Computational theory of mind

Not to be confused with <u>Theory of computation</u> or <u>Pancomputationalism</u>.

In <u>philosophy of mind</u>, the **computational theory of mind** (**CTM**), also known as **computationalism**, is a family of views that hold that the human <u>mind</u> is an information processing system and that cognition and consciousness together are a form of <u>computation</u>. It is closely related to <u>functionalism</u>, a broader theory that defines mental states by what they do rather than what they are made of.^[1]

Warren McCulloch and Walter Pitts (1943) were the first to suggest that neural activity is computational. They argued that neural computations explain cognition. [2] The theory was proposed in its modern form by Hilary Putnam in 1960 and 1961, [3] and then developed by his PhD student, philosopher, and cognitive scientist Jerry Fodor in the 1960s, 1970s, and 1980s. [4][5] It was later criticized in the 1990s by Putnam himself, John Searle, and others.

The computational theory of mind holds that the mind is a computational system that is realized (i.e. physically implemented) by neural activity in the brain. The theory can be elaborated in many ways and varies largely based on how the term computation is understood. Computation is commonly understood in terms of Turing machines

which manipulate symbols according to a rule, in combination with the internal state of the machine. The critical aspect of such a computational model is that we can abstract away from particular physical details of the machine that is implementing the computation. [5] For example, the appropriate computation could be implemented either by silicon chips or biological neural networks, so long as there is a series of outputs based on manipulations of inputs and internal states, performed according to a rule. CTM therefore holds that the mind is not simply analogous to a computer program, but that it is literally a computational system. [5]

Computational theories of mind are often said to require mental representation because 'input' into a computation comes in the form of symbols or representations of other objects. A computer cannot compute an actual object but must interpret and represent the object in some form and then compute the representation. The computational theory of mind is related to the representational theory of mind in that they both require that mental states are representations. However, the representational theory of mind shifts the focus to the symbols being manipulated. This approach better accounts for systematicity and productivity.^[5] In Fodor's original views, the computational theory of mind is also related to the language of thought. The language of thought theory allows the mind to process more complex representations with the help of semantics.

Recent work has suggested that we make a distinction between the mind and cognition. Building from the tradition of McCulloch and Pitts, the *computational theory of cognition* (CTC) states that neural computations explain cognition. The computational theory of mind asserts that not only cognition, but also phenomenal consciousness or <u>qualia</u>, are computational. That is to say, CTM entails CTC. While phenomenal consciousness could fulfill some other functional role, computational theory of cognition leaves open the possibility that some aspects of the mind could be non-computational. CTC, therefore, provides an important explanatory framework for understanding neural networks, while avoiding counterarguments that center around phenomenal consciousness.

"Computer metaphor"

Computational theory of mind is not the same as the computer metaphor, comparing the mind to a modern-day digital computer. [6] Computational theory just uses some of the same principles as those found in digital computing. [6] While the computer metaphor draws an analogy between the mind as software and the brain as hardware, CTM is the claim that the mind is a computational system. More specifically, it states that a computational simulation of a mind is sufficient for the actual presence of a mind, and that a mind truly can be simulated computationally.

'Computational system' is not meant to mean a modern-day electronic computer. Rather, a computational system is a symbol manipulator that follows step-by-step functions to compute input and form output. <u>Alan Turing</u> describes this type of computer in his concept of a <u>Turing</u> machine.

Criticism

A range of arguments have been proposed against physicalist conceptions used in computational theories of mind.

An early, though indirect, criticism of the computational theory of mind comes from philosopher John Searle. In his thought experiment known as the Chinese room, Searle attempts to refute the claims that artificially intelligent agents can be said to have intentionality and understanding and that these systems, because they can be said to be minds themselves, are sufficient for the study of the human mind.^[7] Searle asks us to imagine that there is a man in a room with no way of communicating with anyone or anything outside of the room except for a piece of paper with symbols written on it that is passed under the door. With the paper, the man is to use a series of provided rule books to return paper containing different symbols. Unknown to the man in the room, these symbols are of a Chinese language, and this process generates a conversation that a Chinese speaker outside of the room can actually understand. Searle contends that the man in

the room does not understand the Chinese conversation. This is essentially what the computational theory of mind presents us—a model in which the mind simply decodes symbols and outputs more symbols. Searle argues that this is not real understanding or intentionality. This was originally written as a repudiation of the idea that computers work like minds.

Searle has further raised questions about what exactly constitutes a computation:

WordStar program, because there is some pattern of molecule movements that is isomorphic with the formal structure of WordStar. But if the wall is implementing WordStar, if it is a big enough wall it is implementing any program, including any program implemented in the brain.^[8]

Objections like Searle's might be called insufficiency objections. They claim that computational theories of mind fail because computation is insufficient to account for some capacity of the mind. Arguments from qualia, such as Frank Jackson's knowledge argument, can be understood as objections to computational theories of mind in this way—though they take aim at physicalist conceptions of the mind in general, and not computational theories specifically. [citation needed]

There are also objections which are directly tailored for

computational theories of mind.

Jerry Fodor himself argues that the mind is still a very long way from having been explained by the computational theory of mind. The main reason for this shortcoming is that most cognition is <u>abductive</u> and global, hence sensitive to all possibly relevant background beliefs to (dis)confirm a belief. This creates, among other problems, the <u>frame problem</u> for the computational theory, because the relevance of a belief is not one of its local, syntactic properties but context-dependent. ^[9]

Putnam himself (see in particular *Representation and Reality* and the first part of *Renewing Philosophy*) became a prominent critic of computationalism for a variety of reasons, including ones related to Searle's Chinese room arguments, questions of world-word reference relations, and thoughts about the <u>mind-body problem</u>. Regarding functionalism in particular, Putnam has claimed along lines similar to, but more general than Searle's arguments, that the question of whether the human mind *can* implement computational states is not relevant to the question of the nature of mind, because "every ordinary open system realizes every abstract finite automaton." [10]

Computationalists have responded by aiming to develop criteria describing what exactly counts as an implementation. [11][12][13]

Roger Penrose has proposed the idea that the human mind does not use a knowably sound calculation

procedure to understand and discover mathematical intricacies. This would mean that a normal <u>Turing</u> <u>complete</u> computer would not be able to ascertain certain mathematical truths that human minds can. [14] However, the application of <u>Gödel's theorem</u> by Penrose to demonstrate it was widely criticized, and is considered erroneous. [15]

Pancomputationalism

CTM raises a question that remains a subject of debate: what does it take for a physical system (such as a mind, or an artificial computer) to perform computations? A very straightforward account is based on a simple mapping between abstract mathematical computations and physical systems: a system performs computation C if and only if there is a mapping between a sequence of states individuated by C and a sequence of states individuated by a physical description of the system.^{[16][10]}

Putnam (1988) and Searle (1992) argue that this simple mapping account (SMA) trivializes the empirical import of computational descriptions. [10][17] As Putnam put it, "everything is a Probabilistic Automaton under some Description". [18] Even rocks, walls, and buckets of water—contrary to appearances—are computing systems.

Gualtiero Piccinini identifies different versions of Pancomputationalism. [19]

In response to the trivialization criticism, and to restrict

SMA, philosophers of mind have offered different accounts of computational systems. These typically include causal account, semantic account, syntactic account, and mechanistic account. [20] Instead of a semantic restriction, the syntactic account imposes a syntactic restriction. [20] The mechanistic account was first introduced by <u>Gualtiero Piccinini</u> in 2007. [21]

Notable theorists

This section **does not <u>cite</u> any <u>sources</u>**. Please help <u>improve this section</u> by <u>adding citations to reliable sources</u>. Unsourced material may be challenged and <u>removed</u>. (July 2023) (Learn how and when to remove this message)

- <u>Daniel Dennett</u> proposed the <u>multiple drafts model</u>, in which consciousness seems <u>linear</u> but is actually blurry and gappy, distributed over space and time in the brain. Consciousness is the computation, there is no extra step in which you become conscious of the computation.
- Jerry Fodor argues that mental states, such as beliefs and desires, are relations between individuals and mental representations. He maintains that these representations can only be correctly explained in terms of a language of thought (LOT) in the mind. Further, this language of thought itself is codified in the brain, not just a useful explanatory tool. Fodor adheres to a species of functionalism, maintaining that thinking and other mental processes consist

- primarily of computations operating on the syntax of the representations that make up the language of thought. In later work (*Concepts* and *The Elm and the Expert*), Fodor has refined and even questioned some of his original computationalist views, and adopted LOT2, a highly modified version of LOT.
- David Marr proposed that cognitive processes have three levels of description: the *computational level*, which describes that computational problem solved by the cognitive process; the *algorithmic level*, which presents the algorithm used for computing the problem postulated at the computational level; and the *implementational level*, which describes the physical implementation of the algorithm postulated at the algorithmic level in the brain.
- <u>Ulric Neisser</u> coined the term cognitive psychology in his book with that title published in 1967. Neisser characterizes people as dynamic informationprocessing systems whose mental operations might be described in computational terms.
- Steven Pinker described language instinct as an evolved, built-in capacity to learn language (if not writing). His 1997 book <u>How the Mind Works</u> sought to popularize the computational theory of mind for wide audiences.
- Hilary Putnam proposed <u>functionalism</u> to describe consciousness, asserting that it is the computation that equates to consciousness, regardless of whether the computation is operating in a brain or in a

computer.

See also

- Artificial consciousness
- Cognitivism (psychology)
- Constructivist epistemology
- Enchanted loom
- Simulated reality
- Stimulus-response model
- Stochastic parrot

Alternative theories

- Adaptive systems
- Associationism
- Connectionism
- Enactivism
- Memory-prediction framework
- Neurophenomenology
- Perceptual control theory
- Situated cognition

References

- <u>"Functionalism"</u>. Stanford Encyclopedia of Philosophy. 2023.
- 2. ^ a b Piccinini, Gualtierro & Bahar, Sonya, 2012.

 "Neural Computation and the Computational Theory of Cognition" in Cognitive Science.

- https://onlinelibrary.wiley.com/doi/epdf/10.1111/cogs.1 2012
- 3. Horst, Steven (1999). <u>"Symbols and Computation A Critique of the Computational Theory of Mind"</u>. Minds and Machines. **9** (3): 347–381. <u>doi:10.1023/A:1008351818306</u>.
- 4. Putnam, Hilary, 1961. "Brains and Behavior", originally read as part of the program of the American Association for the Advancement of Science, Section L (History and Philosophy of Science), December 27, 1961, reprinted in Block (1983), and also along with other papers on the topic in Putnam, *Mathematics, Matter and Method* (1979)
- 5. ^ a b c d Horst, Steven, (2005) <u>"The Computational Theory of Mind"</u> in *The Stanford Encyclopedia of Philosophy*
- 6. ^ a b Pinker, Steven. The Blank Slate. New York: Penguin. 2002
- 7. Searle, J.R. (1980), "Minds, brains, and programs" (PDF), The Behavioral and Brain Sciences, **3** (3): 417–457, doi:10.1017/S0140525X00005756, S2CID 55303721
- 8. Searle, J.R. (1992), The Rediscovery of the Mind
- 9. Fodor, J. (2000). <u>The Mind Doesn't Work That Way:</u>
 <u>The Scope and Limits of Computational Psychology.</u>
 The MIT Press. <u>ISBN 978-0-262-56146-4</u>.
- 10. ^ a b c Putnam, H. (1988). Representation and Reality. Cambridge, Massachusetts: The MIT Press. ISBN 978-0-262-66074-7. OCLC 951364040.

- 11. Chalmers, D.J. (1996), "Does a rock implement every finite-state automaton?", Synthese, **108** (3): 309—333, CiteSeerX 10.1.1.33.5266, doi:10.1007/BF00413692, S2CID 17751467, archived from the original on 2004-08-20, retrieved 2009-05-27
- 12. Edelman, Shimon (2008), "On the Nature of Minds, or: Truth and Consequences" (PDF), Journal of Experimental and Theoretical AI, 20 (3): 181–196, CiteSeerX 10.1.1.140.2280, doi:10.1080/09528130802319086, S2CID 754826, retrieved 2009-06-12
- 13. Blackmon, James (2012). "Searle's Wall". Erkenntnis. **78**: 109–117. <u>doi:10.1007/s10670-012-9405-4</u>. <u>S2CID 121512443</u>.
- Roger Penrose, "Mathematical Intelligence," in Jean Khalfa, editor, What is Intelligence?, chapter 5, pages 107-136. Cambridge University Press, Cambridge, United Kingdom, 1994
- 15. "The Computational Theory of Mind". Stanford Encyclopedia of Philosophy. 2020. "Various philosophers and logicians have answered the critique, arguing that existing formulations suffer from fallacies, question-begging assumptions, and even outright mathematical errors [...]. There is a wide consensus that this criticism of CCTM lacks any force."
- 16. Ullian, Joseph S. (March 1971). <u>"Hilary Putnam.</u>

 <u>Minds and machines. Minds and machines, edited by</u>

- Alan Ross Anderson, Prentice-Hall, Inc., Englewood
 Cliffs, New Jersey, 1964, pp. 72–97. (Reprinted from
 Dimensions of mind, A symposium, edited by Sidney
 Hook, New York University Press, New York 1960, pp.
 148–179.)". Journal of Symbolic Logic. 36 (1): 177.
 doi:10.2307/2271581. ISSN 0022-4812.
 JSTOR 2271581.
- 17. Smythies, J. R. (November 1993). "The Rediscovery of the Mind. By J. R. Searle. (Pp. 286; \$22.50.) The MIT Press: Cambridge, Mass.1992". Psychological Medicine. 23 (4): 1043–1046.

 doi:10.1017/s0033291700026507. ISSN 0033-2917. S2CID 143359028.
- 18. <u>"ART, MIND, and RELIGION"</u>. Philosophical Books. **8** (3): 32. October 1967. <u>doi:10.1111/j.1468-</u> <u>0149.1967.tb02995.x</u>. <u>ISSN 0031-8051</u>.
- 19. Piccinini, Gualtiero (2015-06-01), <u>"The Mechanistic Account"</u>, Physical Computation, Oxford University Press, pp. 118–151, <u>doi:10.1093/acprof:oso/9780199658855.003.0008</u>, <u>ISBN 978-0-19-965885-5</u>, retrieved 2020-12-12
- 20. ^ a b Piccinini, Gualtiero (2017), "Computation in Physical Systems", in Zalta, Edward N. (ed.), The Stanford Encyclopedia of Philosophy (Summer 2017 ed.), Metaphysics Research Lab, Stanford University, retrieved 2020-12-12
- 21. Piccinini, Gualtiero (October 2007). <u>"Computing Mechanisms*"</u>. Philosophy of Science. **74** (4): 501–526. <u>doi:10.1086/522851</u>. <u>ISSN 0031-8248</u>.

Further reading

- <u>Block, Ned</u>, ed. (1983). <u>Readings in Philosophy of Psychology</u>. Vol. 1. Cambridge, Massachusetts:
 Harvard University Press. <u>ISBN 067474876X</u>.

 <u>OCLC 810753995</u>.
- Chalmers, David (2011). <u>"A Computational</u>
 <u>Foundation for the Study of Cognition"</u>. Journal of
 Cognitive Science. **12** (4): 323–357.
- <u>Crane, Tim</u> (2016). The Mechanical Mind: A
 Philosophical Introduction to Minds, Machines, and Mental Representation (3nd ed.). London and New York: Routledge. <u>ISBN 9781138858329</u>.

 <u>OCLC 964575493</u>.
- <u>Fodor, Jerry</u> (1975). The Language of Thought. Cambridge, Massachusetts: Harvard University Press. <u>ISBN 0674510305</u>. <u>OCLC 15149586</u>.
- Fodor, Jerry (1995). The Elm and the Expert: Mentalese and Its Semantics (eBook ed.).
 Cambridge, Massachusetts: The MIT Press. doi:10.7551/mitpress/2693.001.0001.
 ISBN 9780262272889. OCLC 9470770683.
- Fodor, Jerry (1998). Concepts: Where Cognitive Science Went Wrong. Oxford Cognitive Science Series. Oxford and New York: Clarendon Press of the Oxford University Press. <u>ISBN</u> <u>9780198236375</u>. <u>OCLC</u> <u>38079317</u>.

- Fodor, Jerry (2000). The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology. Cambridge, MA and London: The MIT Press. <u>ISBN 9780262062121</u>. <u>OCLC 43109956</u>.
- Fodor, Jerry (2010). LOT 2: The Language of Thought Revisited. Oxford and New York: Oxford University Press. <u>ISBN 9780199548774</u>.
 OCLC 470698989.
- Harnad, Stevan (1994). "Computation Is Just Interpretable Symbol Manipulation: Cognition Isn't". Minds and Machines. 4 (4): 379–390. doi:10.1007/bf00974165. S2CID 230344.
- Marr, David (2010) [1981]. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Cambridge, MA and London: The MIT Press. <u>ISBN 978-0-262-51462-0</u>. OCLC 472791457.
- <u>Piccinini, Gualtiero</u> (2015). Physical Computation: A
 Mechanistic Account. Oxford: Oxford University
 Press. <u>ISBN 9780199658855</u>. <u>OCLC 920617851</u>.
- <u>Pinker, Steven</u> (1997). How the Mind Works. New York: W. W. Norton. <u>ISBN</u> <u>978-0393045352</u>. <u>OCLC</u> <u>36379708</u>.
- Putnam, Hilary (1995) [1979]. Mathematics, Matter, and Method. Philosophical Papers, Volume 1 (2nd ed.). Cambridge and New York: Cambridge University Press. ISBN 0-521-29550-5.
 OCLC 258667059.
- Putnam, Hilary (1995). Renewing Philosophy

(Reissue ed.). Cambridge, Massachusetts: Harvard University Press. <u>ISBN</u> <u>9780674760943</u>. <u>OCLC</u> <u>60298037</u>.

- <u>Pylyshyn, Zenon</u> (1984). Computation and Cognition: Toward a Foundation for Cognitive Science.
 Cambridge, Massachusetts: The MIT Press.
 <u>ISBN</u> 0262160986. OCLC 10507591.
- <u>Searle, John</u> (1992). The Rediscovery of the Mind.
 Cambridge, Massachusetts: The MIT Press.
 <u>ISBN 026269154X</u>. <u>OCLC 760581004</u>.
- Zalta, Edward N. (ed.). "The Computational Theory of Mind". Stanford Encyclopedia of Philosophy.

External links

- Computational theory of mind at the <u>Indiana</u>
 Philosophy Ontology Project
- Computational theory of mind at PhilPapers
- Online papers on consciousness, part 2: Other
 Philosophy of Mind, compiled by David Chalmers

•	<u>v</u>
•	<u>t</u>
•	<u>e</u>
	Philosophy of mind
	 G. E. M. Anscombe Aristotle Armstrong

Thomas Aquinas

L. Austin

<u>Alexander Bain</u>

- George Berkeley
- Henri Bergson
- Ned Block
- Franz Brentano
- C. D. Broad
- Tyler Burge
- David Chalmers
- Patricia Churchland
- Paul Churchland
- Andy Clark
- Dharmakirti
- Donald Davidson
- Daniel Dennett
- René Descartes
- Fred Dretske
- Fodor
- Goldman
- Martin Heidegger
- David Hume
- Edmund Husserl
- William James
- Frank Cameron Jackson
- Immanuel Kant
- David Lewis (philosopher)
- John Locke
- Gottfried Wilhelm Leibniz
- Maurice Merleau-Ponty
- Marvin Minsky
- Thomas Nagel
- Alva Noë
- Derek Parfit
- Plato
- Hilary Putnam
- Richard Rorty
- Gilbert Ryle
- John Searle
- Wilfrid Sellars
- Baruch Spinoza
- Alan Turing



	 Michael Tye Vasubandhu Ludwig Wittgenstein Stephen Yablo Zhuangzi more 		
Theories	 Behaviorism Biological naturalism Dualism Eliminative materialism Emergent materialism Epiphenomenalism Functionalism Interactionism Naïve realism Neurophenomenology Neutral monism New mysterianism Nondualism Occasionalism Parallelism Phenomenalism Phenomenology Physicalism Type physicalism Property dualism Representational Solipsism Substance dualism 		
	 Abstract object Chinese room Creativity Cognition Cognitive closure Concept Consciousness 		

	 Hard problem of consciousness 			
	 Hypostatic abstraction 			
	• Idea			
	• Identity			
	• <u>Intelligence</u>			
	Artificial			
	• Human			
	• Intentionality			
	• Introspection			
Concepts	• Intuition			
	Language of thought			
	Mental event			
	Mental image			
	Mental process			
	Mental property			
	 Mental representation 			
	• Mind			
	Mind-body problem			
	• Pain			
	 Problem of other minds 			
	 Propositional attitude 			
	• Qualia			
	Tabula rasa			
	 Understanding 			
	• Zombie			
	<u>Zorrioro</u>			
	Metaphysics			
Related	 Philosophy of artificial intelligence / 			
Related	information / perception / self			
 <u>Category</u> 				
 Philosopher 	<u>s category</u>			
 <u>Project</u> 				
• <u>Task Force</u>				
•	<u>V</u> +			
•	<u>t</u>			
•	<u>e</u>			

Evolutionary psychology

- <u>History</u>
 - Evolutionary thought
- Theoretical foundations
 - Adaptationism
 - o Cognitive revolution
 - o Cognitivism
 - o Gene selection theory
 - Modern synthesis
- Criticism

Evolutionary processes

- Adaptations
- Altruism
 - Cheating
 - Hamiltonian spite
 - Reciprocal
- Baldwin effect
- By-products
- Evolutionarily stable strategy
- Exaptation
- Fitness
 - Inclusive
- Kin selection
- Mismatch
- Natural selection
- Parental investment
 - Parent-offspring conflict
- Sexual selection
 - Costly signaling
 - Male/female intrasexual competition
 - Mate choice
 - Sexual dimorphism
- Social selection

• Affect

- Display
- Display rules
- Facial expression

	Cognition / Emotion	 Behavioral modernity Cognitive module/modular Automatic and contreprocesses Computational theor Domain generality Domain specificity Dual process theory Cognitive tradeoff hypothe Evolution of the brain Evolution of nervous syste Fight-or-flight response Arachnophobia Basophobia Ophidiophobia Folk biology/taxonomy Folk psychology/theory of Intelligence Flynn effect Wason selection task Motor control/skill Multitasking Sleep Visual perception Color vision
	Culture	 Naïve physics Aesthetics Literary criticism Musicology Anthropology Biological Crime Language Origin Psychology Speech Morality Moral foundations Religion Origin Universals
		 Attachment

	Development	 Bonding Affectional/maternal/pateri Caregiver deprivation Childhood attachment Cinderella effect Cognitive development Education Language acquisition Personality development Socialization
Areas	Human factors / Mental health	 Cognitive ergonomics Computer-mediated communication Engineering psychols Human-computer int Media naturalness th Neuroergonomics Depression Digital media use and men Hypophobia Imprinted brain hypothesis Mind-blindness Psychological effects of Intermediate in the second second
		 Activity Adult attachment Age disparity Arousal Concealed ovulation Coolidge effect Desire Fantasy Hormonal motivation

Sex	 Jealousy Mate guarding Mating preferences Mating strategies Orientation Ovulatory shift hypothesis Pair bond Physical/Sexual attraction Sexuality/male/female Sexy son hypothesis Westermarck effect
Sex differences	 Aggression Autism Cognition Crime Division of labour Emotional intelligence Empathising—systemising t Gender role Intelligence Memory Mental health Narcissism Neuroscience Schizophrenia Substance abuse Suicide Variability hypothesis
Academic disciplines	 Behavioral/evolutionary econo Behavioral epigenetics/genetic Affective/behavioral/cognitive/neuroscience Biocultural anthropology Biological psychiatry Cognitive psychology Cognitive science Cross-cultural psychology Ethology Evolutionary biology Evolutionary medicine Functional psychology Neuropsychology

		 Philosophy of mind Population genetics Primatology Sociobiology
Related subjects	Research topics	 Cultural evolution Evolutionary epistemology Great ape language Human-animal communication Missing heritability problem Primate cognition Unit of selection Coevolution Cultural group selection Dual inheritance theory Fisher's principle Group selection Hologenome theory Lamarckism Population Punctuated equilibrium Recent human evolution Species Species Species complex Transgenerational epiger inheritance Trivers-Willard hypothes
	Theoretical positions	 Cultural selection theory Determinism/indeterminism Biological determinism Connectionism Cultural determinism Environmental determinis Nature versus nurture Psychological nativism Social constructionism Social determinism Standard social science Functionalism Memetics Multilineal evolution Neo-Darwinism Neoevolutionism Sociocultural evolution

- Unilineal evolution
- Evolutionary psychologists
- Evolutionary psychology
- Psychology portal
- Evolutionary biology portal