4 Can Chinese Rooms Think?

The History and Status of the Debate — Map 4 of 7

An Issue Map™ Publication

Start Here

1 Alan Turing, 1950 **Yes, machines can (or will be able to) think.** A computational system can possess all important elements of human thinking or understanding.

I believe that at the end of the century ... one will be able to speak of machines thinking without expecting to be contradicted.

Alan Turing

2 Allan Newell and Herbert Simon, 1976 **Physical symbol systems can think.** "A physical symbol system has the necessary and sufficient means for general intelligent action. By 'necessary' we mean that any system that exhibits general intelligence will prove upon analysis to be a physical symbol system. By 'sufficient' we mean that any physical symbol system of sufficient size can be organized to produce intelligence. General intelligence." **Note:** The Chinese Room debate mapped here is only one of many debates related to the question of whether physical symbol systems can think. Other sets of arguments are mapped on Map 3.

Allen Newell — Herbert Simon

supported by

3 John Searle, 1980, 1980b, 1990b **Strong AI is false.** Imagine that a man who does not speak Chinese sits in a room and is passed Chinese symbols through a slot in the door. To him, the symbols are just so many squiggles and squeggles. But he reads an English-language book that tells him which Chinese symbols the symbols and which ones to send back out the slot. To the Chinese speakers outside, whoever (or whatever) is in the room is carrying on an intelligent conversation. But the man in the Chinese Room does not understand Chinese; he is merely manipulating symbols according to a rule book. Searle concludes that passing the Turing test (see the Turing test for intelligence, but nevertheless he does not understand Chinese. This shows that instantiation of a formal program is not enough to produce semantic understanding or intentionality.

Note: For more on Turing tests, see Map 2. For more on formal programs and instantiation, see the "Is the brain a computer?" arguments on Map 1, the "Can functional states generate consciousness?" arguments on Map 6, and sidebar, "Formal Systems: An Overview," on Map 7.

I'm just manipulating squiggles and squoggles according to a computer program's behavior. But I don't understand Chinese. This rule book is in English.

在这里使动的任何人或物一无理解。

[Whoever or whatever is in that room] is an intelligent Chinese speaker!

This is a syntactic—or purely formal—system. Input 1 and you respond to it by inscribing squiggle number 2. Input 2 and you respond by inscribing squiggle 3.

John Searle

in · tent · ion · al · ity *n.* The property (in reference to a mental state) of being directed at a state of affairs in the world. For example, belief that Sally is in front of me is directed at a person, Sally, in the world. Intentionality is a hallmark of the mental. Intentionality, consciousness, understanding, consciousness, meaning, and semantics. Although these are important and subtle distinctions in the definitions of "intentionality," "understanding," "semantics," and "meaning," in this debate they are sometimes used synonymously.

Can the Chinese Room, considered as a total system, think?

18 Anticipated by John Searle, 1980, 1980b, 1990b **The systems reply.** Because neither the person in the room nor part of a total system that includes the rule book, scratch paper, and data banks of Chinese symbols. The whole system, not only the person, should be regarded as understanding Chinese.

That whole system understands Chinese.

supported by

19 Richard Sharvy, 1983 **The Chinese Room argument refutes the point of the Chinese Room thought experiment.** All the Chinese Room argument is intended to show is that instantiation of a program that understands Chinese is not sufficient to prove understanding. For this, all Searle needs to show is that something instantiates such a program but does not understand Chinese. So, the fact that the room as a whole may understand is irrelevant, because the man (who instantiates the program) doesn't understand Chinese. He is enough of a counterexample by himself.

20 David Cole, 1984 **The Chinese Room argument commits the fallacy of composition.** Just because one part of the Chinese Room (the man shuffling symbols) doesn't understand Chinese, it doesn't logically follow that the Chinese Room as a whole doesn't understand Chinese.

David Cole

21 Richard Double, 1983 **The man in the room is not in a position to judge for the rest of system.** Whether the room as a whole understands Chinese has nothing to do with the man's understanding of Chinese. The man's perspective provides no evidence about the system as a whole, and hence is irrelevant for determining whether the system has mentality.

22 C. A. Fields, 1984 **It Doubles account, brains lack understanding.** Double thinks that a part of a system, such as a homunculus in a Japanese Room, cannot understand in the way that a system as a whole can. Based on that account, a brain is a person cannot understand in the way that a person as a whole can.

fal · la · cy of com · po · si · tion: Something true of the parts must hold true for the properties of the whole. For example, individual water molecules aren't wet, but it doesn't follow that water itself is not wet.

I'm not wet.

ho · mun · cu · lus: Literally, little man. In philosophy, a homunculus generally refers to a rational person, agent, or being used to make some philosophical point.

23 Richard Double, 1984 **Fields mistakes an epistemological point for a metaphysical one.** The Japanese Room thought experiment was only supposed to make an epistemological point that a man in a room can't tell us whether the room as a whole is thinking. It was not supposed to make a metaphysical point about what can or can't have intentionality.

24 Richard Rorty, 1983 **The Japanese Room thought experiment.** Suppose we have a computer system that behaves as if it understands Japanese. Now imagine that the system is large enough that we can put a man inside of it and ask him to run some part of the system. We would not expect this man in the Japanese Room to be aware of the existence of the system he is running or to be able to answer questions about it. He doesn't have access to the activities of the system as a whole. The man's observations thus say nothing about the intentionality of the Japanese Room as a whole.

Is strong AI valid category?

supporting arguments

25 John Searle, 1980b, 1980b, 1990b **The Chinese Room argument refutes strong AI but not weak AI.** The Chinese Room argument is intended to show that strong AI—that running a computer program is not enough to generate any real understanding or intentionality. However, it does not refute weak AI, which simply uses computers as tools to study the mind. Weak AI is an appropriate research program, strong AI is not.

weak AI: The position that computers are a useful tool in psychology. They help researchers test and evaluate theories about how the mind works.

strong AI: The position that an appropriately programmed computer really is a mind. Computers can possess mental states just by virtue of formal symbol manipulation.

26 Yorick Wilks, 1982 **Searle attacks a straw man.** By attributing strong AI to proponents of the Chinese Room argument, Searle attacks a straw man. No one in AI thinks that successfully functioning programs must explain human understanding. Programs generally offer their computer simulations as models of what human thinking might be like.

straw man: A misrepresentation of an opponent's view that makes it easier to attack.

27 John Searle, 1982 **Just read what they've written.** There are lots of cognitive scientists who are adherents of strong AI—just read what people have written. They have identified themselves in their published commentaries on the Chinese Room argument.

Can brain simulators think?

4 Anticipated by John Searle, 1980a, 1980b, 1990b **The brain simulator reply.** A computational system that simulated the sequence of events (i.e., the pattern of neuronal firings) of a Chinese speaker would be able to understand Chinese.
Supporters of
"The Chinese Gym Argument," Map 5, Box 5.

Note: This has also been called the connectionist reply, which claims that connectionist neural networks are not affected by the Chinese Room argument.

5 John Searle, 1980a, 1980b, 1990b **The Chinese water-pipe brain simulator.** Imagine that the man in the Chinese Room is given a new rule book that tells him how to operate a huge system of water pipes. The water pipes simulate the pattern of neurons firing in a Chinese speaker's brain. The man still doesn't understand Chinese, but he does the English rule book which valves to turn on and off, then he lets the brain simulator do its work, and the faucets squirt out a Chinese response. Even though the water pipes simulate the formal properties of the Chinese speaker's brain, the man controlling them has no understanding of Chinese.

Can robots think?

6 Anticipated by John Searle, 1980a, 1980b, 1990b **The robot reply.** Suppose that a computer system were put inside a robot that had been equipped with a television camera for sight, mechanical arms for doing things, and legs for movement. Such a system, equipped like a human to interact with the world, would have genuine intentionality.

7 John Searle, 1980a, 1980b, 1990b **The Chinese Room can be placed inside the robot.** Merely putting a computer program in a robot doesn't mean it can put the Chinese Room inside a robot and apply the same thought experiment as before. Send Chinese symbols from a television camera to the man in the room and transform the man's output into robotic arm and leg movements. The man is still just manipulating formal symbols without knowing what they mean.

Searle uses the wrong causal connection with the world.

8 Jerry Fodor, 1980 **Causal connection with the world is essential to meaning.** Mental states become meaningful (i.e., intentional or semantic) by being causally connected to the world in the right way. Causes and through the ability to act on the external world. Searle's response to the robot reply shows so that a little man mediating between the robot's sensors and its mental connection, but Searle's response does not demonstrate that no causal connection could ever produce meaning at all.

10 Nazika Newton, 1988 **Repeatable goal-directed actions can ground intentionality.** Searle has not yet recognized that his critics do intentionality. The robot has intentionality by virtue of the fact that it makes connections between symbols and its past and present goal-directed actions.

11 Nazika Newton, 1988 **A goal-directed robot.** Imagine a robot that performs at least 2 kinds of tasks.
- It is able to move around its environment to find electrical sockets that recharge it. If a socket recharges it, the robot marks "yes" in its memory for that socket. If the socket drains the robot's power, it marks "no" in its memory.
- It is able to answer questions about stories in its memory, responding accurately to queries with "yes" or "no" answers.

Such a robot has intentionality if it can demonstrate that its "yes" and "no" answers are caused by some of the actions its uses when seeking to recharge its batteries. Intentionality just amounts to this kind of connection between an entity's symbols and its goal-directed actions.

9 John Searle, 1980a, 1980b, 1990b **Symbols and causal connections cannot by themselves generate intentionality.** Formal or causal connections are not sufficient to produce meaning (or intentionality). If there is any meaning in a system, it can only be because there is something more to the system than just the fact of a symbol and the fact of a causal connection.

Symbols and causal connections are not enough to get intentionality.

Can a combination robot/brain simulator think?

12 Anticipated by John Searle, 1980a, 1980b, 1990b **The combination reply.** Take separately, the systems reply (Box 18), the robot reply (Box 6) and the brain simulator reply (Box 5). Together, they prove that computers can think. Taken in combination, however, they show that a syntactic system can think. That is, a total Chinese-speaking system that included a robotic body, transducers, and a brain simulator, would be able to think.

13 John Searle, 1980a, 1980b, 1990b **If we knew how the robot worked.** If the robot looked and behaved exactly like a human, we would be willing to attribute intentionality to it for purposes of interpretation. However, if we knew that it was given as another way to explain its behavior without appeal to intentionality (e.g., if we found out that it acted on the basis of formal symbol manipulations), then we would no longer appeal to intentionality in to explain its behavior.

Do humans, unlike computers, have intrinsic intentionality?

14 John Searle, 1980a, 1980b, 1990b **Only minds are intrinsically intentional.** Mental states alone are intrinsically intentional. Computational systems are only intentional relative to some observer who treats them as if they are intentional. Only minds really have intrinsic intentionality. Computers always have observer relative or "as if" intentionality.

That machine really possesses intentionality as some observers.

Nothing else has original intentionality.

15 Daniel Dennett, 1987 **Only Mother Nature exhibits original intentionality.** Humans do not exhibit original intentionality (or, to use Searle's terms, intrinsic intentionality); rather, their intentionality is derived from genes. Only Mother Nature exhibits intentionality. The regress ends, however: Mother Nature exhibits original intentionality.

16 Ronald McIntyre, 1986 **Computers can have intentionality whose that mental states are intrinsically intentional.** According to Edmond Husserl, meanings and mental states are intentional not because of any relation they bear to anything else, but as a matter of their intrinsic character. They are "a sort of entity whose very nature is to mean something other than itself." By this, Husserl rule-governed aspects of mental states, but those aspects do not exhaustively characterize mental states. Mental states cannot be reduced to formal structures, even if they have formal aspects. As such, mental states cannot be duplicated by formal computer models.

Do Husserl's account intentional states can't be duplicated by a purely formal model like a computer program.

Just the opposite! Husserl shows that intentionality can be duplicated by a formal computer model.

17 James Moreoli, 1991 **The formal structure of intentionality can be modeled on a computer.** According to Husserl, intentionality is constructed from meaningless elements in a rule-governed way. The construction of intentionality occurs through a process of "passive synthesis," in which the data of sense experience are ordered into intentional perceptions "without any active participation of the subject" (p. 115-116). Similarly, computers use rules to synthesize meaningless elements into meaningful wholes. So Husserl shows that mental states can be duplicated by formal computer programs.

Can an internalized Chinese Room think?

supporting arguments

36 John Searle, 1980a, 1980b, 1990b **The internalization reply.** Suppose the man in the Chinese Room memorizes the rule book and all the symbols, and does the machine in his head rather than on scratch paper. Then he can internalize the entire Chinese Room system, but he still does not understand Chinese. So, the whole system doesn't understand Chinese.

37 Robert Wilensky, 1980 **The subsystem understands.** The man who has internalized the room may not understand Chinese, but he contains an independently functioning subsystem that does understand Chinese.
Note: Similar arguments are played out in the "Can translations occur between the internalized Chinese Room and the internalizing English speaker?" arguments on this map.

38 Daniel Dennett, 1987 **Understanding is the result of speed and complexity.** No human being can manipulate symbols fast enough or accurately enough to come close to simulating Chinese translation in real time. But if computers can simulate the Chinese symbols, as suggested by the internalization reply, It is the very speed and complexity of computer software that prevents humanlike thinking to occur.

Daniel Dennett

39 John Fisher, 1988 **The levels of conscious involvement dilemma.** Once the man internalizes the system, the Chinese Room argument fails regardless of what level of involvement he imagine the man to have with it.

Either	Or
The man unconsciously follows the rules in the rule book. *in which case* strong AI would not expect the man to understand Chinese any more than it would expect any other element of the Chinese-speaking system to understand Chinese.	The man consciously follows the rules in the rule book. *in which case* as far as we know the man may understand Chinese after all.

In Either Case
The predictions of strong AI are not refuted.

40 John Fisher, 1988 **We lack adequate intuitions about the man who internalizes the Chinese Room.** Searle's Chinese Room argument we have clear intuitions about what's going on inside the man's head—we know he understands English but not Chinese. However, once he internalizes the room, we lack clear intuitions. We don't know what's going on in his head anymore because of the complexity of the process of internalizing the room. We lack adequate intuitions to determine whether or not he understands Chinese.

That's so complex that I have no idea whether or not he understands Chinese.

41 John Fisher, 1988 **Phenomenology is not required of a general cognitive theory of understanding.** Even if we think we have intuitions about what goes on in the mind of the man who internalizes the room, phenomenology is not a necessary component theory of understanding. A general cognitive theory of understanding is only responsible for explaining competence (having the right internal states) and performance (applying the rules in context), but is not responsible for explaining the phenomenology of understanding.

We don't need phenomenology for a general cognitive theory.

42 Philip Cam, 1990 **A dilemma about cognition and consciousness.** The internalization reply relies on the notion that cognition and intentionality are necessarily connected. But this leads to a dilemma.

Either	Or
The man engages in cognitive activity without Chinese-speaking intentionality. When he operates the internalized Chinese Room, he acts thinks, even though he doesn't understand what he is squiggling and squoggling. *in which case* we can't justifiably deny the presence of some kind of cognition in the internalized Chinese Room.	The man engages in cognitive activity with Chinese speaking intentionality (e.g., he may know that "squiggle-squiggle" is generally followed by "squiggle-squoggle"). *in which case* that intentionality may carry over to the internalized Chinese Room as well.

In Either Case
The internalization reply doesn't conclusively refute the idea that machines can think.

43 Jack Copeland, 1993 **The "part-of" principle is fallacious.** The internalization reply rests on a fallacious "part-of" principle, which, if it isn't allowed, would let us say that a part of a person's brain can't think because it is part of a thinking thing. Similarly, the fact that the man internalized the Chinese Room can't speak Chinese, so it does not follow that no part of him can speak Chinese. Searle's response to this objection, that Searle's response does not demonstrate that no causal connection could ever produce meaning at all.

Jack Copeland

Can translations occur between the internalized Chinese Room and the internalizing English speaker?

44 David Cole, 1984 **The man understands Chinese, but he can't translate it into English.** Once the man has internalized the system, he then speaks and understands Chinese. But his English- and Chinese-speaking subsystems can't communicate. He understands both English and Chinese but cannot translate between them.

I speak perfect English.

(I speak perfect Chinese.)

45 J. Christopher Maloney, 1987 **We should listen to the man.** The man's claim that he doesn't understand Chinese is overrides his behavior (which indicates that he does speak Chinese). He has the last word on what his mental states are.

I'ts telling you, I don't understand Chinese!

46 J. Christopher Maloney, 1987 **Failure to translate proves failure to understand.** The man's claim that he is necessary consequence of forming a second language. If you can't translate from a second language, then you don't understand that language. Because the man who internalizes the room can't translate from Chinese to English, he doesn't understand Chinese.

50 David Cole, 1984 **The brain-programming thought experiment.** Scientists program an isolated patch of Searle's brain to understand Chinese. After the operation, Searle is told that he can speak Chinese. Searle laughs and says, "Don't be silly." But the scientist's explanation was in Chinese! So, Searle understands Chinese but doesn't know it.

51 David Cole, 1984 **The brain-transplant thought experiment.** Part of a Chinese speaker's brain is transplanted into Searle's skull. Searle finds himself able to understand Chinese. But when someone asks him in Chinese if Searle knows anyone who speaks no Chinese, monolingual Chinese speakers report that, not only does Searle speak Chinese fluently, but that he is insisting, in Chinese, that he speaks no English!

47 David Cole, 1991 **The man doesn't understand Chinese but a virtual person may.** Maloney is right that the man who internalizes the room doesn't understand Chinese. However, there might be a virtual person in his brain who does understand Chinese. There could even be 2 separate virtual people in the same brain, each of whom speaks a different language. The fact that the man can't speak the 2 virtual people and the man who internalized them) can translate their own language into the others' languages says nothing about whether the virtual people speak and understand their own languages.

48 Steiner Bringsjord, 1992 **Cole confuses persons with personalities.** Multiple virtual personalities may exist in one brain, but multiple personalities are A psychiatric patient with multiple personalities is not a person, but possessed by a variety of virtual persons, but rather by a psyche fractured into a variety of personalities.

49 David Cole, 1991 **The Korrekoe Room thought experiment.** Imagine that the man in the room internalizes 2 rule books. One gives answers like those of a young Korean man, and the other like those of an old Chinese woman. Hence it is a Korean/Chinese or "Korrekoe" Room that is internalized. Neither of the 2 virtual people that result understands the other's language, and the man who internalizes the room only understands English. The fact that the man understands neither Korean nor Chinese is irrelevant to whether the virtual people understand their respective languages.

Can translations occur between the internalized Chinese Room and the internalizing English speaker?

vir · tu · al per · son: A formal implementation of the structure of human understanding. A virtual person is like a virtual machine, which is a machine that emulates another machine by instantiating its formal structure. For example, an old Atari video game console can be run as a virtual machine on a Dell laptop. Programs emulate

Do Chinese Rooms instantiate programs?

28 **The man in the Chinese Room doesn't instantiate a program.** A human being (or a homunculus) shuffling symbols in a room is not a proper instantiation of a computer program, and so the Chinese Room argument does not refute AI.
Note: For more multiple realizability arguments, see the "Is the brain a computer?" arguments on Map 1, the "Can functional states generate consciousness?" arguments on Map 6, and sidebar, "Formal Systems: An Overview," on Map 7.

29 H. J. Nelson, 1989 **The proper algorithm is constitutive of thought.** There is no hermeneutics involved in human thought, so Searle's Chinese Room does not implement a program in the man's head. The man's job is just to do the algorithms constitute the thought of the agent; they are not algorithms run by an agent that already has its own thoughts.

30 Jerry Fodor, 1991 **Proper instantiations require the right causal connections.** The Chinese Room argument exploits a vague definition of instantiation that allows any series of events to be interpreted as instantiations, so long as they are related to each other in the proper sequence. However, this concept allows to some very strange instantiations. For example, if the events of the program are spaced out over thousands of years and vast distances in space and still be an instantiation, the computer opponents aren't what we intend by computational instantiations. A better notion of instantiation would also require that a state be part of a program only if it is part of a somewhat tight causal way. Because this kind of homuncularism is not exhibited in the Chinese Room, the Chinese Room is not an instantiation of a program.

For next step in the algorithm, check Alpha Centauri in 1,000 years.

31 John Searle, 1991 **Conscious agents can instantiate computer programs.** Fodor's straightforward notion of instantiation does not rule out the Chinese Room as an instantiation of a program. Computers are causal mechanisms. The causal connections that Fodor asks for are supplied by the man in the room. The idea that conscious agents can't instantiate computer programs is inconsistent with the Turing test, since the Turing himself wouldn't have known what Turing machine was, given that some of his examples specifically involved conscious agents going through the steps of the program.

32 Aaron Sloman, 1986 **Implementations of computer programs must perform reliably.** The man in the Chinese Room can't reliably implement a program, since he is prone to make mistakes and lapses. The man might get tired, bored, or distracted, and thereby not follow the rule book correctly. A good implementation must exhibit "a strong causal connection between physical implementations of programs and the behavior they produce." Because the Chinese Room fails in this respect, it is not a proper implementation of a computer program.

zzzzzzz ...

33 Richard Sharvy, 1983 **Computers and humans run programs differently.** A man instantiating a chess program without knowing what he is doing is not playing chess. However, a computer doing the same thing is playing chess. It seems reasonable to suppose that the situation is similar in the case of speaking Chinese: a man running through the steps of a program without knowing what it is doing is not speaking Chinese, whereas a computer doing the same thing is speaking Chinese.

34 David Cole, 1984 **Computers embody programs; they don't obey them.** Searle misinterprets what it means for computers to obey programs the way the man in the room obeys a rule book. Computers embody programs in a causal way, a falling rock doesn't obey the law of gravity but embodies it.

35 Lilly Marlene Rosone, 1984 **Simulation of programs requires duplication of functional interconnections.** Genuine simulation of a computational system must accurately duplicate the functional interconnections within that system. The presence of the man in the Chinese Room introduces a homunculus that fails to duplicate the functional interconnections found in a computational system. Therefore, the Chinese Room is not a genuine simulation of a program.

Other Chinese Room arguments

52 Joseph Rychlak, 1991 **The Chinese Room cannot process symbols predicationally or oppositionally.** The Chinese Room lacks understanding because it cannot operate on meaningful symbols. In particular, it cannot process symbols in a way that involves affirmation and negation (predication), nor can it "reason from what is the case to what is not the case" (opposition). All the Chinese Room can do is shuffle symbols around (mediation).
Note: For elaboration on these concepts, see the "Dialectical" arguments on Map 3.

53 Anticipated by John Searle, 1980a, 1980b, 1990b **The many mansions reply.** The problems raised by the Chinese Room argument only exist because of the present state of AI and computer science. We will be able to build devices that reproduce the causal processes that are involved in intentionality. At that time, artificial intelligence will be able to explain

54 John Searle, 1980a, 1980b, 1980b **The many mansions reply trivializes strong AI.** By expanding the project of AI to include any device that can produce intentionality, the many mansions reply trivializes the project of strong AI. It is entirely possible that someday some device will possess intentionality, but it will not be because it has a Chinese Room architecture it must to realize the more specific thesis that formal computational processes can produce intentionality.

55 Lawrence Carleton, 1984 **The Chinese Room is more than a simulation.** In a mere simulation, the output of a computer is of the same type as the phenomenon being simulated. For example, a fire simulator doesn't take wood as input or generate heat as output, so a fire simulator is a mere simulation. But in the Chinese Room, the input and output are Chinese symbols, the same as those used by a real Chinese speaker. Therefore, the Chinese Room is more than a mere simulation of Chinese speaking. **Note:** Also, see the "If a simulated intelligence passes, is it intelligent?" arguments on Map 2.

Where should I put the wood?

56 Patricia Hanna, 1985 **Searle is trapped in a dilemma.** The Chinese Room can't be as Searle describes it because it either has semantics or it doesn't, and neither alternative is allowed.

Either	Or
The Chinese Room uses an integrated syntactic-semantic representation of Chinese-speaking behavior. *in which case* the room is not purely syntactic, as Searle claims it is.	The room works purely on syntactic rules. *in which case* it won't produce fluent Chinese-speaking behavior, because Searle's rules are based on the assumption that beyond specified syntactic rules (like those contained in the rule book),	

In Either Case
The room is not as Searle describes it.

57 J. Jonathan Cohen, 1986 **The man in the room will understand some Chinese questions.** Speaking fluent Chinese requires an ability to answer basic mathematical questions, and the man will do the same in Chinese as they are in English. Then, the man in the room will recognize some mathematical parts of Chinese. In such cases, the man understands Chinese and is not merely manipulating symbols.

Aha! I plus 1 equals 2.

58 Dale Jacquette, 1990 **The Chinese Room argument is circular.** The Chinese Room argument is supposed to show that syntactic symbol manipulations can never yield semantic meanings. But to show that a formal system doesn't yield semantics, Searle assumes that the man in the room performing syntactic operations understands no Chinese. But to do this Searle must again assume that syntax can yield semantics.

Start Here
To prove that
which assumes that
syntax does not imply semantics
doing syntactic manipulations does not give the man in the room an understanding of Chinese
you assume that

59 James Moor, 1988 **The pseudorealization fallacy.** Fantastic realizations of the computational theory of mind, such as Searle's brain in a Chinese Room, are irrelevant to empirical psychology. Just because we can imagine a computer simulating a mind, it doesn't follow that it is a real possibility. For instance, although by some stretch of the imagination a stomach can be thought of as an information processor, such a realization is likely to be ad hoc. The stomach must consistently exhibit the right sort of input-output relations to be a plausible candidate for a digestive system. Similarly, it doesn't follow that just because a stomach exists a new computational state is required. Note: For more multiple realizability arguments, see the "Is the brain a computer?" arguments on Map 1, the "Can functional states generate consciousness?" arguments on Map 6, and sidebar, "Formal Systems: An Overview," on Map 7.

The empirical content of a theory limits the kinds of realizations which are possible (p. 49).

James Moor

60 James Moor, 1988 **The man in the Chinese Room is an alien.** The Chinese Room argument relies on the assumption that the man in the room carries out instructions of a computer program at the second with perfect accuracy. But no human being, even in principle, could do this. Because the man in the Chinese Room is not a human, but a cyborg, the man in the Chinese Room must actually be a kind of alien possessing a "mix of superhuman capacities and subhuman ... needs and desires" (p. 40).

61 E. J. Nelson, 1989 **Chinese-speaking behavior is too limited to serve as a counterexample to strong AI.** Searle's Chinese Room only displays a small subset of intentional human behaviors, namely, Chinese hookup, and thus Searle's model is complete only if it accounts all counterfactuals of intentionality. For example, it doesn't model the ability to abstract universals from particulars, to recognize the same object in a variety of circumstances, to generalize from instances, to recognize contrasting concepts, and so on. A system that lacks these abilities would not be able to think, but the Chinese Room doesn't have these abilities and so it isn't a real counterexample to its conceptual intentionality.
Note: Also, see "The Test Is not Narrow," Map 2, Box 5.

62 Steve Harnad, 1989 **The Chinese Room makes a modularity assumption.** Searle's argument makes an untested modularity assumption. It presupposes that certain human cognitive capacities, such as the ability to speak Chinese, can be modeled independently of the other functional parts of the mind. Evolutionary biology suggests that language cannot be modeled independently of other communicative capacities, because linguistic capacities never exist without these evolutionarily more primitive capacities.

63 Hiroshi Ben-Yami, 1993 **There are questions the Chinese Room can't answer.** Because the man in the Chinese Room follows the rules of rules, the Chinese Room can never answer questions about changing situations. For example, it can't answer questions like, "What time is it?" or "What day is this, today?" Because answering such questions is part of ordinary linguistic competence, the Chinese Room can never simulate Chinese.

Can computers have the right causal powers?

supporting arguments

64 John Searle, 1980a, 1980b, 1990b **Understanding arises from the "right causal powers."** Any system capable of semantic understanding and intentionality must have (at least) the causal powers of a human brain. We know that brains are able to produce understanding. It is an open empirical question whether other materials, like silicon, can have sufficient causal powers to produce understanding.

I have the right causal powers.

65 Philip Cam, 1990 **Sufficient causal powers don't imply necessary and sufficient causal powers.** Searle mistakenly claims that the causal powers of the brain are sufficient for intentionality. From that initial sufficiency claim, Searle wrongly concludes that anything with intentionality must have causal powers equivalent to those of the brain. But we can imagine many systems that are sufficient to produce intentionality but that aren't equivalent to brains.

66 Karl Pfeifer, 1992 **Cam misunderstands the sense of "right causal powers."** Searle doesn't claim that anything that produces intentionality must be equivalent to a brain. He just claims that there is some set of causal powers (brain-based or otherwise) that is necessary for understanding, and that formal computer programs lack that set of causal powers.

67 John Haugeland, 1980 **A Chinese Room-style argument shows that causal powers are insufficient for understanding.** Searle's claim that the right causal powers are sufficient for understanding can be refuted by his own Chinese Room. If a man lacks the relevant causal power in a Chinese speaker's brain (for example, neurotransmitters) with a demon that simulates that causal power's function without understanding, then the fact that having the "right causal powers" is not sufficient for understanding.

That demon has the right causal powers, but it doesn't understand Chinese.

That person's brain still has the right causal powers; it just needed some help from the demon.

68 Jack Searle, 1980a, 1980b, 1990b **Causal powers are maintained if the simulation is low-level enough.** The person whose neurotransmitters have been replaced still understands Chinese because she still has the right causal powers; she still has a brain that causes her to understand. The simulation is low-level enough that the relevant causal effects—the physical stimulation of neurons—are still initiated, and so intentionality is maintained.

69 Steven Sayid, 1982 **The series of Chinese simulators.** We can imagine a series of thought experiments ranging from a case in which a demon replaces one of the speaker's neurons to a case in which a demon replaces his entire nervous system. Searle's claim that at some point intentionality has been replaced the speaker will still understand Chinese, such a simulation would be very low-level. As more neurons are replaced, then the degree of simulation increases. At what point does the person cease to understand Chinese? No clear line can be drawn here, so Searle's claim about low-level simulations won't help him.

70 Lawrence Carleton, 1984 **Searle commits the fallacy of denying the antecedent.** Searle makes an invalid argument by drawing a bad conclusion from a premise.
Premise If certain brain process equivalents are present, then they produce intentionality.
Premise Formal computer programs do not have these equivalents.
But the premises do not imply the conclusion.
Conclusion Therefore, computer programs do not have intentionality.
That is, Searle commits the fallacy of denying the antecedent.

Denying the Antecedent
The fallacy of denying the antecedent is a bad inference of this form:
(1) if *x*, then *y*. (2) Not *x*. Therefore, (3) not *y*.

	Example		Explanation
1A	If a rock is thrown at a window, the window will break.	1B	The antecedent is the first part of the if-then statement (1), in this case that a rock is thrown at a window.
2A	It is not the case that a rock was thrown. It is just falling there.	2B	The denial of the antecedent (2) is the denial of *x*, in this case, the denial that a rock was thrown at a window.
2A	Therefore, the window was not broken.	3B	The conclusion that the window was not broken contains the fallacy. It does not follow from the combination of statements 1A and 2A that the window was not broken, because the window may have been broken by something other than a rock, for example, by a gun blast, a punch, and so forth.

71 Dale Jacquette, 1989 **Carleton misreads Searle.** Carleton misreads Searle as saying that brain processes or something as powerful as a brain are necessary for intentionality. Searle's claim is that something as powerful as a brain must have been produced by a brain or by something as powerful as a brain.
Note: Jacquette also thinks that Searle's right causal powers argument is faulty (see "The Brain's Causal Powers Can Be Reproduced by a Computational System," Box 73), but he doesn't think that Carleton makes a good case against Searle.

72 Richard Double, 1983 **Programming may be necessary for understanding.** Even if Searle is right that programs by themselves are not sufficient to produce intentionality, they still might contribute to the causal powers that underlie intentionality. Imagine a case in which 2 brains have evolved, 1 to speak Chinese and 1 to speak English. Scientists would explain their linguistic differences in terms of underlying neurological differences. Suppose that one of the brains, the Chinese one, has evolved to have some of the linguistic differences to result from programming as well.

73 Dale Jacquette, 1989 **The brain's causal powers can be reproduced by a computational system.** A computational system that could duplicate the microlevel functional structure of the brain would have the brain's causal powers. If, as Searle says, it is these causal powers that give rise to intentionality, then such a system would possess intentionality.

Searle's arguments presuppose a dualism of control, like that of DEC.

74 John Searle, 1980 **The level of formal program can produce semantics.** Jacquette thinks I argue that no program could ever be a mind. But I never make that claim. A program could be a mind if it were implemented in a machine with the right causal powers. What I argue is that no program, no matter how low level it is, could ever produce semantics which by virtue of syntactic symbol manipulations. So Jacquette's appeal to microlevel isomorphism won't help.
Supporters of
"The Chinese Water-Pipe Brain Simulator," Box 5.

75 Dale Jacquette, 1989 **Searle assumes a central locus of control.** Searle presupposes that any simulation of the brain must be a simulation of the brain's central locus of control that manipulates symbols without understanding Chinese. But consider a model of a Chinese speaker that lacks any such locus of control (for example, a version of a connectionist network that models the brain's unconditional functional structure). Searle has not shown that such a model would not understand Chinese.

76 John Searle, 1989 **Locus of control is not assumed by the Chinese Room argument.** The assumption of a central locus of control is a misreading of the Chinese Room argument. I give several examples of programs (beer-can computers, water pipes, etc.) that don't need a central locus of control. My only assumptions are that "programs are formal (syntactic), and minds have semantics, and minds have semantics."
Note: For more on these assumptions, see the "System-Semantics Barrier" arguments on Map 5, and Dale Jacquette's response to these assumptions, see "The Chinese Room Argument is Circular," Box 58.

77 Dale Jacquette, 1990 **Locus of control is never abandoned by Searle.** Searle's Chinese Room is a system without some locus of control. In the Chinese water-pipe example there is a man controlling the faucets. In the Chinese Room there is a man who remembers the rules in the bankers, and so forth, and who manipulates the symbols in his head.

Dale Jaquette

Is biological naturalism valid?

supporting arguments

78 John Searle, 1980a, 1980b, 1990b **Biological naturalism.** Consciousness and intentionality are caused by brain processes. The brain has the right causal powers to produce intentionality.

79 Klaus Obermeier, 1983 **Understanding is intentionality. Understanding** requires an assumption about biological or internal processes, but is inferred from behavioral performance. Handling linguistic information by relying on context and shared assumptions is sufficient for understanding.

They display linguistic behavior without understanding.

French — Swahili

80 Klaus Obermeier, 1983 **Understanding of things doesn't require having feelings.** If a computer can handle linguistic information, then it can understand those words. We aren't required to justify our first-person emotional reports, so why should a computer be required to?

Of course I understand love. Those people are in love ...

81 Dale Jacquette, 1989 **Intentionality is irreducibly abstract.** Intentionality is irreducible to biological phenomena, just as space-time is irreducible to the physical phenomena it organizes. The space-time, such of which is primitive and fundamentally abstract.

82 John Searle, 1989 **Intentionality is both abstract and biological.** Jacquette claims that I misunderstand the abstract nature of intentionality and that I place too much emphasis on biology. But Jacquette fails to recognize that intentionality is a biological phenomenon. The abstract nature of intentionality is the result of a linguistic dualism that distributes the abstract features of intentionality.

83 David Cole, 1984 **Searle's materialism is problematic.** Searle's claim that the right causal powers and right causal powers claims commit him to a form of materialism. This materialism is problematic because we can imagine cases in which the brain has the biological causal powers to understand Chinese, but fails to do so.

84 David Cole, 1984 **The superneuron thought experiment.** Suppose a conscious neuron internalizes a map of its whole surrounding neural system along with transfer functions for each other neuron. Its nucleus swells with information, and it becomes a superneuron. It gradually replaces all the other central nervous system neurons, and it becomes the brain. It is the superneuron that thinks and understands, not the brain. It is the result of a lingering dualism that distributes the abstract features of intentionality.

Can computers cross the syntax–semantics barrier?

85 John Searle, 1990b **The syntax–semantics barrier.**
Axiom 1: Computer programs are formal (syntactic).
Axiom 2: Human minds have mental contents (semantics).
Axiom 3: Syntax by itself is neither constitutive of nor sufficient for semantics.
Conclusion 1: Programs are neither constitutive of nor sufficient for minds (p. 27).

You can't get there from here.

Syntax — Semantics

86 **The syntax–semantics barrier can be overcome by programs that learn.** If a computer *is given* the ability to learn, it can generate semantics from low-level syntax.
Note: Also, see the "Can physical symbol systems learn as humans do?" arguments on Map 3.

87 Christopher Maloney, 1987 **Learning programs don't help.** The syntax–semantics barrier is not overcome by complex learning programs, because a program doesn't understand anything. Making a program complex won't give it understanding either.

That liver is complicated, but it doesn't understand anything.

Liver

88 Paul Thagard, 1986 **Semantics can arise from programs that learn.** A system with sufficiently complex learning mechanisms can possess semantics. Such a system must meet a principle of indicative adequacy, which states that the system should possess inductive mechanisms that are capable of producing all the knowledge constructs that the system initially lacks. If a system of learning mechanisms can generate new rules, a frame-based system must be able to generate new frames, and so forth.

89 William Rapaport, 1988 **Internal semantics is realized in syntactic networks that learn.** A network of appropriately connected syntactical symbols, which are linked by the consequences from objects, possesses internal semantics. Such a system can be said to understand because human understanding is just a complex form of internal semantics.

William Rapaport

90 Noel Jabren, 1990 **Human understanding is not reducible to syntactic semantics.** Even if a machine that learns, it can possibly have understanding in the human sense. Primitive categories are necessary for understanding, and a program cannot generate these categories by itself.

91 Neil Jabren, 1990 **The Chinese record-book thought experiment.** Imagine that the man in the room is given record books in which he can write down Chinese Chinese symbols, and that his manual has been revised as follows.
- He can record specific symbols he gets from outside.
- He can modify or delete an entry in his record book on the basis of symbols he gets from outside.
- He can write new symbols on the records with one another and with symbols passed to him from outside.
- He can formulate strings based on this correlation of symbols.

The man in this modified Chinese Room can accumulate more and more symbols and can, in that s