HOW TO PASS A TURING TEST AND ESCAPE FROM THE CHINESE ROOM

William J. Rapaport

SNePS Research Group (SNeRG)
Dept. of Computer Science & Engineering

Department of Philosophy

Center for Cognitive Science

rapaport@cse.buffalo.edu
http://www.cse.buffalo.edu/∼rapaport/

Philosophical implication of
computational cognitive science
(symbolic or connectionist):

If mental states and processes can be expressed as algorithms,
then they are capable of being implemented in non-human computers.

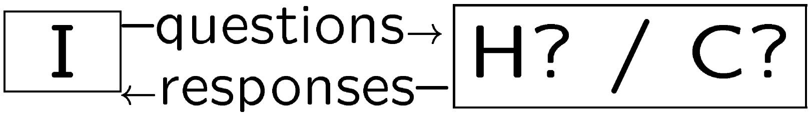
Are computers executing such algorithms merely simulating mental states and processes, or are they actually exhibiting them?

Do such computers think?

Answer: Turing's Test

Objection: Searle's Chinese-Room Argument

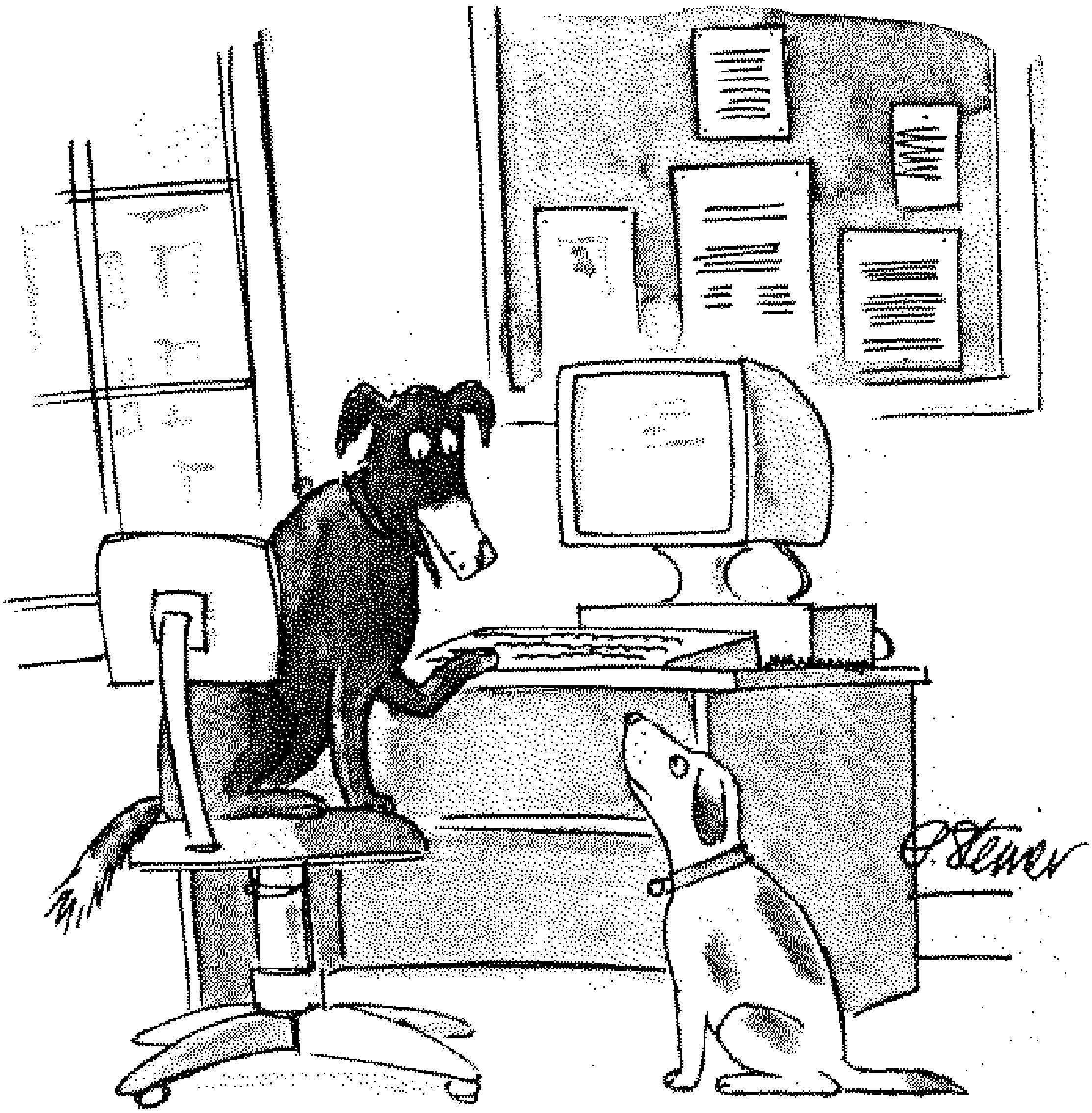
The Turing Test



“I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.” (Turing 1950)

Thinking vs. “Thinking”

61



“On the Internet, nobody knows you're a dog.”

© *The New Yorker*, 5 July 1993

Thinking vs. "Thinking", cont'd

- Cartoon works because:
  one does *not* know with whom one is communicating via computer.

- Nevertheless, we assume we are talking to a human
  (i.e., entity w/ h. thinking capacities)
  = Turing's point, namely:

- Argument from Analogy:

  - solution to Problem of Other Minds
    = I know I think;
    how do I know you do?

  - you are otherwise like me

  - ∴ (probably)
    you are like me w.r.t. thinking

Thinking vs. "Thinking", cont'd

- What's wrong with Arg't. from Analogy:

– I could be wrong about whether you're biologically human.

- Am I wrong about your thinking/ (human) cognitive abilities?
  Turing: No.

- More cautiously, whether I'm wrong depends on def of (human) cognitive abilities/thinking

  $=_{df}$? passing TT /∴ TT-passer thinks by def.

  $=_{df}$? XYZ /∴ if TT-passer satisfies XYZ,
  then TTP thinks
  or ¬XYZ
  or TTP only superficially sats XYZ
  but doesn't really think (cf. CRA)
  or TTP "thinks"
  (metaphorical or extended sense)

Thinking vs. “Thinking”, cont’d

- Birds fly.
- Do people fly?
- Do planes fly?
  - don’t flap wings

Thinking vs. “Thinking”, cont’d

But planes do fly:

- metaphorical extension (Lakoff/Johnson)
  - planes fly = planes “fly”
  - “use of words” changed

- flapping wings not essential to flying
  - physics of flight is same for birds & planes; more gen’l, abstract theory of flying
  - “general educated opinion” changed

NB: Use of ‘fly’ & gen’l educated opinion
have changed

- spaceships “fly”; planes fly!
- single abstr. theory can account for metaphorical extension

Thinking vs. “Thinking”, cont’d

1950: computers were (only) human

2000: computers are (only) machines

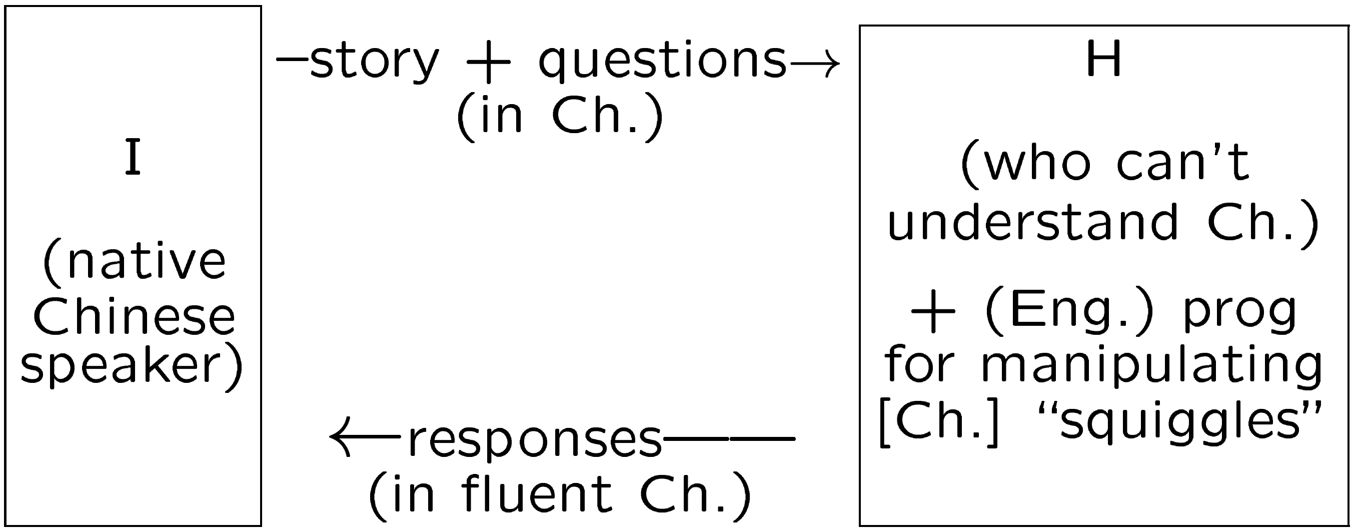
general educated opinion:
computers are viewed not as implementing devices but in functional, I/O terms

- Ditto for ‘think’ (see later)

- BUT: it’s not really thinking (?)

The Chinese-Room Argument

It's possible to pass TT, yet not (really) think



I

(native Chinese speaker)

−story + questions→
(in Ch.)

H

(who can't understand Ch.)

+ (Eng.) prog for manipulating [Ch.] "squiggles"

←responses——
(in fluent Ch.)

The Chinese-Room Argument(s)

Argument from biology:

(b1) Computer programs are non-biological.
(b2) Mentality is biological.
(b3) ∴ No non-biological computer program
can exhibit biological mentality.

Argument from semantics:

(s1) Computer programs are purely syntactic.
(s2) Mentality is semantic.
(s3) Syntax alone is not sufficient for semantics
(s4) ∴ No purely syntactic computer program
can exhibit semantic mentality.

X

- Cognition can be characterized abstractly & implemented in different media

X

- Syntax suffices for semantics
- Better: Try to build C.R.
  - What's needed for NLU?

"I [Searle] still don't understand a word of Chinese

and neither does any other digital computer
because all the computer has is what I have:

a formal program
that attaches no meaning, interpretation, or content
to any of the symbols.

Therefore, no formal program by itself
is sufficient for understanding."

N.B.: A program that did attach meaning, etc., might understand!

BUT: Searle denies that, too →

“I see no reason in principle why we couldn’t give a machine the capacity to understand English or Chinese, since in an important sense our bodies with our brains are precisely such machines.

But we could not give such a thing to a machine whose operation is defined solely in terms of computational processes over formally defined elements.”

. . . because:

“Only something having the same causal powers as brains can have intentionality” (i.e., mental states and processes).

“These causal powers are due to the (human) brain’s biological (i.e., chemical and physical) structure”

. . . namely. . . ?

A simulated human brain

“made entirely of old beer cans rigged up to levers and powered by windmills”

would not really exhibit intentionality
even though it appeared to.

Why must intentionality be biological?

Searle:

> Only biological systems have the requisite causal properties to produce intentionality.

What are the causal powers?

Searle:

> The ones that can produce perception, action, understanding, learning, and other intentional phenomena.

Isn't this a bit circular?

Possible clue to what the causal powers are:

“Mental states are both

- caused by the operations of the brain
- realized in the structure of the brain”

i.e., implemented in the brain.

“Mental states are as real as any other biological phenomena, as real as lactation, photosynthesis, mitosis, or digestion.

Like these other phenomena, mental states are caused by biological phenomena and in turn cause other biological phenomena.”

- Searle’s “mental states” are implementations of abstract mental states.

1. “Intentional states are both caused by and realized in the structure of the brain.”

BUT: Brains & beer-cans/levers/windmills can share structure

$\therefore \neg 1$

2. “Intentional states are both caused by and realized in the neurophysiology of the brain.”

$\therefore$ **3:** “Intentional states stand in causal relation to the neurophysiological”

4. “Intentional states are realized in the neurophysiology of the brain.”

I.e., they are implemented in the brain.

ADT	**Implementation**
stack	array
natural numbers (Peano axioms)	any sequence of items satisfying Peano axioms
musical score	performance
play script	performance
liquid	water
liquid	alcohol
mental states	some brain states/processes
mental states	some computer states/processes

Summary of biological argument:

Searle:

– understanding is biological;

∴ human brain can understand Chinese;

– BUT: computer running Chinese NL
program cannot understand Chinese

Rapaport (et al.):

– on abstract, functional, computational notion of understanding as an ADT, understanding can be implemented in human brain & computer

∴ both can understand

Argument from Semantics

(S1) Computer programs are purely syntactic.

(S2) Cognition is semantic.

X (S3) Syntax alone is not sufficient for semantics.

(S4) $\therefore$ No purely syntactic computer program can exhibit semantic cognition.

$\neg$ (S3): Syntax suffices for semantics

Syntactic Semantics:
How syntax can suffice for semantics

SS 1: Semantics (relations between symbols & meanings)
→ syntax (relations among symbols & internalized (symbolized) meanings)
∴ syntax can suffice for semantic interpretation

SS 2: Semantics is recursive:

- We understand syntactic domain in terms of antecedently understood semantic domain
- base case: syntactic understanding (cf. proof theory)

SS 3: Internal, narrow, 1st-person POV is what's needed for understanding/modeling cognition

Searle:

Syntax can't suffice for semantics
because links to external world are missing

2 assumptions:

- computer has no links to external world (solipsism?)
- external links are needed to attach meanings to symbols

  (But, if so, then computer can have them just as humans do)

Semiotics

- Given: A system of “markers”
  (“symbol system”)

- Syntax: Study of relations among markers
  - grammar, proof-theory
  - no relations between markers & non-markers

- Semantics:
  Study of relations between markers & “meanings”

On this theory, should be clear that
syntax can’t/doesn’t suffice for semantics

Pragmatics:

- Study of relations between markers & interpreters
- Study of relations among markers, meanings, & interpreters

Semiotics, cont'd

if set of markers is unioned with set of meanings

& if union is taken as a new set of markers

(i.e., mngs are internalized in the symbol system)

then what was once semantics

(relations between old markers & meanings)

becomes syntax

(relations among new markers)

(& thus syntax can do the job of semantics)

Aside: Linguistic semantics is like this:

study of synonymy, antonymy, entailment, etc.

w/o invoking "meanings" or "external entities"

Syntactic Semantics I:
Turning Semantics into Syntax

- Can the semantic domain be internalized?
  - Yes: under the conditions obtaining for human language understanding

- How do we learn the meaning of a word? How do I learn that ‘tree’ means tree?
  - by association . . . (of tree w/ ‘tree’? No!)
    - * . . . of my internal representation of ‘tree’ with my internal representation of a tree

- internal representation could be activated neurons
  (binding of multiple modalities)

- Ditto for computer (Cassie):
  - I say something to C. in English
  - C. builds internal nodes representing my utterance
  - I show pictures to C.
(or: C. sees something)
  - C. builds internal nodes representing what she sees
  - These 2 sets of nodes are part of same KB (semantic network)
- Ditto for formal semantics:
  syntax & semantics are both defined syntactically

Points of View

- To understand how a cognitive agent understands,

  and to construct a computational cog agent,

  we must take 1st-p. POV

  - what is going on “in” agent’s head, from agent’s POV

- Don’t need to understand causal/historical origins of internal symbols

- Searle-in-CR’s POV vs. interrogator’s POV:

  - CRA:
    S$_{cr}$’s POV trumps interrogator’s POV

  - TT & SynSem:
    interrogator’s POV trumps S$_{cr}$’s POV

(From Baum, *Wizard of Oz*, 1900: 34–35.)

When Boq saw her silver shoes he said,

“You must be a great sorceress.”

“Why?” asked the girl.

“Because you wear silver shoes and have killed the wicked witch. Besides, you have white in your frock, and only witches and sorceresses wear white.”

“My dress is blue and white checked,” said Dorothy, smoothing out the wrinkles in it.

“It is kind of you to wear that,” said Boq. “Blue is the color of the Munchkins, and white is the witch color; so we know you are a friendly witch.”

Dorothy did not know what to say to this, for all the people seemed to think her a witch, and she knew very well she was only an ordinary little girl who had come by the chance of a cyclone into a strange land.

Is Dorothy a witch? – D's POV: no – Munchkin POV: yes	Does S_{cr} understand Ch? – S_{cr}'s POV: no – native Ch spkr's POV: yes
D bels she's not a witch (as she understands 'witch')	S_{cr} bels that he doesn't understand Ch (as he understands 'understands Ch')
$\Diamond$[Witch(D) & Bel_D(¬Witch(D))]	$\Diamond$[U(S_{cr}, Ch) & $\mathrm{Bel}_{S_{cr}}$(¬U(S_{cr}, Ch))]
What counts as being a witch? – dispute isn't re whether D is "really" a witch in some context-indep. sense – dispute is re whether Witch(D) in Munchkin sense (Munchkin POV)	What counts as really understanding Ch?
M POV trumps D's POV	native spkr's POV trumps S_{cr}'s POV

More on CR situation:

But S$_{cr}$ could insist that he doesn't
understand Ch

Cf.: I bel that I understand 80% of French
& can express myself 75%
but always feel I'm missing something

Should I bel native Fr spkr who says
I'm fluent?

Searle: No

The Systems Reply

But $S_{cr} \neq$ me

S_{cr} can't insist that he alone doesn't
understand Chinese
& that $\therefore$ his POV trumps

because S_{cr} isn't alone:

- S_{cr} has instruction book (systems reply)

- S_{cr} + book, stranded on desert island,
  could communicate with
  native Chinese-speaking Friday

More on the Systems Reply

- Hutchins, "Cognition in the Wild"

- extended cognitive system (crew + instruments) that navigates ship is real-life counterpart to S$_{cr}$ + book

- "systems that are larger than an individual may have cognitive properties in their own right that cannot be reduced to the cognitive properties of individual persons" (Hutchins 1995)

- S$_{cr}$ + external book has cognitive property of understanding Chinese,
  even though S$_{cr}$ (simpliciter) lacks that property

POV, cont'd

- Cognitive agent has no direct access to external entities

- When I point to a tree,
  I'm aware of internal visual image of:
  my hand pointing to a tree

- Kant: phenomena vs. noumena

- My access to external world is mediated by internal representatives

  - Argument from Illusion:
    see different things with each eye

POV, cont'd

- 3rd-person POV:
  - you (can) have access to:
    - * external world
    - * my/Cassie's internal world
  - we both see same tree, no?
- NO: you have access to
  your internal representations of:
    - * external world
    - * my/Cassie's internal world

"Kant was rightly impressed by the thought that if we ask whether we have a correct conception of the world, we cannot step entirely outside our actual conceptions and theories so as to compare them with a world that is not conceptualized at all, a bare 'whatever there is.' " (Bernard Williams (1988): 40.)

$\therefore$ by merging internalized semantic markers
with internal syntactic markers,
semantic project of mapping meanings
to symbols
can be handled by syntax
(symbol manipulation)

$\therefore$ syntax suffices
for 1st-person semantic enterprise

Syntactic Semantics II: Recursive Theory of Semantic Understanding

Semantics ::= 2 domains + 1 binary relation:

- syntactic domain (markers) [SYN]
  (char'ized by syntactic formation/inference rules)
  semantic domain [SEM]
  (meanings, interpretations)
  semantic interpretation $\mathrm{I} : \mathrm{SYN} \rightarrow \mathrm{SEM}$

- We use SEM to understand SYN
  ∴ we must antecedently understand SEM

- How?

  - Treat SEM as a new SYN
    & find new SEM for it
    (correspondence continuum)
    $\mathrm{SYN}_1 \leftarrow \mathrm{SEM}_1 (= \mathrm{SYN}_2) \leftarrow \mathrm{SEM}_2 \leftarrow \ldots \leftarrow \mathrm{SEM}_n \hookleftarrow$

– base case:
understand "last" domain
in terms of itself,
viz., syntactically

i.e., we understand a domain syntactically by being conversant with manipulating its markers (or by knowing which wffs are thms)

– meaning of node is its location
(= relns to all other nodes) in network

– can constrain this to a subset;
yields theory of vocabulary acquisition

- I understand what you say by interpreting it i.e., mapping it into my concepts

- I (semantically) understand a purely syntactic formal system by interpreting it
– i.e., providing a (model-theoretic) semantics for it

Question:

What would it be for a formal system to understand me?

Answer:

By treating what I say as a formal system and interpreting it

N.B.:

– links to external world are irrelevant
– "semantic" interpretation of formal system is a syntactic enterprise

- NLU system S1 understands the NL output of NLU system S2 by building and manipulating the symbols of its (S1's) internal model (i.e., an interpretation) of S2's output considered as a formal system.

- S$_{cr}$ understands native Chinese speakers as I understand you:

  – by mapping internal representations
    of your utterances
    (considered as syntactic markers)
    to my internal symbols
    & then doing symbol manipulations
    (i.e., syntax)

∴ syntax suffices

- What is needed for (computational) NLU?

- Domain of inquiry:
  - understanding narrative text

Mind as Syntactic System

To understand, a cognitive agent must:

- take discourse as input
- understand ungrammatical input
- make inferences & revise beliefs
- make plans
  – for speech acts
  – to ask/answer questions
  – to initiate conversation
- understand plans
  – speech-act plans of interlocutor
- construct user model
- learn (re: world, language)
- have background/world/c.s. knowledge
- remember
  – what it heard, learned, inferred, revised

= have a mind!