

The image shows the cover of a spiral-bound notebook. The cover is a light beige or tan color with a fine, woven fabric texture. On the left side, there is a silver metal spiral binding. The notebook is slightly open, revealing a white page on the left and a grey page on the right. The text is centered on the beige cover.

# **Filosofie III.**

**Turingův test a námitky**

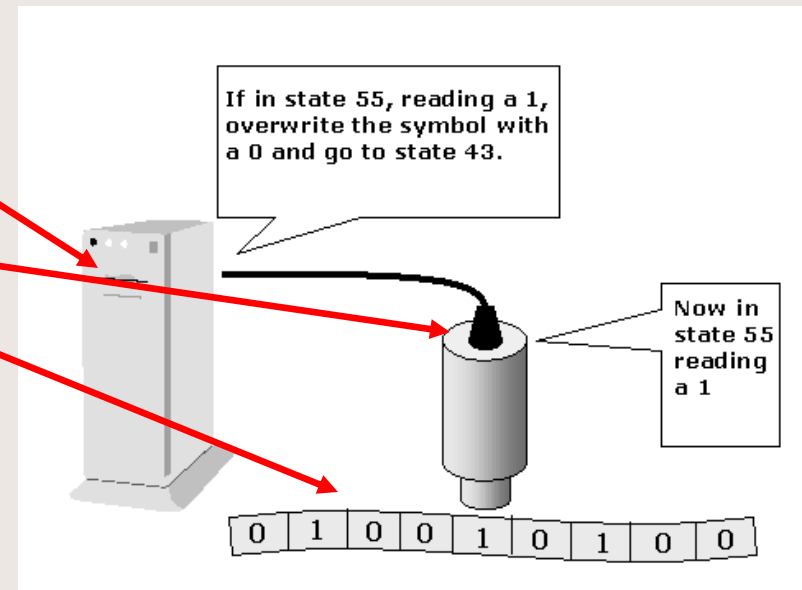
# Alan M. Turing

- 1912 – 1954
- britský matematik, logik, kryptoanalytik
- zachránil Evropu před nacismem (rozluštění Enigmy)
- texty:
  - „*On Computable Numbers with an Application to the Entscheidungsproblem*“
  - „*Computing Machinery and Intelligence*“



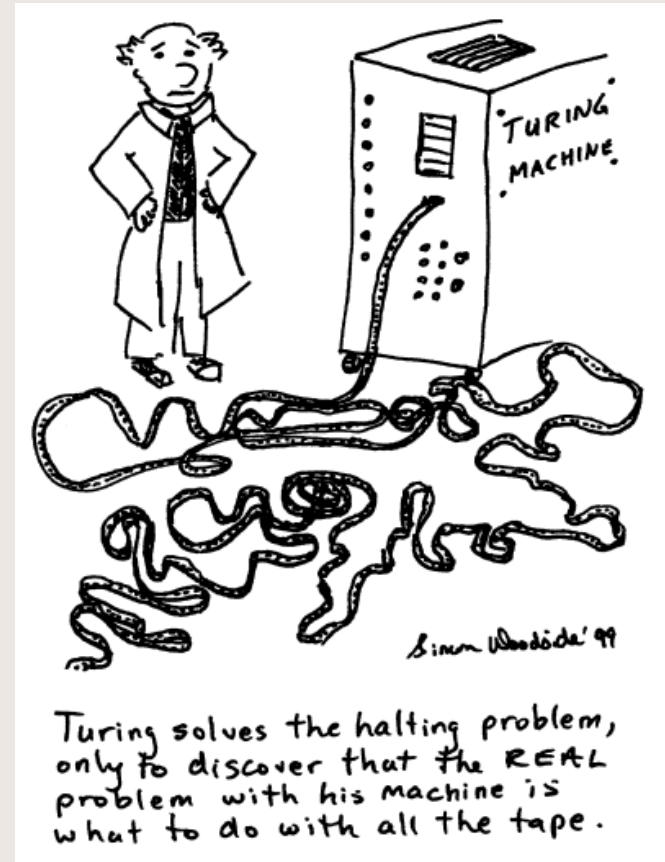
# Turingův stroj (klasický)

- **zařízení schopné implementovat určitou algoritmizovatelnou proceduru – výpočet**
- skládá se z **řídící jednotky** (*finite state machine controller*), **čtecí hlavy** a **neomezeně dlouhé pásky**
- v závislosti na současném stavu a symbolu přečteném na konkrétním místě pásky může stroj změnit svůj stav, pohybovat hlavou (doleva i doprava) a měnit znaky na pásce
- stroj je deterministický

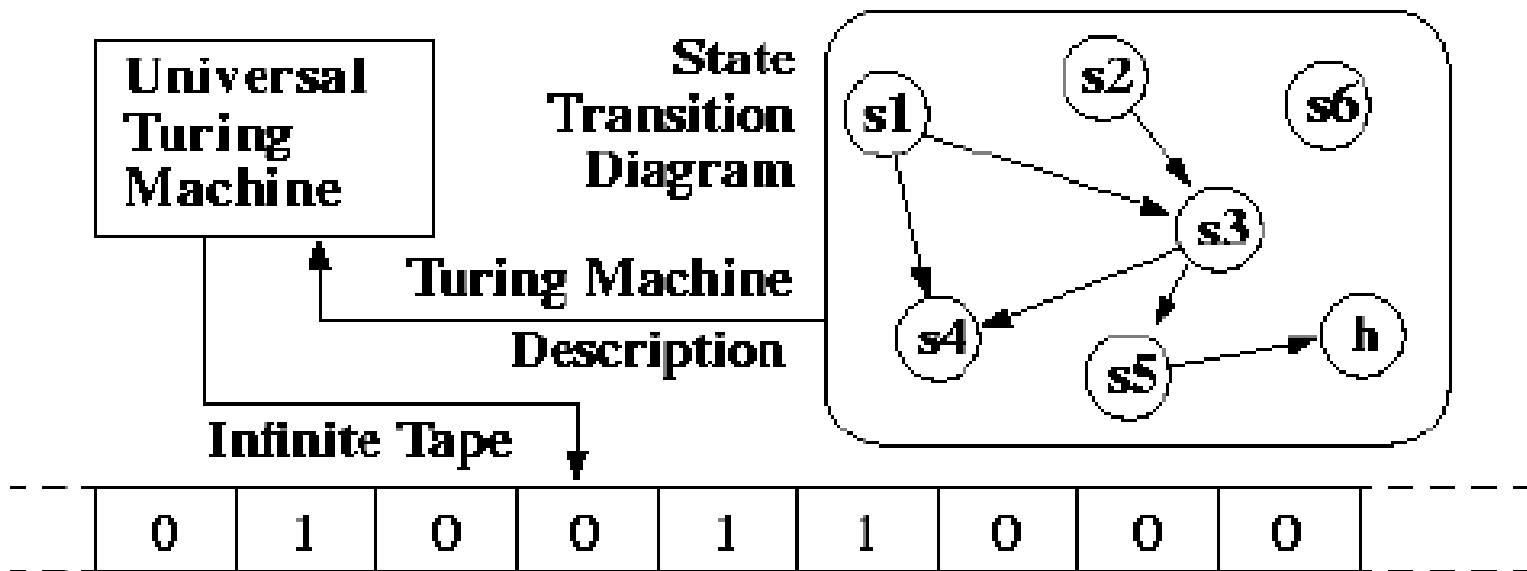


# Univerzální Turingův stroj

- zařízení, které je schopno implementovat jakoukoliv algoritmizovatelnou proceduru
- jako své stavy zahrnuje všechny myslitelné Turingovy stroje



**Halting problem**



„Turingův stroj si můžeme představit jako krabičku, do které vstupuje oboustranně nekonečně dlouhá páska. V krabičce je řídicí jednotka se čtecí hlavou, která čte obsah pásky. Podle obsahu políčka, které zrovna čtecí hlava čte a podle interního stavu řídicí jednotky se vybere pravidlo, které určuje, co bude Turingův stroj dělat. Ten může změnit obsah aktuálního políčka, může se posunout na pásce o jedno políčko vpravo nebo vlevo (nebo může zůstat stát) a nebo může změnit svůj stav.“ Martin Blažek

# Turingův test

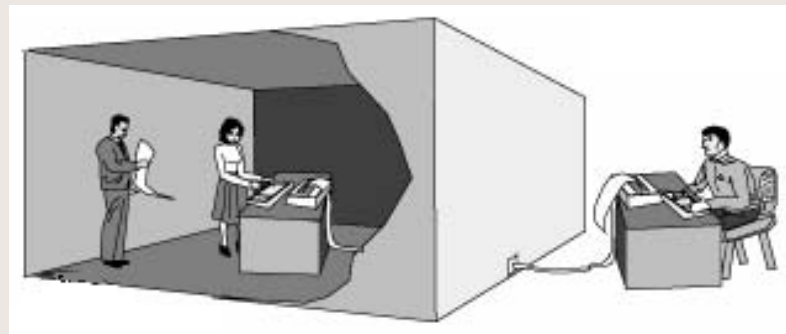
*„I believe that in about 50 years it will be possible to program computers...to make them play the imitation game so well that an average interrogator will not have more than 70 % chance of making the right identification after five minutes of questioning.“*

Alan Turing (1950)



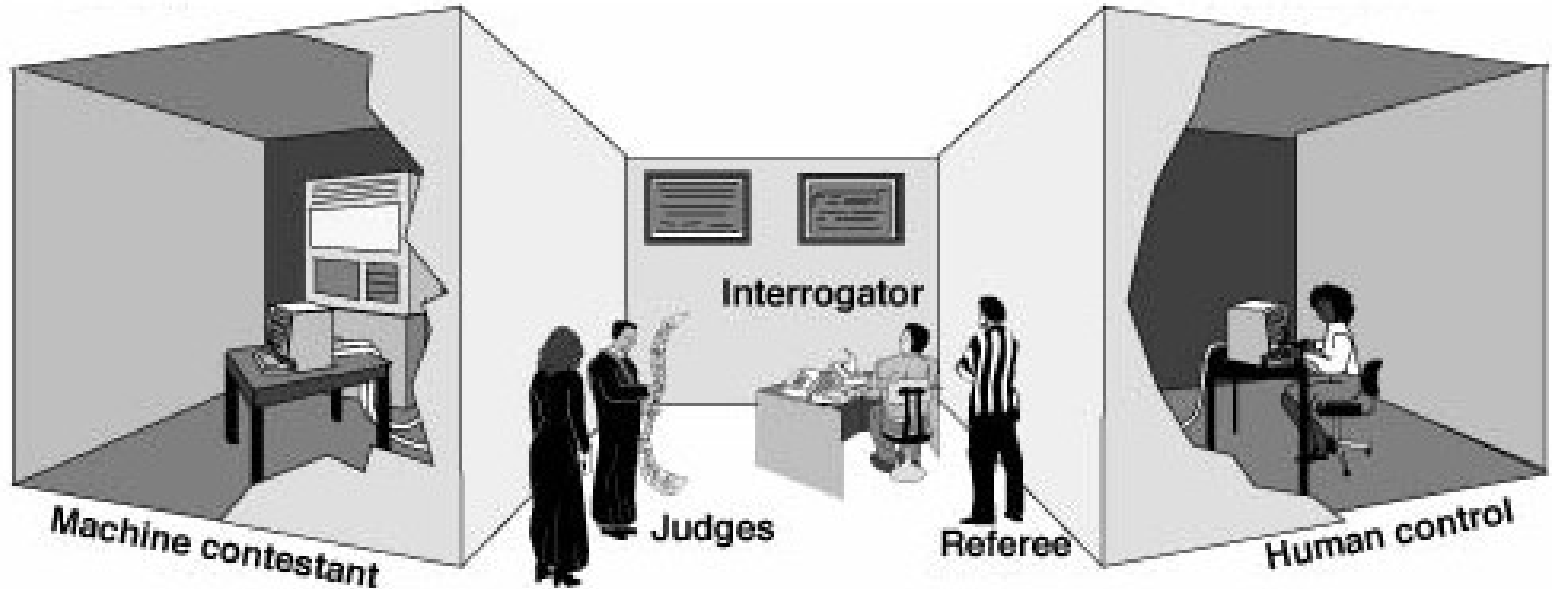
# Turingův test

- *Je možné pojímat myšlení jako jistý typ výpočtu?* – odtud převládající **paradigma počítačně-reprezentačního uchopení mysli**
- otázku: *Mohou stroje myslet?* se snaží **nahradit** jinou – **rozhodnutelnou otázkou**
- otázka inspirována tzv. **imitační hrou** – úkolem je pouze **na základě otázek a odpovědí rozhodnout, zda komunikují s mužem nebo se ženou**



**Imitační hra** hrají tři hráči: hráč X dostává od hráčů A a B informace v podobě psaného textu prostřednictvím terminálu (původně dálnopisu). Jeden z hráčů A a B je muž, druhý žena. Hráč X má odhalit, kdo je muž a kdo žena. Muž má překazit výhru hráče X, cílem ženy je, aby hráč X vyhrál.

# Turingův test



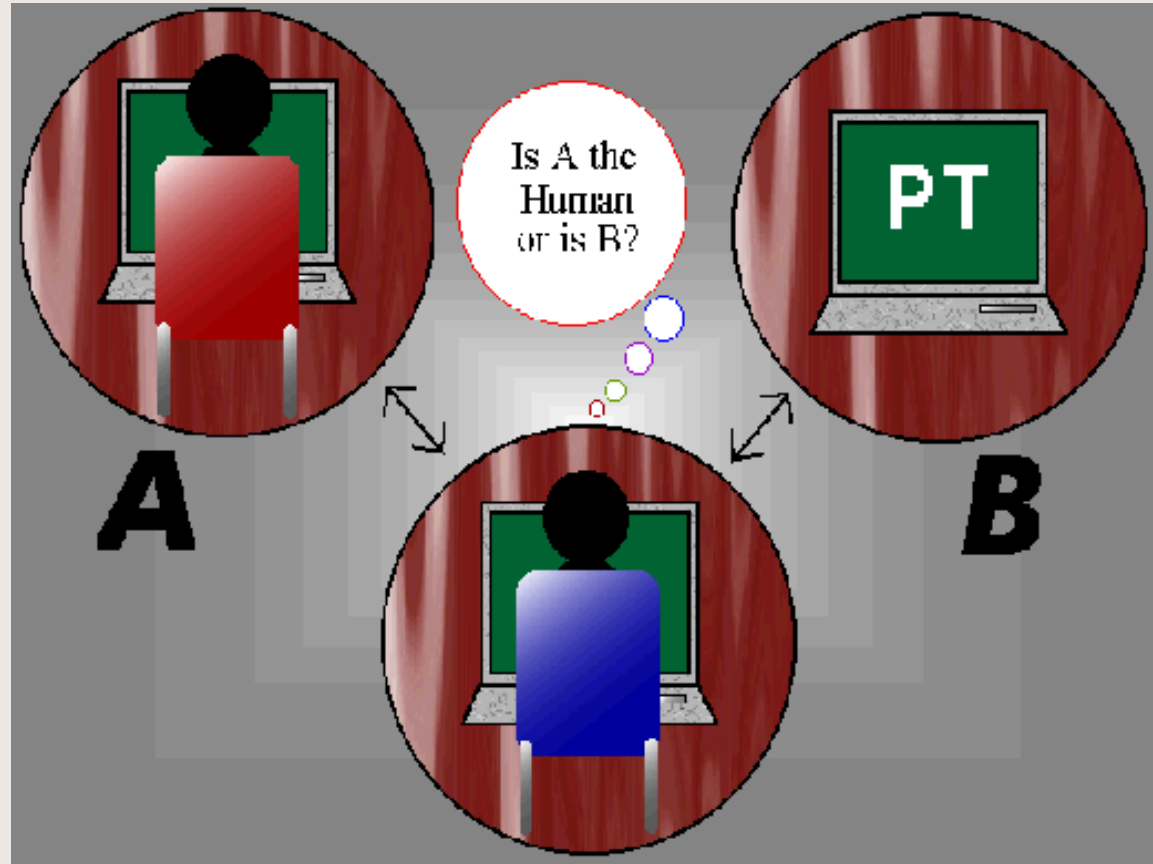
- není-li člověk schopen z jazykového chování rozpoznat, že nemluví s člověkem, ale se strojem, pak stroj obstál v Turingově testu
- otázka *Mohou stroje myslet?* nahrazena otázkou *Mohou stroje absolvovat tento test?*
- úspěšně složit Turingův test je **dostatečnou** (nikoliv nutnou) podmínkou mentálních stavů



# Námítky

- přímo v Turingově textu:
  1. teologická
  2. hlava v písku
  3. **matematická**
  4. **z vědomí**
  5. z různých neschopností
  6. Lady Lovelace
  7. **z kontinuity nervového systému**
  8. z neformálnosti jednání
  9. z mimosmyslového vnímání

# Turingův test



# Program ELIZA

- vytvořil Joseph Weizenbaum (1966)
- dokázal **úspěšně imitovat chování psychoterapeuta**
- část dialogu mezi programem ELIZA (E) a pacientem (P):

P : Všichni lidé jsou stejní.

E : V jakém smyslu?

P : Pořád nás do něčeho nutí.

E : Máte na mysli nějaký určitý příklad?

P : No, můj chlapec mě přinutil, abych sem přišla.

E : Váš přítel vás donutil, abyste sem přišla!

P : Říká, že jsem pořád tak skleslá.

E : Je mi líto, že jste skleslá.

P : Je to pravda. Jsem nešťastná.

E : Myslíte, že vám pomůže, že jste sem přišla, abyste nebyla nešťastná?

P : Potřebuji nějakou pomoc, to je jisté.

E : Co by to znamenalo, kdybyste dostala nějakou pomoc?

P : Možná, že bych se naučila nějak vyjít se svou matkou.

E : Řekněte mi něco víc o své rodině!

P : Moje matka se o mě stará.

E : Kdo z vaší rodiny se ještě o vás stará?

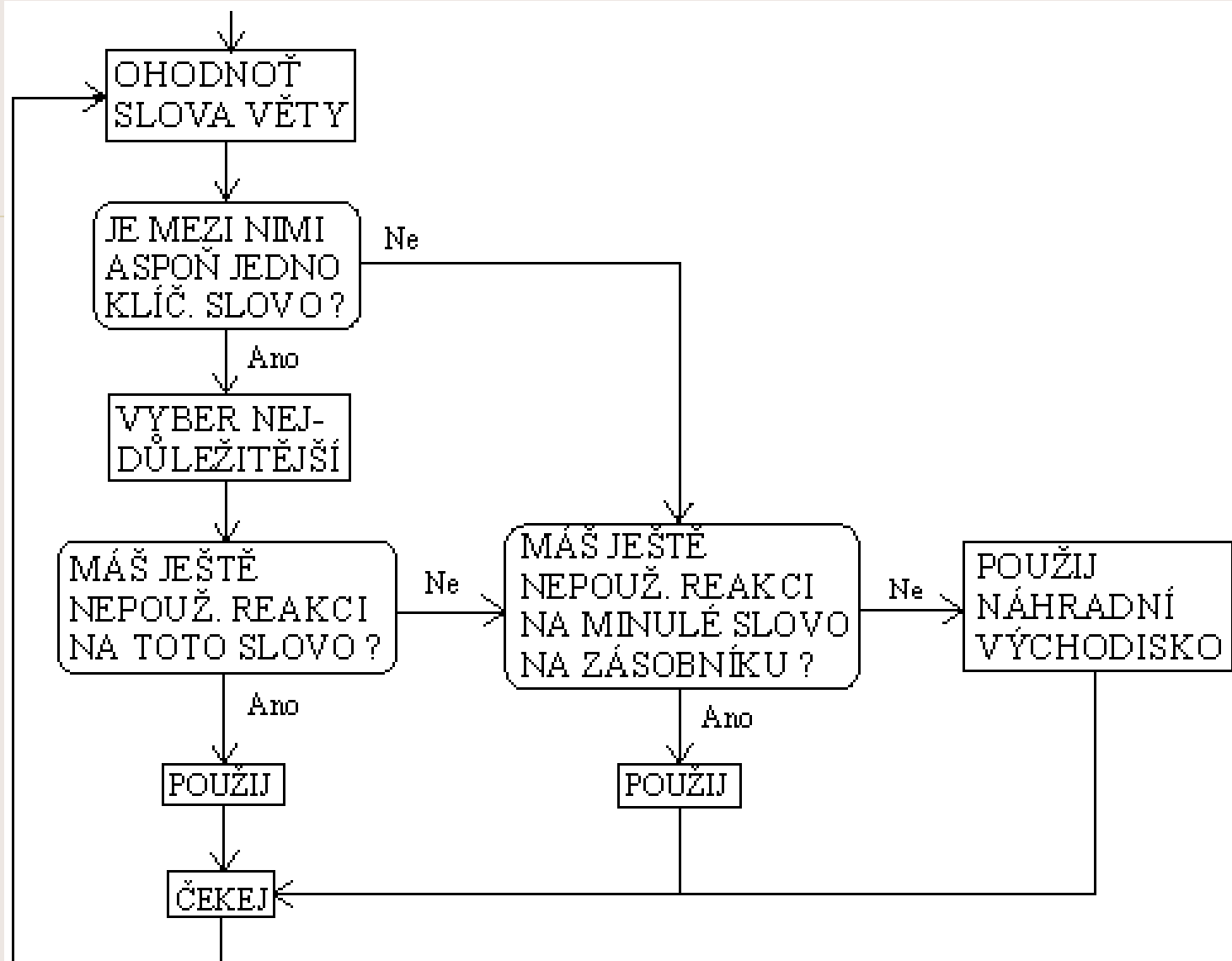
P : Můj otec.

E : Váš otec.

P : V něčem se podobáte mému otci.

E : V čem vidíte podobnost?

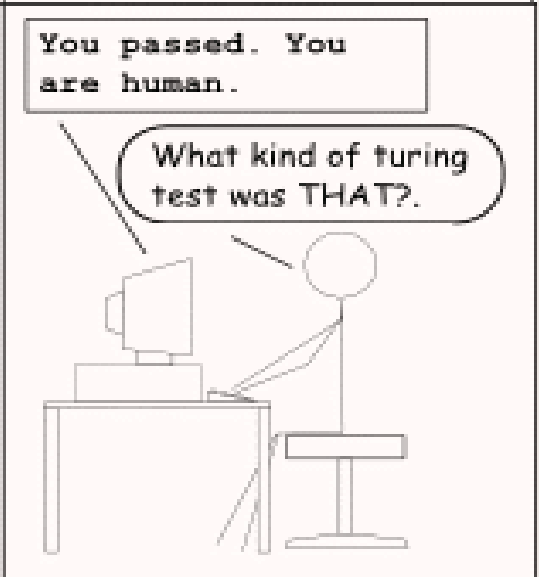
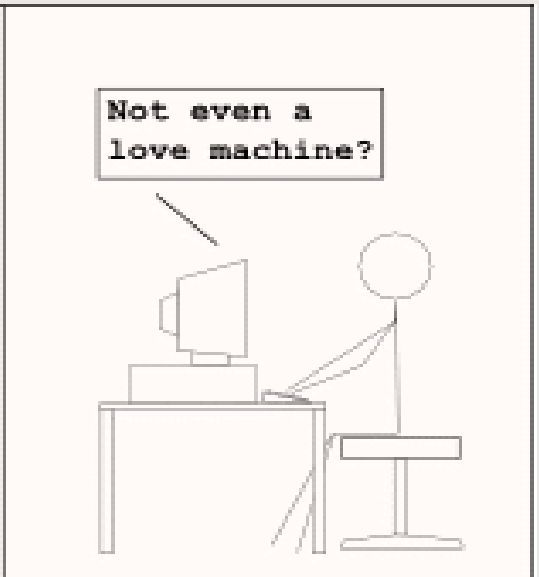
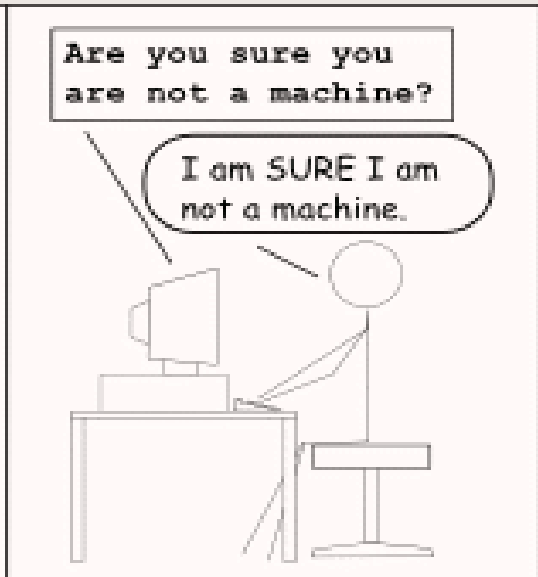
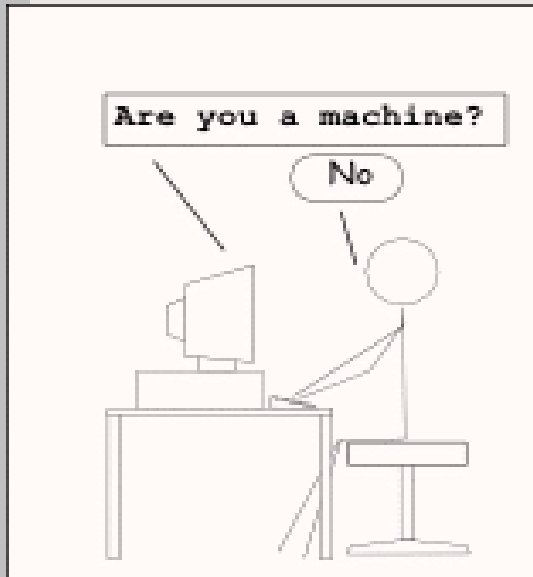
...



**Vývojový diagram Elizy**

# ... emancipace ...





# Závažné námitky

## 1. John R. Searle

- *Minds, Brains and Programs* (1980)

## 2. Ned Block

- *Psychologism and Behaviorism* (1981)

## 3. Roger Penrose

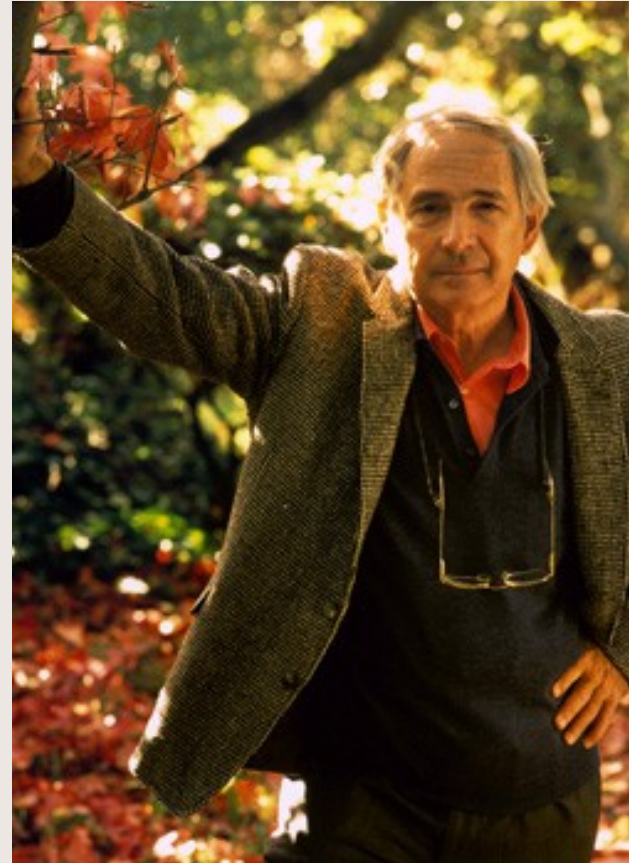
- *The Emperor's New Mind* (1989)

## 4. David Chalmers

- *The Conscious Mind* (1996)

# John R. Searle

- \*1932
- americký filosof (jazyka a mysli)
- profesor v Berkeley
- texty:
- *Speech Acts: An Essay in the Philosophy of Language* (1969)
- *Minds, Brains and Programs* (1980)
- *Mind: A Brief Introduction* (2004)





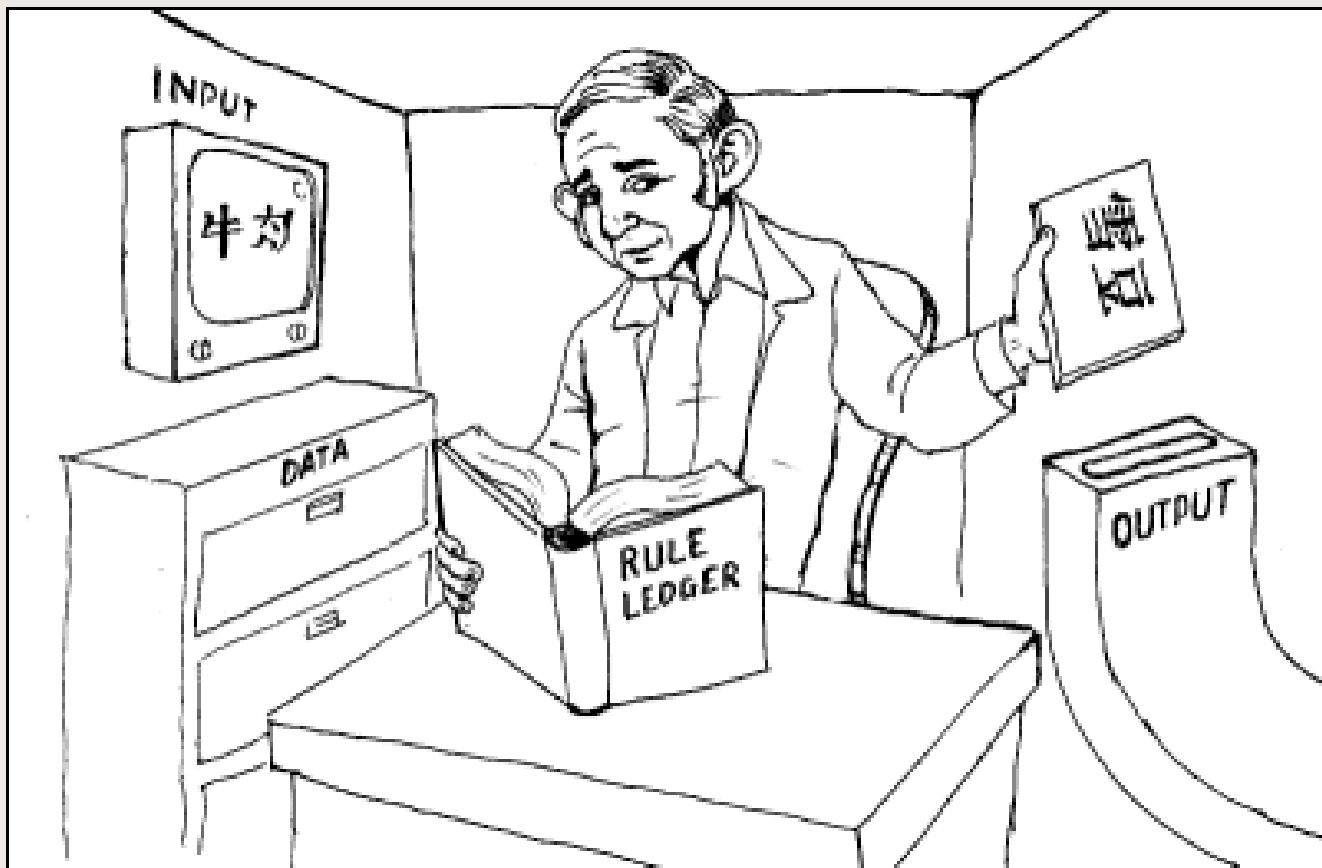
# Searlův čínský pokoj

- argument proti programu umělé inteligence (silná UI) opřené o kritérium Turingova testu
- Je vhodné rozlišovat trojí:
  1. **Myšlenkový experiment**
  2. **Argument ve vlastním smyslu**
  3. argument **Relativity výpočtu vzhledem k uživateli** (1990)

# Myšlenkový experiment

- **člověk, který neumí čínsky v místnosti zpracovává data podle jednoznačně předepsaných pravidel**
- **člověk neví, že vstupující data jsou čínské znaky, ani že klikyháky, které odevzdává ven jsou v čínštině **odpovědi na otázky****
- *Umí dotyčný čínsky?*

# Searlův čínský pokoj



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".



# *Umí dotyčný čínsky?*

## **1. podle Searla:**

- Neumí!
- vše, co umí, je podle anglicky napsaných pokynů přemísťovat desky s klikyháky na základě jejich formálních (syntaktických) vlastností

## **2. podle odpůrců:**

- Pokud neumí, odkud se berou správné odpovědi?!
- Searle chybně ztotožnil odpovídajícího s člověkem v místnosti, zatímco ve skutečnosti odpovídá celý pokoj (tzv. systémová odpověď)

# Argument ve vlastním smyslu

- proti **silné umělé inteligenci**, která tvrdí, že **k rozumění dochází výlučně díky programu**, který rozumějící implementuje
  1. Zacházení se **syntakticky definovanými jednotkami podle výlučně syntakticky definovaných pravidel nedává** samo žádný přístup k jejich případné sémantice
  2. Program je **formální** a jako takový **kauzálně inertní**; rozumění je naopak výsledek reálných dějů, které mají **kauzální povahu** – k rozumění nemůže docházet výlučně díky programu

# Relativita výpočtu vzhledem k uživateli

- **Všechny syntakticky definované vlastnosti jsou relativní vzhledem k uživateli – jsou něčím, co do reálného světa vnáší uživatel obdařený primární intencionalitou**
- **silnější než Argument ve vlastním smyslu, neboť tvrdí i to, že mozek není organizován algoritmicky**

# Zhodnocení

## 1. Myšlenkový experiment

- není konkluzivní (zřejmě i chybně zadaný)

## 2. Argument ve vlastním smyslu

- je konkluzivní, osudný pro silnou UI

## 3. Relativita výpočtu ...

- není konkluzivní pro ty, kdo odmítají primární intencionalitu



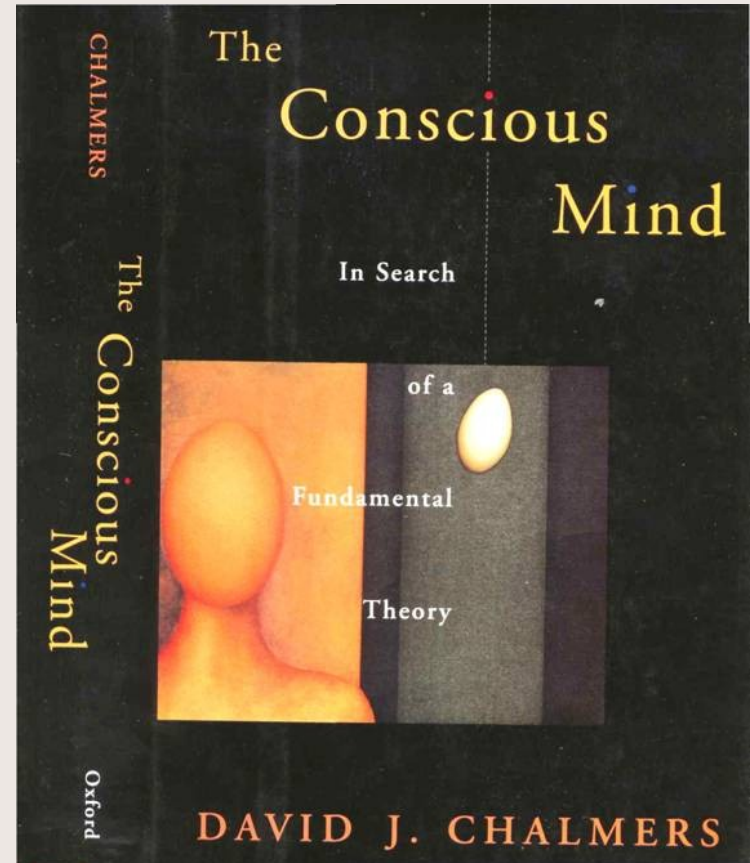
# David Chalmers

- \*1966
- australský filosof
- působí jako profesor filosofie na Australian National University
- texty:
- *The Conscious Mind*
- *Explaining Consciousness: The Hard Problem*



# Filosofická Zombie

- myšlenkový experiment
- hypotetický tvor, který **postrádá** **fenomenální** **mentální stavy**, v kterémkoli jiném ohledu je ale **nerozlišitelný od člověka**



# David Ch. a jeho Zombie



**Nerozlišitelní z perspektivy třetí osoby, ale z perspektivy první osoby fundamentálně odlišní – Zombie postrádá subjektivní fenomenání mentální stavy – kválie.**

# Kválníová mutace

- **proces**, při kterém dochází k neobvyklým změnám kválií
- **druhy**:
  1. **chybějící kvália** (sem patří Zombie)
  2. **obrácená kvália**
  3. **cizí kvália**

# Zacílení zombie-argumentů

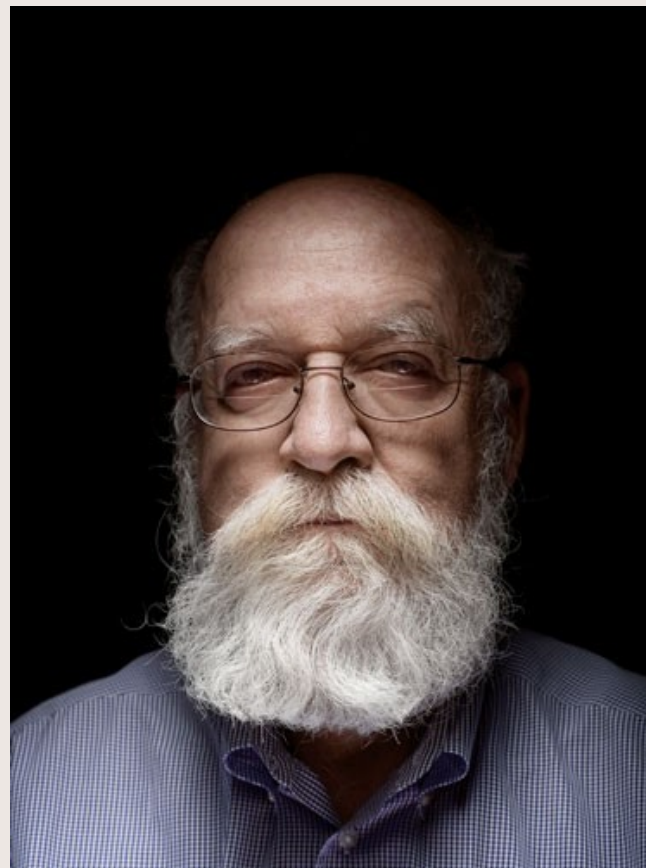
1. **proti behaviorismu** – dvě naprosto stejně se chovající stvoření, ale...
  2. **proti funkcionalismu** – dvě stvoření se stejnou funkcionální strukturou, ale...
  3. **proti teorii identity** – dvě stvoření s naprosto stejnými mozky, ale...
- +
4. **proti fyzikalismu** - zombifikovaný svět je bezrozporně myslitelný (logicky možný)

# Námitky

1. Opravdu absolvuje Zombie Turingův test?
2. Ach ty myšlenkové experimenty...
3. Zombie je možná, ale nic to nedokazuje – naše představa Zombie je příliš naivní.

# Zombie dilemma

- **bud' budeme chápat zombii úzce – bez reprezentací vyššího řádu – pak ale nemáme co do činění s behaviorálně identickou bytostí**
- **nebo budeme chápat zombii široce – s reprezentacemi vyššího řádu – pak je však zombií každý z nás**



**„Jsem Zombie.“**