Dreyfus on expertise

- Don’t make decisions
- Don’t solve problems
- Do what works
- No decomposition of situation into discrete elements
- Pattern recognition extends to plan as well as diagnosis
- know-how (skills and abilities) rather than know-that (facts, rules)

"One has to abandon the traditional view that a beginner starts with specific cases and, as he becomes more proficient, abstracts and interiorises more and more sophisticated rules. It turned out that skill acquisition moves in just the opposite direction – from abstract rules to particular cases" (Dreyfus and Dreyfus, *Mind Over Machine*, p108).

- "Current AI is based on the idea, prominent in philosophy since Descartes, that all understanding consists in forming and using appropriate representations. Given the nature of inference engines, AI's representations must be formal ones, and so common-sense understanding must be understood as some vast body of precise propositions, beliefs, rules, and procedures. Thus formulated, the problem has so far resisted solution. We predict it will continue to do so" (Dreyfus and Dreyfus, *Mind Over Machine*, p99).

Dreyfus identified several basic assumptions about the nature of human knowledge which grounded much cognitive science, certainly what he called GOFAI.

- A belief that the mind functions like a digital computer using symbolic manipulations – **psychological assumption** (163ff)
- A belief that computer programs could be understood as formalising human thought – **epistemological assumption** (189)
- An assumption about the data about the human world that we employ in thought, namely that it consists of discrete, determinate, and explicit pieces that can be processed heuristically – **ontological assumption** (206). http://ejap.louisiana.edu/EJAP/1996.spring/wrathall.kelly.1996.spring.html

John Searle (1932-)

- Wrote an influential paper called “Minds, Brains, Programs” in *Behavioral and Brain Sciences* in 1980, accompanied by various replies.
- The Chinese Room thought experiment has come to be one of the most widely discussed and disputed in philosophy. It is an argument against “Strong AI”, but also much of cognitive science, functionalism, computationalism, and behaviorism

If you see this shape,

"什 麼 "

followed by this shape,

"帶 來 "

followed by this shape,

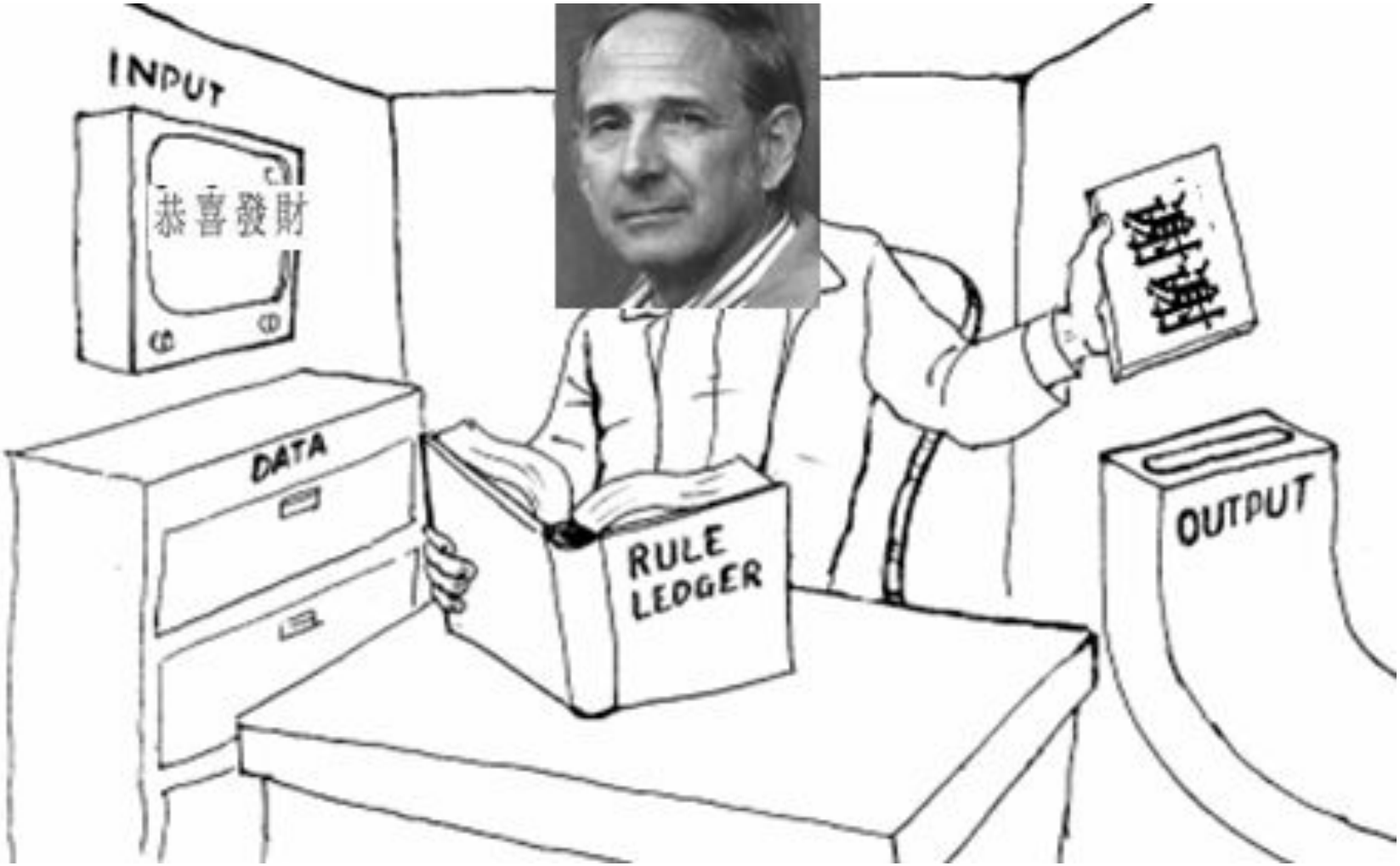
"快 樂 "

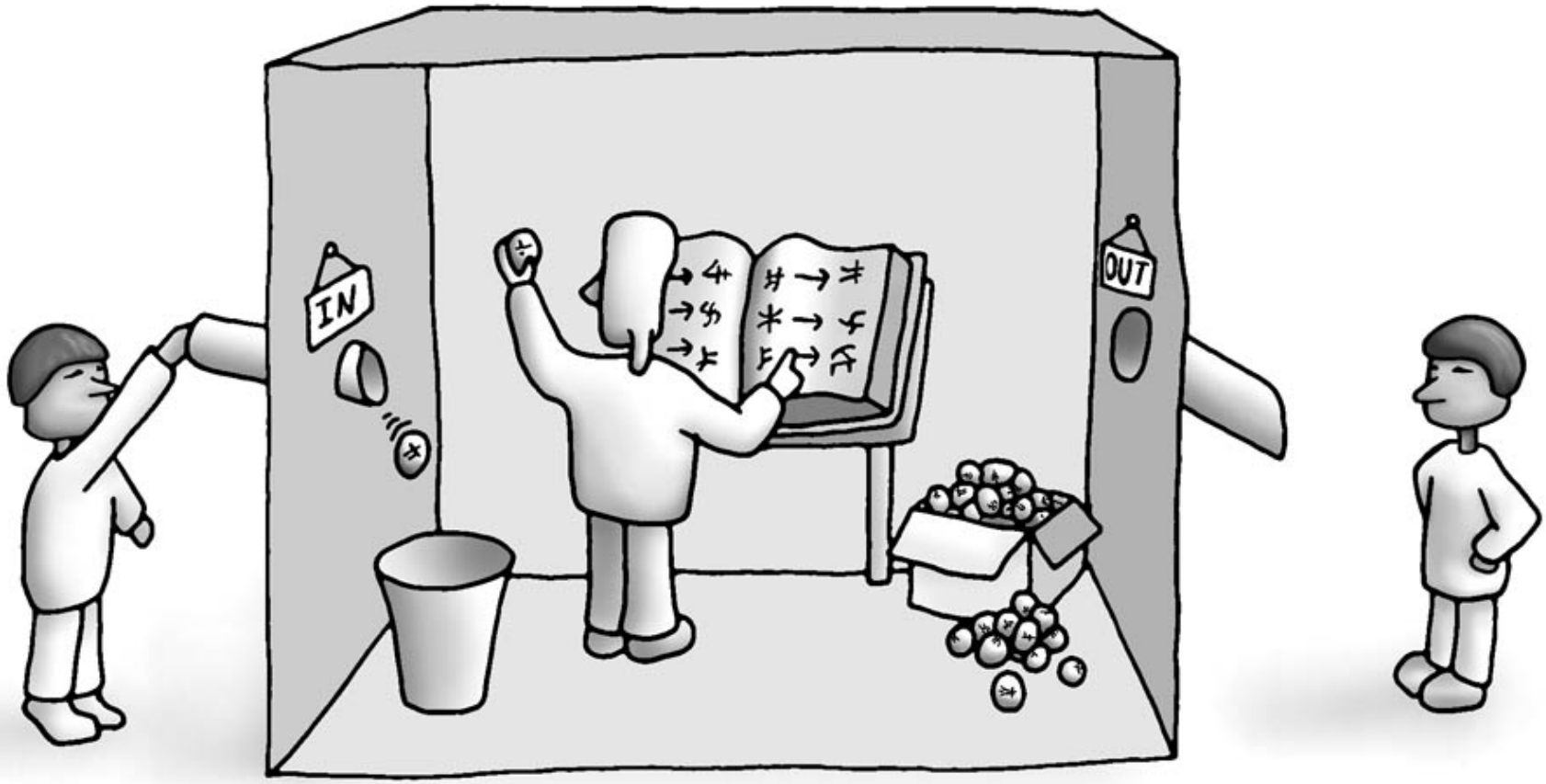
then produce this shape,

"爲 天 "

followed by this shape,

"下 式 ".


INPUT
恭喜發財
DATA
RULE
LEDGER
OUTPUT



jolyon.co.uk

“Imagine a native English speaker who knows no Chinese locked in a room full of boxes of Chinese symbols (a data base) together with a book of instructions for manipulating the symbols (the program). Imagine that people outside the room send in other Chinese symbols which, unknown to the person in the room, are questions in Chinese (input). And imagine that by following the instructions in the program the man in the room is able to pass out Chinese symbols which are correct answers to the questions (output). The program enables the person to pass the Turing Test for understanding Chinese but he does not understand a word of Chinese” (Searle, 1999)

Chinese room argument

(1) John cannot understand Chinese merely because he is able, given relevant input symbols, to manipulate them and produce output symbols according to rules he is given.

(2) All a computer can do is, given relevant input symbols, to manipulate them and produce output symbols according to the rules it is given.

(C_1) So a computer can never understand Chinese.

(3) If the classical cognitive science program is right, the strong AI project is possible in principle.

(4) If strong AI is possible in principle, a computer can understand Chinese.

(C_2) So the classical program is wrong.

Three kinds of reply

1. Some critics concede man in room doesn't understand Chinese, but hold that there is some other thing that does understand (the whole system, say).
2. Others concede Searle's point but argue that a variation on computer system could understand (robot)
3. Others deny Searle's claim that the man in room necessarily does not understand Chinese (perhaps our intuitions are unreliable, say, or it depends what we mean by 'understand' - or problem of other minds).

Searle material

- http://library.latrobe.edu.au/search~S5?/aSearle%2C+John/asearle+john/1%2C2%2C28%2CB/frameset&FF=asearle+john+r&14%2C%2C27 (library e-reserve)
- http://www.iep.utm.edu/c/chineser.htm
- http://plato.stanford.edu/entries/chinese-room/
- See Copeland, J., *The Philosophy of Artificial Intelligence* (reserve)
- See Crane, T., *The Mechanical Mind* (reserve)

The image of John Searle on slide 4 is from Wikipedia

The image of the Chinese Room is by Jolyon Troscianko, website: jolyon.co.uk

The formal presentation of John Searle's Chinese Room argument is borrowed from James Chase (Utas).