Stanford Cart). These projects had limited success. In the middle eighties, Rodney Brooks of MIT was able to build robots that had superior ability to move and survive without the use of symbolic reasoning at all. Brooks (and others, such as Hans Moravec) discovered that our most basic skills of motion, survival, perception, balance and so on did not seem to require high level symbols at all, that in fact, the use of high level symbols was more complicated and less successful.

In a 1990 paper Elephants Don't Play Chess (<http://people.csail.mit.edu/brooks/papers/elephants.pdf>), robotics researcher Rodney Brooks took direct aim at the physical symbol system hypothesis, arguing that symbols are not always necessary since "the world is its own best model. It is always exactly up to date. It always has every detail there is to be known. The trick is to sense it appropriately and often enough."^[12]

Connectionism

Embodied philosophy

George Lakoff, Mark Turner and others have argued that our abstract skills in areas such as mathematics, ethics and philosophy depend on unconscious skills that derive from the body, and that conscious symbol manipulation is only a small part of our intelligence.

See also

- Artificial intelligence, situated approach

Notes

1. [Newell & Simon 1976](#), p. 116 and [Russell & Norvig 2003](#), p. 18
2. [Nilsson 2007](#), p. 1
3. [Dreyfus 1979](#), p. 156, [Haugeland](#), pp. 15–44
4. [Horst 2005](#)
5. [Dreyfus 1979](#), pp. 130–148
6. [Haugeland 1985](#), p. 112
7. [Dreyfus 1979](#), pp. 91–129, 170–174
8. Reconstructing Physical Symbol Systems David S. Touretzky and Dean A. Pomerleau Computer Science Department Carnegie Mellon University Cognitive Science 18(2):345–353, 1994. <https://www.cs.cmu.edu/~dst/pubs/simon-reply-www.ps.gz>
9. [Dreyfus 1979](#), p. 156
10. [Dreyfus 1972](#), [Dreyfus 1979](#), [Dreyfus & Dreyfus 1986](#). See also [Russell & Norvig 2003](#), pp. 950–952, [Crevier, 1993 & 120–132](#) and [Hearn 2007](#), pp. 50–51
11. [Searle 1980](#), [Crevier 1993](#), pp. 269–271
12. [Brooks 1990](#), p. 3

References

- [Brooks, Rodney \(1990\)](#), "Elephants Don't Play Chess" (<http://people.csail.mit.edu/brooks/papers/elephants.pdf>) (PDF), *Robotics and Autonomous Systems*, **6** (1–2): 3–15, CiteSeerX 10.1.1.588.7539 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.588.7539>), doi:10.1016/S0921-8890(05)80025-9 (<http://doi.org/10.1016%2FS0921-8890%2805%2980025-9>), retrieved 2007-08-30.
- [Cole, David \(Fall 2004\)](#), "The Chinese Room Argument", in [Zalta, Edward N.](#) (ed.), *The Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/archives/fall2004/entries/chinese-room/>).
- [Crevier, Daniel \(1993\)](#), *AI: The Tumultuous Search for Artificial Intelligence*, New York, NY: BasicBooks, ISBN 0-465-02997-3
- [Dreyfus, Hubert \(1972\)](#), *What Computers Can't Do*, New York: MIT Press, ISBN 978-0-06-011082-6
- [Dreyfus, Hubert \(1979\)](#), *What Computers Still Can't Do*, New York: MIT Press.
- [Dreyfus, Hubert](#); [Dreyfus, Stuart \(1986\)](#), *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, Oxford, U.K.: Blackwell
- [Gladwell, Malcolm \(2005\)](#), *Blink: The Power of Thinking Without Thinking*, Boston: Little, Brown, ISBN 978-0-316-17232-5.
- [Haugeland, John \(1985\)](#), *Artificial Intelligence: The Very Idea*, Cambridge,

Mass.: MIT Press.

- [Hobbes](#) (1651), *Leviathan*.
- [Horst](#), Steven (Fall 2005), "The Computational Theory of Mind", in [Zalta](#), Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/archives/fall2005/entries/computational-mind/>).
- [Kurzweil](#), Ray (2005), *The Singularity is Near*, New York: Viking Press, ISBN 978-0-670-03384-3.
- [McCarthy](#), John; [Minsky](#), Marvin; [Rochester](#), Nathan; [Shannon](#), Claude (1955), *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (<https://web.archive.org/web/20080930164306/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>), archived from the original (<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>) on 2008-09-30.
- [Newell](#), Allen; [Simon](#), H. A. (1963), "GPS: A Program that Simulates Human Thought", in [Feigenbaum](#), E.A.; [Feldman](#), J. (eds.), *Computers and Thought*, New York: McGraw-Hill
- [Newell](#), Allen; [Simon](#), H. A. (1976), "Computer Science as Empirical Inquiry: Symbols and Search", *Communications of the ACM*, **19** (3): 113–126, doi:10.1145/360018.360022 (<https://doi.org/10.1145%2F360018.360022>)
- [Nilsson](#), Nils (2007), [Lungarella](#), M. (ed.), "50 Years of AI" (<https://ai.stanford.edu/%7Enilsson/OnlinePubs-Nils/PublishedPapers/pssh.pdf>) (PDF), *Festschrift, LNAI 4850*, Springer, pp. 9–17 | contribution= ignored ([help](#))
- [Russell](#), Stuart J.; [Norvig](#), Peter (2003), *Artificial Intelligence: A Modern Approach* (<http://aima.cs.berkeley.edu/>) (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2
- [Searle](#), John (1980), "Minds, Brains and Programs" (<https://web.archive.org/web/20150923204614/http://www.class.uh.edu/phil/garson/MindsBrainsandPrograms.pdf>) (PDF), *Behavioral and Brain Sciences*, **3** (3): 417–457, doi:10.1017/S0140525X00005756 (<https://doi.org/10.1017%2FS0140525X00005756>), archived from the original (<http://www.class.uh.edu/phil/garson/MindsBrainsandPrograms.pdf>) (PDF) on 2015-09-23
- [Turing](#), Alan (October 1950), "Computing machinery and intelligence" (<https://web.archive.org/web/20080702224846/http://loebner.net/Prizef/TuringArticle.html>), *Mind*, **LIX** (236): 433–460, doi:10.1093/mind/LIX.236.433 (<https://doi.org/10.1093%2Fmind%2FLIX.236.433>), archived from the original (<http://loebner.net/Prizef/TuringArticle.html>) on 2008-07-02

Retrieved from "https://en.wikipedia.org/w/index.php?title=Physical_symbol_system&oldid=931345065"

This page was last edited on 18 December 2019, at 10:42 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.