E-LOGOS
ELECTRONIC JOURNAL FOR PHILOSOPHY
ISSN 1211-0442 4/2011

VŠE
University of Economics
Prague

Různé pohledy na otázku: Mohou stroje myslet?

Ondřej Vadinský

Abstract

This paper deals with the question, whether machines can think. Changes in thematisation of the question by several philosophers and scientists such as Cartesian dualism, Turing test, Searle's Chinese room argument, artificial neural networks of the Curchlands and Rapaport's abstraction and implementation and syntactic semantics are followed. The paper then shows what are the differences and similarities of such thematisations. Then it asks what is the nature of the relation between thought and its manifestation, which is the cornerstone of the Turing test. Furthermore the paper accents the need of combining suitable distinctive capacities with appropriate complexity to reach adequate analogy in the Chinese room argument. In the end the paper deals with relationship among duplication, abstraction and implementation of thought.

Abstrakt

Tato práce se zabývá otázkou, zda mohou stroje myslet. Sleduje tedy proměny její tematizace u vybraných filosofů a myslitelů, jimiž jsou: Descartův dualizmus, Turingův test, Searlův čínský pokoj, umělé neuronové sítě manželů Churchlandových a Rapaportova abstrakce a implementace a syntaktická sémantika. Poté, co ukáže, jak se tyto tematizace liší a v čem se naopak shodují, ptá se práce po povaze vztahu mezi myšlením a jeho projevy, který je ústřední pro oprávněnost Turingova testu. Dále práce akcentuje nutnost kombinace vhodné rozlišovací schopnosti s reprezentativní mírou komplexity pro adekvátnost myšlenkového experimentu s čínským pokojem. Konečně se práce také zabývá vztahy mezi duplikací, abstrakcí a implementací fenoménu myšlení.

Key words

Artificial intelligence, Turing test, Chinese room, artificial neural network, abstraction and implementation, thought.

Klíčová slova

Umělá inteligence, Turingův test, čínský pokoj, umělé neuronové sítě, abstrakce a implementace, myšlení.

Obsah

1 Úvod

Otázka, zda mohou stroje myslet, rozhodně nepatří k těm jednoduše zodpověditelným. Není to ani otázka nová, která by se vynořila jen v posledních několika dekádách provázených překotným rozvojem výpočetní techniky překonávající mnohá omezení, která se dříve zdála neotřesitelná. Přesto jde o otázku, která právě díky prudkému rozvoji výpočetní techniky získává v současnosti na aktuálnosti.

Esej se snaží zachytit postupné vynoření této otázky, které je patrné už v díle René Descarta. Nejde ale jen o samotnou otázku, zda mohou stroje myslet, ale také o způsob, jak něco takového ověřit. Za asi nejznámější metodu takového testu lze považovat Turingovu imitační hru, známou jako Turingův test, kterou tato esej také uvádí. Opozici přístupů klasické umělé inteligence ukáže esej na Searlově myšlenkovém experimentu s čínským pokojem. Další vývoj pak esej ukazuje především na poznatcích Williama Rapaporta v oblasti počítačového rozumění jazyku a přístupu manželů Churchlandových, který prosazuje umělé neuronové sítě.

V závěru práce autor shrnuje postoje jednotlivých myslitelů, aby ukázal, v čem se shodují a v čem naopak liší. Poté autor uvažuje nad správností zpětné implikace Turingova testu a ptá se po povaze vztahu mezi myšlením a jeho projevy. Dále autor rozvádí námitku manželů Churchlandových proti Searlovu čínskému pokoji na příkladě esenciální vlastnosti filmu. Nakonec se pak autor zabývá vztahy mezi duplikací, abstrakcí a implementací.

2 Descartes

2.1 Dualizmus tělo – mysl

Tématikou vztahu člověk – stroj, respektive člověk – zvíře se Descartes zabývá v části své Rozpravy o metodě věnované přírodním otázkám. Na úvod této části se Descartes věnuje přírodním zákonům a jejich zaručenosti „Bohem“.

Pak Descartes přechází k popisu živých tvorů a lidí. Descartes důsledně rozlišuje mezi funkcemi těla, kterými se lidé podobají zvířatům a funkcemi duše, čili mysli, kterými se od nich odlišují: „Neboť zkoumaje funkce, jež mohou tak být v těle, nalezl jsem tam přesně všechny ty, jež v nás mohou být, aniž na to myslíme a aniž k nim něčím přispívá naše duše, . . . , a jež se všechny rovnají těm, o nichž se dá říci, že se nám jimi nerozumná zvířata podobají: ale nemohl jsem takto nalézt mezi těmito funkcemi ani jedinou z těch, jež závisíce na myšlení, jsou jediné, jež nám náležejí jakožto lidem. Avšak nalezl jsem je tam potom všecky, když jsem předpokládal, že Bůh stvořil rozumnou duši a spojil ji s tělem jistým způsobem, jejž jsem popsal.“ [1, str. 35] Pak Descartes obšírně vysvětluje soudobé pojetí fungování lidského těla, především pak krevního oběhu. Dochází také k přirovnání lidského těla ke stroji vytvořeném rukama „Božíma“, viz [1, str. 32 – 40]. To ho přiměje zmínit se o tom, jak lze rozlišit člověka od stroje:

> *„A tu jsem se zvláště zastavil u důkazu, že kdyby existovaly takové stroje, jež by měly orgány a vnější vzhled opice nebo jiného nerozumného zvířete, měli bychom důvod se domnívat, že by byly ve všem stejné povahy jako tato zvířata; kdežto kdyby existovaly stroje, podobající se našim tělům a napodobující naše úkony potud, pokud by to mravně bylo možné, měli bychom vždy dva velice vážné důvody, abychom poznali, že proto ještě nejsou skutečnými lidmi. První důvod je, že by nikdy nemohly užívat slov ani jiných znaků, skládajíce je jako činíme my, abychom své myšlenky vyložili jiným. Neboť lze dobře chápat, že stroj může býti udělán tak, aby pronášel slova, ba dokonce aby pronášel některá ve spojení s tělesnými úkony, souvisejícími s nějakými změnami jeho orgánů: jako například když se ho dotkneme na určitém místě, aby se zeptal, co mu chceme říci, když na jiném místě, aby křičel, že ho to*

> *bolí, a podobně; nemůže však být udělán tak, aby slova různě sestavoval a takto odpovídal na vše, co se řekne v jeho přítomnosti, jak to i nejtupější lidé mohou činit. A druhý důvod je, že i kdyby vykonávaly určité věci stejně dobře nebo snad i lépe než kdokoli z nás, selhaly by nevyhnutelně v jiných, při nichž by vyšlo najevo, že nejednaly s vědomím, nýbrž toliko podle sestavení svých orgánů; neboť rozum je všestranný nástroj, kterého lze užívat ve všech možných případech, kdežto tyto orgány musí mít nějaké zvláštní uzpůsobení pro každý úkon jednotlivý, a proto je morálně nemožné, aby rozmanitost těchto orgánů v jednom stroji stačila přivést jej k tomu, aby jednal za všech okolností života stejně, jako jednáme my vlivem svého rozumu.“* [1, str. 41]

Stejná metoda umožňuje podle Descarta rozlišit člověka od zvířete. Přičemž řečí nemyslí Descartes jen řeč mluvenou, ale například i znakovou. Ovšem je třeba nepovažovat mylně za slova přirozené pohyby vyjadřující toliko vášně. Zvířata pak podle Descarta nemají méně rozumu než lidé, ale nemají ho vůbec. A znovu Descartes akcentuje, že sám fakt, že je zvíře v nějaké činnosti schopnější než člověk, nevypovídá o jeho rozumnosti či duchu, ale pouze o tom, že jeho tělo je tomuto úkonu vhodně přizpůsobeno. Závěrem této části hovoří Descartes o duši. Zdůrazňuje, že ta je nehmotná a zcela jiné povahy než duše zvířat. Vztah duše k tělu je užší, avšak podle Descarta duše nepodléhá destrukci jako tělo, viz [1, str. 41 – 43].

3 Turing

3.1 Imitační hra

Úvodem svého článku si Turing klade otázku, zda mohou stroje myslet. Přirozeně se pak ptá po tom, co je vlastně stroj a co to znamená myslet. Následně se brání snaze hledat definice těchto slov založené na tom, jak se tato slova běžně používají. Navrhuje tedy výše uvedenou otázku převést na jednoznačnější problém. Tím má být podle Turinga tzv. imitační hra. Tu Turing přibližuje následovně: účastní se jí muž, žena a „vyšetřovatel“ [1]. „Vyšetřovatel“ – v oddělené místnosti od ostatních – má za úkol zjistit, kdo z dvojice je muž a kdo žena, přičemž muž se snaží „vyšetřovatele“ zmást a žena mu naopak pomoci [2]. „Vyšetřovatel“ smí pokládat otázky. Komunikace by měla probíhat nějakým způsobem, který zabrání identifikaci, tj. např. strojopisně, telegrafem, přes prostředníka … Tuto úlohu lze pak snadno přeměnit tak, že místo muže dosadíme stroj. Pak se tedy Turing ptá, zda se „vyšetřovatel“ bude mýlit tak často jako v původní verzi hry – tedy zda dokáže stroj člověka věrně imitovat. Viz [2].

V další části článku Turing zvažuje výhody a nevýhody imitační hry. Jako hlavní výhodu vidí možnost oddělit od sebe fyzické a intelektuální charakteristiky člověka a stroje. Výhodou dotazovací metody podle Turinga je i to, že dokáže adresovat téměř libovolnou oblast lidské intelektuální tvorby, sám Turing uvádí tvorbu sonetu, hru šachů nebo matematické výpočty. Poměrně silnou námitku vidí naopak Turing ve značné obtížnosti vzájemné imitace člověka a stroje v některých oblastech – např. přesnost a rychlost matematických výpočtů. Turing se také nezabývá aplikací teorie her na tento případ, pouze předpokládá, že nejlepší strategií pro stroj je imitace člověka. Viz [2].

Následně Turing rozebírá stroj účastnící se imitační hry. Především zde vytyčuje podmínku, že by mělo jít o tzv. elektronický, či digitální počítač a ne například o uměle vypěstovaného člověka. Viz [2].

[1] V originále „interogator“.

[2] Zde Turing poznamenává, že prostá odpověď domnělé ženy: „Já jsem žena, neposlouchej ho,“ nic neodhalí, protože muž může odpovídat stejně tak, aby „vyšetřovatele“ zmátl.

3.2 Digitální počítače

Turing pak podává definici digitálního počítače, která vychází z představy člověka provádějícího zadané početní operace ručně, podle pevných pravidel. Dále popisuje hlavní části takového digitálního počítače – paměť, výpočetní jednotku a ovládací mechanizmy zaručující korektní vykonávání instrukcí a v celku detailně popisuje vnitřní fungování počítače při vykonávání jednotlivých instrukcí a složitějších programů. Zdůrazňuje také roli generátoru náhodných čísel a zabývá se teorií počítače s nekonečnou kapacitou paměti. Na příkladu návrhu Babbageova analytického stroje [3], fungujícího čistě na mechanickém principu, ukazuje, že spojitost mezi člověkem a digitálním počítačem neleží v rovině elektrické, která se v obou případech stará o přenos signálů. Podobnost digitálního počítače a člověka vidí Turing spíše v jejich matematické analogii funkce. Viz [2].

Nyní Turing popisuje digitální počítač jako diskrétní stavový automat [4], z čehož pak vyvozuje možnost rozumně přesné predikce budoucích stavů stroje na základě znalosti vstupních signálů a všech vnitřních stavů. Tento fakt pak klade do kontrastu s Lapaceovou teorií chaotičnosti vesmíru. [5] K pojetí digitálního počítače jako stavového automatu dodává Turing, že počet vnitřních stavů takovýchto počítačů je obrovský, predikci budoucích stavů pak tedy Turing také přenechává digitálnímu počítači. Digitální počítač pak může napodobovat libovolný jiný stavový automat tak, že by ho „vyšetřovatel“ nedokázal od původního automatu odlišit. Turing pak označuje digitální počítač jako univerzální stroj a říká, že není potřeba vytvářet různé specializované stroje pro jednotlivé úlohy, ale stačí použít vhodně naprogramovaný digitální počítač. Zmiňuje také, že takto uvažovány jsou digitální počítače ekvivalentní. Vlastní úlohu pak Turing posouvá do nové polohy: jsou představitelné digitální počítače, které by obstály v imitační hře? Viz [2].

Turing předestírá svou vlastní představu o pokroku v oblasti digitálních počítačů tak, že budou za určitou dobu schopny do určité míry obstát v imitační hře. Viz [2].

3.3 Protiargumenty k imitační hře

Pak Turing prezentuje několik protiargumentů ke své teorii. Začíná teologickou námitkou, podle níž je schopnost myslet spjata s duší člověka, kterou mu dal „Bůh“ a kterou stroj nemá a tudíž nemůže myslet. Podle Turinga je toto samotné tvrzení problematické, protože vlastně popírá všemohoucnost takového „Boha“. Turing také poukazuje na nespolehlivost podobných teologických argumentů. Viz [2].

Další uváděnou námitkou je „strkání hlavy do písku“ před důsledky myslících strojů. Turing uvádí, že lidé rádi shledávají sebe jako výjimečné, všemu nadřazené a mající nutnost být všemu nadřazení. I tato námitka přijde Turingovi natolik nepodložená, že ji nepokládá za hodnu nějakého většího vyvracení. Viz [2].

Matematická námitka proti Turingově teorii spočívá v matematickologické limitaci diskrétních stavových automatů. Jako příklad uvádí Turing Gödelův teorém o existenci takových logických formulí, které jsou nerozhodnutelné v rámci logického systému. Zde Turing připouští, že jsou úlohy, které stroj nemůže vyřešit. Turing se zamýšlí, co vlastně z této námitky vyplývá pro jeho základní otázku. Dochází k závěru, že z ní plyne pouze to, že se stroje mohou v některých otázkách mýlit stejně jako se v jiných mýlí lidé. Viz [2].

[3] V originále „Analytical engine“.

[4] V originále „discrete-state machine“.

[5] V Laplaceově případě záleží při předvídání stavů vesmíru na drobných odchylkách v poloze jednotlivých elektronů, které mohou mít nedozírně odlišné dopady. Podobným jevem se zabývá teorie chaosu a takzvaný efekt motýlích křídel.

Protiargument vědomí spojuje lidské myšlení se schopností vytvořit umělecké dílo, prožívat emoce a se sebeuvědoměním. Protiargument vědomí lze chápat značně solipsisticky. Takovéto chápání však ve svém nejextrémnějším pojetí znemožňuje prokázat, zda si i jiný člověk než solipsista sám sebe uvědomuje. Turing se tedy zaměří na méně radikální pojetí protiargumentu vědomí. Ukazuje pak, že obdoba imitační hry se používá jako tzv. *viva voce* ke zjištění, zda student tématu skutečně porozuměl, nebo zda se pouze něco naučil nazpaměť. Zakomponovat tuto metodu do Turingovy imitační hry ale není problém. Protiargument vědomí je tedy v podstatě rozšířením původního Turingova testu, než jeho zamítnutím. Turing zde zároveň uvádí, že na zodpovězení otázky, zda stroje myslí, není potřeba rozřešit všechny záhady mysli. Viz [2].

Protiargument „neschopnosti“ [6] spočívá ve vyjmenování různých schopností, kterých stroje podle zastánce takového argumentu nejsou schopni. Turing vidí jejich základ ve špatně použité indukci. Nad některými se pozastavuje. Například tvrzení, že stroje nemohou dělat chyby [7], Turing zcela vyvrací: Stroj lze samozřejmě naprogramovat tak, aby vrátil různě náhodný, chybný výsledek. Na tvrzení, že stroj nemůže být subjektem svého myšlení, Turing nahlíží spíše jako, zda může být stroj předmětem svých operací – tedy zda například může stroj upravit svůj vlastní program na základě zpětné vazby tak, aby byl v dané operaci efektivnější. [8] Připomínka k diverzitě v chování stroje souvisí podle Turinga s kapacitou paměti stroje. Viz [2].

Námitku lady Lovelaceové k Babbageovu analytickému stroji lze shrnout tak, že tento si nenárokuje nic sám o sobě vymyslet, pouze dělá to, k čemu ho naprogramujeme, tedy o čem i lidé sami vědí, jak to udělat. Zkrátka lze říci, že stroj nevytvoří nic nového. Novější varianta této námitky přisuzuje strojům, že nemohou člověka nikdy překvapit. Zde Turing argumentuje, že nás stroje často překvapují a ač to lze do značné míry přisuzovat našemu chybnému odhadu jejich reakce, jde pořád o překvapení. Viz [2].

Protiargument spojitosti nervového systému člověka, který by pak byl obtížně modelovatelný diskrétním digitálním počítačem odráží Turing následovně: diskrétní digitální počítač je schopen s určitou přesností tak dobře imitovat spojitý počítač, že lidský „vyšetřovatel“ jen těžko pozná rozdíl. Viz [2].

Protiargument neformality lidského chování říká, že člověk neřídí své chování konečnou množinou pravidel pro jednání v určitých situacích, z čehož plyne, že člověk není stroj. Turing rozebírá vědomost a nevědomost takových pravidel. Dochází k tomu, že existují nevědomé zákonitosti lidského reagování a v jejich smyslu pak člověk jedná strojově. Dále Turing rozporuje možnost, že jsme schopni poznat, zda nějaká množina je či není úplnou množinou takových pravidel či zákonitostí. Viz [2].

Protiargument extrasenzorálního vnímání – například telepatie – považuje Turing za značně silný. Fenomén sám o sobě je statisticky významný, ale vědecky nepochopený. Lze tedy spekulovat, zda by extrasenzorální vnímání nedokázalo Turingův test zneplatnit ve smyslu, že by odstranilo bariéru vnímání, na které je založen. Přirozenou reakcí by pak bylo zpřísnění podmínek testu. Viz [2].

3.4 Učící se stroje

V závěrečné části svého článku prezentuje Turing myšlenku učících se strojů. Nejprve se však vrací k námitce lady Lovelaceové. Lidskou schopnost originální tvorby přirovnává k řetězové reakci vyvolané neutronem, který se srazil s nadkritickým množstvím látky. Většina lidských

[6] V originále Arguments from Various Disabilities.
[7] Jde o chyby v úsudku.
[8] To považuje Turing za možné, a dnešní stav oblasti genetických algoritmů mu v tomto dává zapravdu.

myslí podle Turinga reaguje na příchozí nápad jakoby útlumem. Malá část ale naopak rozběhne „řetězovou reakci“ a vyprodukuje mnoho dalších souvisejících nápadů a teorií. Lze-li takový fenomén pozorovat u lidské mysli, ptá se Turing, je pak možné vytvořit stroj tak, aby se i u něj rozběhla taková „řetězová reakce“? Turing také zmiňuje analogii cibulových slupek jako možné vysvětlení funkce mysli. Skupinu funkcí vysvětluje tato analogie jako mechanickou a tedy nenáležící skutečné mysli, tedy jen jakoby slupku cibule. Analogie se pak táže, kam toto dojde, zda k nějakému skutečnému jádru cibule, nebo k poslední prázdné slupce. Viz [2].

Turing vidí největší nedostatky strojů své doby v oblasti programování. Vložení veškerého kódu považuje Turing za velice náročné, zamýšlí se tedy nad tím, jak vlastně vznikne dospělá lidská mysl. Pode prosté úvahy jde o dětskou mysl, zdokonalenou učením a zkušenostmi. Naprogramovat strojový mechanizmus odpovídající dětské mysli, by mělo být podle Turinga značně snazší. Úloha se tak rozpadá na propojené procesy naprogramovaní umělé dětské mysli a procesu učení. Turing zde navrhuje proces řízené evoluce stroje. Metodu učení pak zakládá na komunikaci a kombinaci odměn a trestů. Rozebírá také různé přístupy ke složitosti vytvářené umělé dětské mysli, především o přínosu indukce a inference. Stav stroje během procesu učení pak bude takový, že ani učitel nebude schopen říct, co se děje uvnitř. Zde Turing zmiňuje, že inteligentní chování by se mělo odchylovat od zcela disciplinovaného chování při výpočtech, ale ne tolik, aby budilo zdání nahodilosti. S čímž souvisí i další Turingova připomínka, zmíněná už výše, že naučené procesy nedávají výsledek s absolutní jistotou, respektive, že mohou být odnaučeny. Ve svých úvahách však Turing zdůrazňuje roli náhodnosti při prohledávání prostoru řešení, považuje ji za lepší než systematické prohledávání. Turing uzavírá svůj článek úvahou, v jaké oblasti s učením strojů začít. Doporučuje vyzkoušet, jak čistě abstraktní oblast jako hraní šachů, tak i možnost naučit stroj komunikovat v přirozeném jazyce a vybavit ho obdobou lidských smyslových orgánů, aby pak mohl být zařazen do běžného učebního procesu. Viz [2].

4 Searle

4.1 Silná a slabá umělá inteligence

Searle ve svém článku nejprve rozlišuje umělou inteligenci na slabou a silnou. V oboru slabé umělé inteligence je podle Searla hlavní rolí počítače být mocným nástrojem při zkoumání lidské mysli. Oproti tomu v silné umělé inteligenci není počítač jen nástroj pro řešení problému, ale jeho program je oním hledaným řešením – naprogramovaný počítač má totiž kognitivní stavy [9]. Searle pak nevidí problém v tvrzeních slabé umělé inteligence, ale soustředí se na zmíněné tvrzení silné umělé inteligence – a sice že naprogramovaný počítač má v podstatě kognitivní stavy a je tedy vysvětlením lidské kognice – reprezentované mimo jiné Schankem, Winogradem, Weizenbaumem a také Turingem. Viz [3].

Nyní Searle stručně shrne Schankův program: Program má za cíl simulovat lidskou schopnost porozumění příběhům. Tato schopnost se u člověka demonstruje tím, že je schopen odpovědět na otázky týkající se onoho příběhu, ač v něm nejsou poptávané informace explicitně řečeny. Schankovy stroje používají reprezentaci implicitních znalostí o nějakém tématu k zodpovězení zmíněného typu otázek tak, jak se čeká od člověka. Příznivci silné umělé inteligence pak říkají, že stroj vlastně rozumí danému příběhu a jeho program vysvětluje lidskou schopnost porozumění příběhu. Tato tvrzení přijdou Searlovi jako nepodložená. Viz [3].

[9] V originále „cognitive states“.

4.2 Experiment s čínským pokojem

Searle navrhuje test takovéto teorie: Pokládá si otázku, jaké by to bylo, kdyby jeho mysl skutečně fungovala tak, jak tvrdí teorie silné umělé inteligence. Searle se tedy uzavírá do čínského pokoje, dostává nějaký čínský text – skript, kterému vůbec nerozumí. Pak dostává další čínský text spolu s anglickými instrukcemi – příběh, což mu umožní přiřazovat k sobě formální symboly prvních dvou textů na základě jejich tvaru. Nyní dostává třetí čínský text s anglickými instrukcemi – otázky, což mu umožní propojit dosud obdržené čínské texty a reagovat na nějakou skupinu znaků ze třetího textu jinou skupinou znaků, opět na základě tvaru – odpovědi. Anglická pravidla pak jsou programem. Kromě toho dostává Searle ještě příběhy v angličtině, kterým rozumí, pak také anglické otázky k těmto příběhům, na které pak odpovídá v anglickém jazyce s porozuměním jako každý rodilý mluvčí. Po nějaké době je pak Searle natolik dobrý v manipulaci s čínskými znaky a programátoři natolik dobří v psaní instrukcí, že jeho odpovědi z vnějšího pohledu jsou naprosto k nerozlišení od odpovědí Číňanů. Tedy nikdo z jeho odpovědí nepozná, že neumí ani slovo čínsky. Jeho odpovědi v čínštině i angličtině jsou stejně dobré. V případě čínštiny však jen manipuluje s formálně specifikovanými neinterpretovanými znaky, tedy Searle v případě čínštiny v podstatě je počítačovým programem. Viz [3].

Ze svého myšlenkového experimentu Searle vyvozuje následující: Ač má vstup i výstup neodlišitelný od Číňana, nerozumí v situaci, kdy je jakoby počítačem, ani slovo čínsky. Tedy ani počítač nerozumí tomu, co zpracovává, protože počítač nemá k dispozici nic víc než Searle v případě čínského pokoje, ve kterém Searle nerozumí tomu, co zpracovává. O druhém tvrzení silné umělé inteligence pak Searle vyvozuje: Počítač a program [10] nezakládají postačující podmínky rozumění, protože fungují bez porozumění. Podle Searla není program ani nutnou podmínkou rozumění. Rozumění a výpočetní operace nad čistě formálně definovanými elementy podle Searla zcela nesouvisí. Viz [3].

Nyní se Searle věnuje rozumění. Zabývá se hlavně případy, kdy se rozumění uskutečňuje a kdy nikoliv. [11] Podle Searla je rozumění vlastní lidem, nikoliv strojům, ač jim ho lidé často přisuzují, protože jimi rozšiřují svou intencionalitu. Toto přisouzení je podle Searla zcela metaforické, tedy strojové „rozumění“ instrukcím je svou podstatou něco zcela jiného než lidské rozumění jazyku. Searle se zde ohrazuje především proti srovnání lidského a strojového rozumění jako stejných principů, a říká, že strojové rozumění – ve smyslu srovnatelném s lidským – není. Viz [3].

4.3 Protiargumenty k čínskému pokoji

Searle nyní uvádí námitky proti svému čínskému pokoji a vyvrací je. Námitka systému: Rozumění neleží v jednotlivci ale v systému, jehož je částí. Searlovo řešení této námitky spočívá ve zvnitřnění systému do individua: Jednotlivec se naučí pravidla z knihy, čínské znaky z tabulek a bude provádět všechny výpočty zpaměti. Podle Searla pak jednotlivec stále nerozumí čínštině a ani systém, protože v systému není nic, co by nebylo v jednotlivci. Když tuto námitku rozebírá dále, zdůrazňuje Searle, že v angličtině rozumí obsahu, na který je tázán, ale v čínštině ví jen, jaké formální manipulace má provést s nějakými znaky. Tedy, že celý jeho čínský pokoj má ukázat, že pouhé takové manipulace s formálními symboly nezakládají rozumění. Searle zde říká, že správný vstup, výstup a program neznamenají vždy rozumění. Tedy, pokud systém projde Turingovým testem, neimplikuje to, že systém rozumí tomu, co zpracovává. Turingovým testem totiž projde jak systém rozumějící – Searle v případě anglických příběhů – tak i systém nerozumějící – Searle v případě čínských příběhů. Viz [3].

[10] Program spočívající jen ve formálních manipulacích s formálně danými znaky, či symboly.

[11] Námitky o různých stupních a úrovních rozumění, či nefaktičnosti rozumění a jeho nutnosti podrobit se úsudku, Searle připouští, ale nepovažuje tyto projevy za relevantní pro řešenou otázku.

Robotická námitka spočívá ve vytvoření robotického těla řízeného počítačem, které by pak mohlo vnímat a pohybovat se, a tak mít i rozumění. Tato námitka tedy tacitně přiznává, že rozumění je o něčem víc, než jen o manipulaci s formálními symboly. Searle však říká, že vnímání a pohybování se nezakládá rozumění. Rozšiřuje svůj příklad s čínským pokojem o více čínských znaků na vstupu (vnímání) a více čínských znaků na výstupu (pohybování) a tvrdí, že člověk v roli počítače stále nerozumí ničemu z toho, jen manipuluje se symboly. Viz [3].

Námitka simulátoru mozku předpokládá vytvoření programu simulujícího jednotlivé neurony aktivní v mozku Číňana. Pak tedy by měl stroj rozumět čínštině, nebo jí nerozumí ani Číňan. Na úrovni synapsí by mělo být vše stejné. Searle na tuto námitku reaguje následující úpravou svého čínského pokoje: Muž v něm na základě anglických instrukcí otevírá a zavírá kohoutky složitého systému potrubí. Muž ani potrubí čínštině podle Searla stále nerozumí. Podle Searla totiž simulátor mozku simuluje špatné aspekty mozku, tedy jen formální strukturu sekvence aktivování neuronů na synapsích, namísto jeho schopnost vytvářet intencionální stavy. Viz [3].

Kombinační námitka spojuje tři výše uvedené námitky: Jde o robota řízeného počítačem, který je naprogramován jako simulátor lidského mozku a při jeho vnímání jako uceleného systému, bychom mu měli připsat už snad konečně intencionalitu. Searle takovému systému – stejně jako čínskému pokoji – dovoluje přisoudit intencionalitu jen zvnějšku. Podle Searla takové přisouzení intencionality nesouvisí s formálním programem – jak tvrdí silná umělá inteligence – ale s podobností s naším chováním, ze které mu intencionalitu mylně přisuzujeme. Pokud se ale opět uzavřeme do čínského pokoje uvnitř takového robota, nahlédneme, že stále jde jen o pouhou formální manipulaci s neinterpretovanými symboly. Stejného omylu přisouzení intencionality se člověk podle Searla dopouští, když ji vkládá do chování zvířat. Viz [3].

Námitka jiných myslí říká: To, že lidé rozumí čínštině (nebo něčemu jinému), lze usoudit jen z jejich chování. Projde-li stroj testem chování, pak i jemu je třeba přisoudit rozumění. Searle říká, že jde o to, co vlastně lidem přisuzuje, a že to to nemůže být pouhý výpočetní proces a jeho výstup, protože ten může existovat bez kognitivních stavů. Viz [3].

Námitka mnoha pokojů říká, že se Searle ve svém experimentu zaměřil na současný stav technologie, a že v budoucnu bude možné nasimulovat i jím zmiňované kauzální procesy potřebné pro intencionalitu, ať už jde o cokoliv. Searlova připomínka pak spočívá v tom, že takováto námitka příliš odchází od původního pojetí silné umělé inteligence, a jelikož neobsahuje ověřitelnou hypotézu, nemůže na ni Searle reagovat. Viz [3].

4.4 Rozumění podmíněné intencionalitou

Nyní se Searle vrací k otázce, která vysvítá z jeho myšlenkového experimentu: Co je to, co rozlišuje případ s anglickými příběhy od příběhů čínských? Tedy z čeho se to rozumění skládá a proč ho nelze předat stroji? Searle nejprve nevylučuje možnost předat stroji lidské rozumění jako takové, protože jak uvádí, lidské tělo a mozek je v zásadě také strojem. Uvádí ale silné námitky proti tomu, pokud by strojové operace měly být pouze výpočetním procesem nad formálně definovanými prvky, tedy formálním počítačovým programem. Searle říká, že jeho rozumění není dáno tím, že by byl instanciací nějakých počítačových programů – což je –, ale tím, že je určitým druhem organismu s jistou biologickou strukturou, která je za jistých podmínek schopna vytvořit intencionální fenomény, tedy například ono rozumění. A jen to, co má zmíněné kauzální schopnosti, může mít zmíněnou intencionalitu. Žádný čistě formální model nebude nikdy dostatečným podkladem pro intencionalitu, protože žádné formální vlastnosti samy o sobě neustavují intencionalitu, ani samy o sobě nemají jiné kauzální schopnosti, než přejít do dalšího formálního stavu. Podle Searla nejde o stín formálnosti v sekvencích probíhajících synapsí, ale o skutečné vlastnosti těchto sekvencí. Viz [3].

Podle Searla stroje mohou myslet, neboť i člověk je v podstatě strojem. Dále Searle říká, že i lidmi stvořené stroje mohou myslet, pokud se nám povede vytvořit umělý nervový systém obdobný tomu našemu. Stejně tak může podle Searla myslet digitální počítač – tedy takový představovaný spuštěním počítačového programu, tedy stejně jako jí jsme my. Podle Searla ovšem nemůže myslet myslet nic, co by bylo z podstaty jen počítačem s instanciací správného programu. Něco takového podle Searla není dostatečnou podmínkou rozumění. Bezvýznamové formální manipulace symbolů prostě nenesou sémantiku, pouze syntaxi, tedy nemají intencionalitu samy o sobě, jen jim ji lidé z vnějšího pohledu přisuzují. Právě neschopnost formálního programu přidat systému nějakou intencionalitu ukazuje Searlův příklad s čínským pokojem. Viz [3].

Searle zdůrazňuje rozdíl mezi programem a jeho realizací a bourá rovnici: „Mysl se má k mozku, jako program k hardwaru." Jednak rozdíl mezi programem a jeho realizací umožňuje, aby program měl mnoho různých realizací, z nich některé nemají vůbec žádnou intencionalitu – např. zmíněné potrubí. Navíc ani objekt se správnými předpoklady – monolingní angličan – žádnou další intencionalitu od programu nezíská – čínsky se nenaučí. Za druhé, program je čistě formální, kdežto intencionální stavy nejsou. Intencionální stavy jsou definovány kontextem, nikoliv formou. A konečně mentální stavy jsou produktem mozku, ale programy nejsou produktem počítače. Viz [3].

Dále Searle zdůrazňuje, že simulace není duplikace. U simulace stačí mít správný vstup a odpovídající výstup. Simulace nic neříká o skutečném obsahu. Searle se pak ptá, proč lidé zaměňují počítačovou simulaci s duplikací. Přisuzuje to zmatku kolem tzv. zpracování informací. Říká se, že lidský mozek, stejně jako počítač, zpracovává informace. Faktem však je, že z hlediska počítače o informace nejde – počítač jen zpracovává formální symboly. Searle zde také zmiňuje, že počítači neschází nějaké informace vyššího řádu o významu symbolů, se kterými pracuje – i to by podle Searla byly jen další bezvýznamové symboly. Searle pak navrhuje podmínit definici zpracování informací intencionalitou. Další příčinu vidí Searle v behavioralismu, na jehož základě přiřazujeme intencionalitu jen na základě vnějších projevů, příkladem čehož je podle Searla Turingův test. Jako poslední důvod pak Searle uvádí zakořeněnou představu o dualismu mysl – mozek (tělo). Podle Searla však není mysl nezávislá na svém biochemickém podkladu, tedy mozku. Mozek podle Searla produkuje intencionalitu a nedělá to instanciací nějakého programu. Podle Searla jedině stroj může myslet, a to takový stroj, který má kauzální schopnosti mozku. Viz [3].

4.5 Úrovně izomorfismu a jejich formálnost

Jacquette tvrdí, že Searlův čínský pokoj je srovnatelný s rodilým čínským mluvčím jen na makroúrovni funkcionality, tedy stejnými výstupy a vstupy. Na mikroúrovni si tyto dva případy neodpovídají vůbec. Muž čtoucí anglické instrukce a slepě manipulující s čínskými znaky, je zcela odlišný od elektrochemických signálů v těle Číňana. Jacquette tedy považuje Searlův experiment za nerelevantní právě vzhledem k rozdílné mikroúrovni. Pro zdůraznění tohoto rozdílu prezentuje Jacquette analogii s mlčenlivým Číňanem a odpadkovým košem. Tento Číňan, ač o vstupní větě celý den úporně přemýšlí, ji nakonec pouze zopakuje. Papírek s napsanou větou lze stejně tak dobře hodit ráno do koše a večer ho z něj vyjmout. Mikroúrovňově izomorfní systém k rodilému čínskému mluvčímu podle Jacquetta nejspíše nemá žádné ústředí, kde by probíhala transformace vstupu na výstupy, jak to naznačuje Searlův čínský pokoj. Transformace je nejspíše distribuovaná. Viz [4, str. 606 – 610].

Searle odpovídá, že rozdíly v mikro a makroúrovni nejsou důležité. Obojí je totiž realizováno formálním programem a ten není dostatečnou podmínkou pro intencionalitu, tedy ani pro myšlení, bez ohledu na izomorfismus. Searle na tomto místě dává jasnou podobu své myš-

lence: Není pravda, že pro všechny programy platí, že program implikuje mysl.[12] Důležitá je podle Searla čistě formální a syntaktická podoba programu, nikoliv to, zda je zpracováván centrálně či distribuovaně. Viz [5, str. 701 – 704].

4.6 Logické a biologické vlastnosti intencionality

Searlova teorie intencionality tvrdí, že zpracování informací není inteligentní, pokud není také intencionální, a že pouhé spuštění programu zpracování informací není dostatečné pro vytvoření intencionality, pokud není implementovaná hmotným systémem se správnými kauzálními schopnostmi. Toto kritizuje Dale Jacquette, který se přiklání k brentanovskohusserlovskému pojetí intencionality jako kauzálně neredukovatelné abstraktní relace mysli a myšleného. Viz [4, str. 610 – 622].

Searle však nevidí rozpor mezi dvěma následujícími stanovisky. Svou povahou je intencionalita podle Searla biologický fenomén způsobený procesy probíhajícími v mozku a realizovaný jeho strukturou. Zároveň intencionalita zahrnuje mnoho abstraktních relací, které pracují s abstraktními fenomény. Podle Searla totiž nejde o dvě odpovědi na stejnou otázku ale o stanoviska k poněkud odlišným problémům. První z nich zařazuje fenomény intencionality do zbytku světa, tedy ukazuje jejich vztah k ontologii. Druhý se zabývá logickou strukturou intencionality. Tyto dva aspekty jsou pak jádrem Searlovy koncepce intencionality. Biologickou povahu intencionality vidí Searle jako zřejmý fakt – jde tedy o úvodní axiom, na kterém staví svou teorii, nikoliv o závěr. Příčinu odporu vůči tomuto stanovisku vidí Searle v zažitém karteziánském dualizmu, který vyčleňuje mysl mimo svět biologických fenoménů odehrávající se v našem mozku. Bez karteziánského dualizmu je intencionalita podle Searla již zcela naturalizovaná. Podle Searla je možné, aby něco mělo jak nereduktivní logické vlastnosti intencionálního fenoménu, tak i bylo fenoménem biologickým. Podle Searla je dále třeba rozlišovat kauzální redukci sloužící k vysvětlení vyššího fenoménu a ontologickou eliminační redukci. Jacquettovy připomínky se týkají kauzální redukce, ale vyšší fenomén při této redukci nezaniká – tedy je ontologicky nereduktivní. Intencionalita podle Searla není z ontologického hlediska abstraktní, ale jde o součást reálného světa. Tento biologický fenomén však má logické vlastnosti, kterými reprezentuje věci a dění reálného světa, – a z tohoto pohledu je abstraktní. Zde však Searle varuje před „omylem posledních tří set let západní filosofie“ a to domněnkou, že cokoliv nesoucí tyto abstraktních vlastností není a nemůže být součástí reálného fyzického světa. Jako vhodnějšího kandidáta na neredukovatelný charakteristický rys duchovna vidí Searle spíše vědomí. Viz [5, str. 704 – 708].

5 Churchlandovi

5.1 Neporovnatelnost úrovní reality

Úvodem svého článku Churchlandovi zopakují vývoj na poli výzkumu umělé inteligence. Ukazují tak formulaci ústřední otázky klasické umělé inteligence: Může stroj, který manipuluje fyzickými symboly podle syntaktických pravidel, myslet? Důvody pro kladnou odpověď na tuto otázku pak spočívají v Church-Turingově tezi vypočitatelnosti. Z ní plyne, že digitální počítač se správným programem, dostatkem paměti a času dokáže vypočítat jakoukoliv pravidly řízenou funkci se vstupem a výstupem. Počítač tak může systematicky reagovat na své okolí a tedy i splnit Turingův test. Klasická umělá inteligence si dává za cíl najít vhodnou funkci lidského reagování na okolní prostředí a zapsat ji programem. Klasická umělá inteligence zaměřená na

[12] V originále „It is not the case that (necessarily (program implies mind)).“ [5, str. 703].

porovnání vstupů a výstupů a abstrahující od architektonických detailů stroje slavila z počátku značné úspěchy. Viz [6, str. 32 – 33]

Pak se Churchlandovi zaměřují na argumenty proti klasické umělé inteligenci. Kromě evolučně překonaného dualizmu uvádějí Dreyfusovu a Searlovu kritiku. Dreyfus klasické umělé inteligenci vyčítá absenci báze základních znalostí a schopnost získat z ní relevantní údaje k aktuální situaci. Opodstatněnost této kritiky se postupně ukázala, jak začaly počítače relativně hůře řešit složitější úkoly, ve kterých je lidský mozek snadno překonával. Pří úlohách jako rozpoznávání objektů se také začal objevovat nedostatek v architektonickém návrhu počítačů. Searlova kritika se pak zaměřila na to, zda vůbec může být manipulace se symboly podstatou vědomé inteligence. Churchlandovi se zaměřují na dva úzce související aspekty Searlova čínského pokoje: Searlův axiom, že syntaxe o sobě nezakládá ani nepostačuje pro sémantiku a řádový rozdíl v rychlosti zpracování instrukcí mezi čínským pokojem a počítačem. Na podporu svého argumentu uvádějí Churchlandovi mnoho případů, kdy si myslitelé nedokázali např. představit, že by pouhé částice o sobě zakládaly objektivní fenomén světla. Toto demonstrují paralelou s luminózním pokojem, která je vystavěna stejně jako Searlův čínský pokoj, jen se týká elektromagnetizmu a světla. [13] Oscilující elektromagnetické síly jsou přesto podstatou světla, i když magnet, se kterým pohybuje člověk v temné místnosti, žádné světlo nevytváří. Rychlosti oscilace jsou totiž v obou případech neporovnatelné. Vlnová délka a intenzita elektromagnetických vln je tak řádově příliš vzdálena schopnostem lidského vnímání. Podle Churchlandových si tedy Searlův axiom o vztahu syntaxe a sémantiky žádá vysvětlení, které čínský pokoj zdaleka nepodává. Viz [6, str. 33 – 35]

5.2 Masivní paralelizace neuronových sítí

Svou vlastní kritiku klasické umělé inteligence zakládají Churchlandovi na výkonnostních selháních klasických modelů a poznatcích a modelech odvozených z biologické struktury mozku. Jako příčinu těchto selhání vidí nevhodnost funkční architektury klasických počítačů pro úkoly, které běžně řeší lidský mozek. Při porovnání architektury počítače a mozku odhalili tři zásadní rozdíly: nervový systém je paralelní; neuron je oproti CPU značně jednodušší a navíc analogový; zpracování není jednosměrné, ale má zpětnou vazbu, která prostřednictvím změny senzorů upravuje vstup. Z těchto poznatků vycházející umělé neuronové sítě pak jsou v některých oblastech značně výkonnější, odolnější proti chybám a rychleji přistupují k uloženým informacím, než klasické modely. Rozlišily se tedy dva výpočetní obory: klasické zpracování s malým vstupním vektorem, na kterém je potřeba provést mnoho operací (především matematické výpočty), a paralelní zpracování s rozsáhlým vstupním vektorem a menším množstvím na něm prováděných operací (každodenní operace prováděné živými tvory). Neuronové sítě navíc primárně neoperují v módu manipulací se symboly, i když se to mohou naučit. Churchlandovi pak nevidí žádný důvod, proč by umělá neuronová síť vytvořená podle struktury mozku nemohla myslet. Viz [6, str. 35 – 37]

Churchlandovi chápou mozek jako počítač, který zpracovává složité funkce, avšak počítač radikálně odlišný od běžného digitálního, sériově pracujícího programovatelného počítače. Teorii významu je pak podle nich třeba ukotvit v neuronové struktuře mozku, a proto je potřeba zjistit o jeho fungování mnohem více. Churchlandovi se shodují se Searlem v tom, že lze vytvořit umělou inteligenci na základě znalosti neuronové struktury mozku. Takový počítač by

[13] Axiomy luminózního pokoje jsou následující:

1. Elektřina a magnetizmus jsou síly.
2. Esenciální vlastností světla je jas.
3. Síly o sobě nezakládají ani nepostačují pro vznik jasu.

Závěr je pak Searlovský: Elektřina a magnetizmus nezakládají ani nepostačují pro vznik světla.

pak měl mít, jak tvrdí i Searle, všechny relevantní, jak explicitně říkají Churchlandovi, kauzální schopnosti mozku pro vznik vědomé inteligence. Stejně jako Searle i Churchlandovi odmítají Turingův test jako postačující podmínku vědomé inteligence. Kromě shody ve vstupech a výstupech je také důležité, aby se uvnitř stroje děl ten správný druh věcí. Churchlandovi však hovoří především o podobnosti v architekturách stroje a mozku, tedy nasazení masivní paralelizace. Viz [6, str. 37]

6 Rapaport

6.1 Argument abstrakce a implementace

Rapaport se ve svém článku [7] říká, že rozdíl mezi simulovaným myšlením počítače a lidským myšlením spočívá v rozdílném médiu implementace abstraktního myšlení. Svou podstatou však jde v obou případech o myšlení. Při tom vychází ze Searlova pojetí intencionálních stavů jako způsobených neurofiziologií mozku a realizovaných v ní. Právě tuto realizaci chápe Rapaport v termínu počítačové vědy jako implementaci abstraktního datového typu. Viz [7, str. 341].

Pro ujasnění terminologie poskytuje Rapaport shrnutí teorie datové abstrakce a implementace. A sice: „Program popisuje akce prováděné s objekty. . . . objekty jsou reprezentovány datovými strukturami . . . Datové struktury lze klasifikovat do různých datových typů. Abstraktní datový typ je pak formální datová struktura spolu s různými charakteristickými operacemi, které s ní lze provádět. Implementace abstraktního datového typu je obvykle konkrétní datová struktura v programu . . . “ [7, str. 341]. Přičemž typickou vlastností pro abstraktní datový typ je existence více způsobů implementace. Tato teorie je podobná aristotelskému chápání druhu a jedince daného druhu. Esenciální vlastnost abstraktního datového typu však může být pouze akcidentální vlastností jeho implementace. Zároveň se dvě různé implementace abstraktního datového typu mohou lišit i jinými než pouze akcidentálními vlastnostmi. Při přesunu mimo informační vědu přechází Rapaport od pojmu abstraktní dataový typ k pojmu abstrakce. Jako příklad vztahu abstrakce a implementace pak uvádí vztah hudební partitury a hudby podle ní zahrané nebo nahrané. Viz [7, str. 342].

Nyní Rapaport přechází k interpretaci Searlova pojetí intencionality respektive inetencionálních stavů. A realizované pomocí B chápe Rapaport jako abstrakci A implementovanou v B. Dále se zaměří na vztah tohoto nového pojetí realizace a Searlova pojetí způsobení. Rapaport dochází k závěru, že si tyto dva vztahy neodporují, avšak vztah způsobení nutně nepodmiňuje vtah realizace. Z příkladu zásobníku realizovaného polem – tedy reálného zásobníku „přivedeného k existenci“ reaálným polem – vyvozuje, že je-li abstrakce A realizována pomocí B, pak je reálné A způsobeno B, respektive reálné A je „převlečené“ B. O implementaci abstrakce dále Rapaport uvádí, že není její redukcí ani eliminací a je tedy konzistentní se Searlovými neredukovatelnými a neeliminovatelnými mentálními fenomény, pokud jsou tyto chápany jako abstraktní. Rapaport vidí rozpor v Searlově tvrzení, že stavy mysli jsou jak způsobeny operacemi mozku tak realizovány v jeho struktuře. Ze Searlova tvrzení, že stavy mysli jsou reaálné jako všechny jiné biologické fenomény, pak vyvozuje, že Searle pouze nevyjádřil své rozdílné chápání těchto mentálních stavů. Rapaport tedy opravuje Searla následovně: „. . . implementované stavy mysli jsou způsobeny operacemi mozku a . . . abstraktní stavy mysli jsou realizovány v jeho struktuře.“ [7, str. 343]. Zde Rapaport zdůrazňuje, že abstrakce může být implementována v různých médiích a že i když se tyto implementace svou vnitřní strukturou liší, jde ve všech případech o realizaci abstrakce. Tedy abstraktní rozumění je podle Rapaporta implementované v lidech a lze ho implementovat v počítačích a v obou případech půjde o skutečné rozumění. Viz [7, str. 342 – 343].

Rapaport tedy shrnuje: „Strojové rozumění je, skutečně, rozumění, stejně jako je jím i rozumění lidské: Obě jsou instancemi abstraktní (funkcionální či aritmetické [14]) charakterizace rozumění. Strojové rozumění je způsobeno počítačem (nebo počítačovým programem), ve kterém je realizované abstraktní rozuměníi, tedy strojové rozumění je instancí abstraktního rozumění, které je realizováno počítačem.“ [7, str. 344]. Searlovy kauzální schopnosti mozku nejsou podle Rapaporta skutečnou kauzalitou, ale složitějším vztahem. [15] Podstatný je vztah realizace. Viz [7, str. 343 – 344].

6.2 Syntaktická sémantika

Rapaportova esej [8] se zaměřuje na oblast rozumění přirozenému jazyku. K tomu je potřebná určitá sémantická interpretace přirozeného jazyka, která se však podle Rapaporta získává toliko manipulací se symboly na syntaktické úrovni – čehož je schopný i počítač. A tedy počítače podle Rapaporta v principu mohou rozumět přirozenému jazyku. Viz [8, str. 81].

Úvodem Rapaport poukazuje na rozdíly mezi počítačem, programem a procesem. Počítač ani program sám o sobě nic dělat nemůže, teprve proces – tedy program běžící na určitém počítači. Toto rozlišení je třeba mít napaměti kdykoliv se hovoří o počítačích a rozumění, či programech a rozumění. Rozumění přirozenému jazyku je znakem inteligence, v případě programu pak umělé. Rozumění přirozenému jazyku je podle Rapaporta nutnou i postačující podmínkou pro úspěch v Turingově testu. Rozumění přirozenému jazyku znamená jistě rozumění jednotlivým výrazům, ale také vyhodnocovat „vnořená přesvědčení“ [16]. Právě uvažování o názorech jiných je potřeba k jejich napodobení – tedy k úspěchu v Turingově testu. Viz [8, str. 81 – 84].

Nyní se Rapaport ptá, co to znamená rozumět. Rozumění spojuje s významem, tedy sémantikou. K většině sémantiky je podle Rapaporta potřeba pouze syntaxe, ke zbytku pak kauzální vazba [17] k vnějšímu světu. Tyto vazby jsou dány vnímáním a mají především vizuální a zvukovou podobu. Pro rozumění jazyku však podle Rapaporta stačí pouze syntaktická sémantika. Je tomu tak především proto, že účastníci konverzace nemají přístup ke kauzálním vazbám svých protějšků, pouze ke své vlastní reprezentaci těchto vazeb. Z hlediska počítače mluví Rapaport o jeho „mysli“, kterou je znalostní báze o vnějším světě. Základ báze je do počítače vložen jeho tvůrcem. Programy ji dále obohacují o informace získané konverzací v přirozeném jazyce, nebo odvozováním nad obsahem báze. Viz [8, str. 84 – 86].

Searlovi přičítá Rapaport pojetí významu jako vztahu mezi symbolem a tím, co symbolizuje. Rozvinutím robotické odpovědi na Searlův čínský pokoj dochází Rapaport k tvrzení, že význam je pro systém dán vztahy symbolů k jeho vnitřní reprezentaci toho, co symboly zastupují.[18] Tedy Searlovo tvrzení, že počítač je čistě syntaktickou entitou, neznamená, že nemůže rozumět přirozenému jazyku, nýbrž právě naopak. Rapaport dále přistupuje k metafoře přirovávající program rozumějící přirozenému jazyku k herci v improvizační hře se zapojením publika. Scénář takové hry – tedy program – se musí měnit podle podnětů od publika. A aby ho herec dokázal měnit, musí rozumět vstupům publika a být schopen vytvořit novou relevantní výpověď, k čemuž potřebuje schopnost plánovat. To, co pak bude herec – počítač – shopen říct, záleží také na tom, co ví, tedy na jeho bázi znalostí. Nyní Rapaport přistupuje k rozdílu mezi explicitní a implicitní znalostí. Člověk ani počítač podle Rapaporta explicitně neví vše, ví jen to, čeho si je vědom. Znalost jazyka je u člověka i u počítače primárně tacitní. S koverzací v přirozeném jazyce pak souvisí jak explicitní tak implicitní učení. Pro rozumění přirozenému

[14] V originále „computational“.

[15] Tento vztah Rapaport upřesňuje jako Castañedův vztah konsubstanciace, případně navrhuje další jako Cambridgskou závislost či inverzní instanciaci.

[16] V originále „nested beliefs“.

[17] V originále „causal link“.

[18] Vnitřní reprezentaci vnějšku může robot získat svými senzory. Jde stále o symboly, ale o symboly jiného druhu.

jazyku je tedy modifikovatelná báze znalostí esenciální. V Searlově pojetí čínského pokoje není jasné, zda je tato modifikovatelná báze znalostí přítomná. Viz [8, str. 86 – 91].

Jako jiný příklad chápání rozumění uvádí Rapaport Dretskeho.[19] V reakci na Dretskeho volání po nutnosti symbolů mít význam pro systém, který s nimi manipuluje, uvádí Rapaport vnitřní a vnější sémantiku. Vnitřní sémantiku chápe Rapaport jako danou vnitřními reprezentacemi vnějšího světa v sémantické síti. Bere ji jako nezávislou na vnější sémantice, tedy významu připsaném z venčí někým dalším. Podle Dretskeho pak počítače nechápou to, co dělají, podle Rapaporta díky sémantice vycházející ze syntaxe ano. Zde je třeba zdůraznit Rapaportovo tvrzení, že se vnitřní význam liší jak mezi jednotlivými lidmi, tak i mezi lidmi a počítači. Pro úspěšnou komunikaci je pak ale potřeba, aby si tyto vnitřní významy byly dostatečně podobné – tedy aby si byly dostatečně podobné sémantické sítě, což opět akcentuje nutnost učení a komunikace. Viz [8, str. 91 – 100].

Rapaport nyní zpřesňuje své tvrzení takto: vnitřní sémantika je nutná pro porozumění jazyku, vnější sémantika je nutná pro vzájemné porozumění. Rapaport odpovídá na otázku, co to znamená rozumět přirozenému jazyku následovně: alespoň částečně poskytnout sémantickou interpretaci k syntaxi. Co pak ale znamená pro dva systémy (ať už lidské nebo umělé) rozumět jeden druhému? Rapaport zvažuje tři následující případy: Člověk, který má rozumět jinému člověku, musí podle Rapaporta sémanticky interpretovat jeho výpovědi přiřazením ke svým konceptům ve své sémantické síti. Při tom samozřejmě může dojít k chybné interpretaci – přiřazení na špatné místo sítě. Stejný koncept mohou mít různí lidé přiřazený na různých místech. Pokud koncept nejde přiřadit přímo, snaží se ho subjekt přiřadit tak dobře, jak to dokáže. Takto pojaté rozumění – přiřazování výpovědí k symbolům sématické sítě – je pak podle Rapaporta syntaktický proces. Člověk, který má rozumět formálnímu jazyku – či systému, může podle Rapaporta uplatnit buď sémantické, či syntaktické rozumění. Syntaktické rozumění vychází z přímého manipulování symboly formálního jazyka na základě jeho syntaktických pravidel. Sémantické rozumění se uskutečňuje přes sémantickou interpretaci této syntaxe. Stejně dobrých sémantických interpretací je ale podle Rapaporta mnoho, přičemž nelze určit, která je ta zamýšlená. Interpretace formálního jazyka je pak ve své podstatě opět mapování jeho termínů na koncepty sémantické sítě. Jak syntaktické tak sémantické rozumění je tedy svou podstatou syntaktický proces. Formální systém, který by měl rozumět člověku, bude podle Rapaporta opět sémanticky interpretovat lidskou výpověď, tedy přiřazovat slova do své sémantické sítě. Program tedy bude čistě syntaktickým způsobem přiřazovat významy k symbolům. Rapaport tedy shrnuje, co to znamená pro dva systémy rozumět si navzájem: „Systém S_1 rozumějící přirozenému jazyku rozumí výpovědi systému S_2 v přirozeném jazyce tak, že vytváří a manipuluje symboly svého interního modelu (interpretace) výstupu systému S_2, přičemž uvažuje tento výstup jako by byl formálním systémem.“ [8, str. 104]. K internímu modelu Rapaport dále poznamenává, že jde o systém reprezentace znalostí s podporou usuzování. Dále upřesňuje vztah interní a externí sémantiky. Interní sémantika systému S_1 zahrnuje to, jak systém S_1 rozumí systému S_2. Externí sémantika systému S_1 pak zahrnuje to, jak systém S_2 rozumí systému S_1. Externí sémantika systému S_1 je podle Rapaporta interní sémantikou systému S_2. Referenční sémantika – vztah slova k odkazované věci – do rozumění podle Rapaporta nevstupuje přímo, ale je vždy zprostředkován reprezentací. Viz [8, str. 100 – 105].

[19] Dretske uváděný Rapaportem tvrdí, že strojům chybí něco podstatného, co jim znemožňuje být racionálními agenty. Strojům pak vyhrazuje teoretické myšlení – v originále „pure thought“ – a lidem běžné myšlení – v originále „ordinary thought“. Rapaport reaguje uplatněním principu abstrakce a implementace a strojům by přiznal implementaci abstraktního teoretického myšlení, jejíž provedení u lidí se označuje jako běžné myšlení. Dretske pak říká, že počítače ani neprovádí sčítání. Sčítání chápe jako operaci s čísly, kdežto počítače operují se zástupnými znaky – které lidé interpretují jako čísla. Počítače pak tedy nemohou sčítat. To, co počítače provádí, je pouze behaviorálně identické s lidským sčítáním. Rapaport nyní rozebírá, kdo tedy interpretuje ony zástupné znaky. Při přístupu zevnitř je rozhodující interpretace počítače, pří přístupu z vnějšku pak člověka. Rapaport se také zaměří na to, zda lidé, když sčítají provádí operaci s čísly, nebo jen s jejich zástupnými znaky. Rozvíjením Dretskeho úvahy dochází k tomu, že pak ani lidé nesčítají, a opět nabízí svůj argument abstrakce a implementace jako cestu z této slepé uličky.

Nyní Rapaport přechází k popisu jedné možné realizace systému pro rozumění přirozenému jazyku: SNePS/CASSIE. Pro více informací viz [8, str. 106 – 111]. Rapaport rozlišuje dva druhy významu. Primárně je význam dán pozicí uzlu v sémantické síti. Uzel lze nahlížet jako skupinu vlastností, tedy obsah. Tento interní význam slova se mění v čase, jak se mění sémantická síť, tedy ho lze nahlížet i jako rozsah (použití daného slova). Definiční význam slova pak zahrnuje podmnožinu takových významů slova, ze kterých lze ostatní odvodit. Jde tedy o interní obsahový význam. Oba tyto druhy významu se podle Rapaporta skládají z interně manipulovatelných symbolů. Pro tvorbu interní sémantické sítě je podle Rapaporta velice důležitá souslednost vět – každá následující věta je interpretována ve světle interpretace té předchozí. Viz [8, str. 111 – 113].

Rapaport shrnuje své stanovisko následovně: Systém rozumí jazyku prostřednictvím vnitřního modelu vnějšího světa, o kterém jazyk vyjadřuje informace. Přitom systém využívá funkci sémantické interpretace výpovědí v daném jazyce do svého modelu světa a funkci generující výpovědi v jazyce z tohoto modelu. Tyto dvě funkce – parser a generátor – spolu s modelem světa – znalostní bází – vytváří základ pro rozumění přirozenému jazyku. Model světa daného systému má také vnější lidskou interpretaci, která ale nemá vliv na to, jak systém rozumí. Dále existuje kauzální vztah mezi vnějším světem a jeho modelem (výpověď v jazyce, ze které se tvoří model, je o světě), který podle Rapaporta není limitován jen na biologické entity, jak to tvrdí Searle. Rozumění jazyku však na tomto kauzálním vztahu nezávisí, vztah je externí. Když je vytvořena reprezentace světa, přestává být kauzální vztah vůbec relevantní pro rozumění, může ale mít jistý smysl pro komunikaci. Sémantika, která je pro rozumění jazyku potřeba je dodána jeho syntaxí, tedy je komputerizovatelná. Viz [8, str. 119 – 120].

Argument syntaktické sémantiky Rapaport dále upřesňuje a rozšiřuje v článku [9]. Sémantické rozumění zde označuje jako rekurzivní vysvětlení významu termíny dalších oblastí až k základním termínům, kterým je rozuměno v rámci nich samotných. Jsou tu tedy naznačeny dva typy rozumění: Korespondeční rozumění (relativní vzhledem k rozumění jinému, tedy externě) a koherenční rozumění (nerelativní, resp. relativní samo k sobě, tedy interně). Korespondenční rozumění vychází ze shody dvou oblastí, koherenční pak ze syntaxe. Korespondenční rozumění se po delší či kratší rekurzi musí dopracovat ke koherentnímu rozumění. Podle Rapaporta se poruzumění novému neznámému rovná snaze toto přirovnat k něčemu známému (tedy najít použitelnou korespondenci). Řetězené termíny vysvětlené (tedy modelované) pomocí termínů, které jsou zase vysvětlené dalšími termíny, vystupují vůči svým předchůdcům v sémantické roli, zatímco vůči následovníkům v roli syntaktické. Chápání řetězu ale lze otočit, vymezení syntaktické a sémantické role je tedy vzájemně relativní. Rapaport chápe vztah mezi sémantikou a korespondencí jako konotaci. Syntaktickou roli nemusí zastávat jazyk, musí ale být rozčlenitelná do symbolů. Viz [9, str. 49 – 60].

Při rozumění se tedy uplatňuje modelování světa. Rapaport zde říká, že modelu je nutné rozumět přednostně, před porozuměním světu, a musí tedy řešit, jak rozumět prvotnímu modelu. Rapaportovo stanovisko je, že si subjekt na prvotní model zvykl. Viz [9, str. 65 –67].

Je-li rozumění mapováním symbolů syntaktické domény do domény sémantické,[20] pak zvyknutí si je uchopení vlastní syntaktické domény jako by byla sémantická – tedy mapování domény samé do sebe. Jde tak tedy o vztah syntaktický mezi symboly jedné domény, ale i sémantický – korespondenci mezi dvěma různými rolemi této domény. Rapaport zde také poznamenává, že není vůbec důležité, zda je prvotní význam daný zvyknutím úplný či vůbec správný – při dalším výskytu symbolu dochází k jeho upřesnění, až vykrystalizuje poměrně odpovídající význam. Jednotlivé prvky mentální sítě, které určují významy ostatních, podle Rapaporta nemají samy žádný význam jen ze své podstaty. Tím, jak se objevují v dalších a dalších kontextech, teprve nějaký – stále se upřesňující – význam získávají. S takovýmto pojetím rozumění a tvorby významů souvisí problém kruhovosti definic a ukotvení termínů. Definice

[20] Jedná se o role ve vztahu nikoliv absolutní sémantickou či syntaktickou doménu.

prvku domény – slovo ve slovníku – jinými prvky téže domény – jinými slovy ve slovníku – je kruhová. Podle Rapaporta je ale důležitá rozlehlost takového kruhu. Kruh tvořený dostatečným počtem prvků je i tak informativní, tedy ukotvený. Malý kruh však ukotvení postrádá. Ukotvení se podle Rapaporta vždy děje prostřednictvím vnitřních reprezentací vnějšího. Celá síť významů je tedy uzavřená, tudíž tak může vzejít sémantika ze syntaxe. Viz [9, str. 74 – 81].

6.3 Úhly pohledu

Rapaport používá argument úhlů pohledu v několika polohách. Ve svém článku [9] se zabývá rolí úhlu pohledu při rozumění něčemu. Ve svém dřívějším článku [8] a své úvaze [10] pak řeší úlohu úhlů pohledu při stanovení, zda někdo něčemu rozumí.

Rozumění něčemu chápe Rapaport jako modelování, viz předchozí sekci. Tento model pak vidí v souladu s B. C. Smithem jako abstrakci situace z reálného světa, tedy nutné zlo, které potřebným způsobem zužuje pohled na věc. Modely jsou tedy neodmyslitelně částečné. Myšlení samo o sobě částečné není, ale je třeba brát v úvahu, že pracuje s částečnými modely – stejně tak komputerizace. Myšlení tak předkládá úhel pohledu, který zachycuje jen fragment kontinuálního celku. Mezi modelem a světem vzniká jakási propast. Jak tuto propast překonat? Chápe-li Rapaport svět jako sémantiku a model jako syntaxi, navrhuje zaujmout nezávislý vnější pohled. Tímto je pro něj určitý jazyk, který má stejný přístup k oběma oblastem: světu (sémantice) i modelu (syntaxi). Z pohledu modelu a jeho termíny není totiž vztah modelu a světa popsatelný. Viz [9, str. 60 – 65].

Rapaport rozlišuje dva způsoby učení se významům termínů: Zakušením toho, k čemu se termín vztahuje – tj. z vnějsího fyzického kontextu a lexikální učení slov z lingvistického kontextu, v jakém jsou použity. Při učení se tento kontext zvnitřňuje. Rozumění je tak rekurzivní ve světle všeho toho, čemu již bylo rozuměno. Pro popsání vztahu mezi větami a nelingvistickými fakty, je podle Rapaporta potřeba zaujmout „pohled třetí osoby“. Z tohoto úhlu pohledu je vidět, jak mysl mluvčího, tak vnější svět, které je potřeba usouvztažnit. Třetí osoba ale přistupuje, jak k mysli mluvčího, tak ke vnějšímu světu přes svou reprezentaci. Takto může vytvořit sémantickou korespondenci těchto dvou oblastí. Ty jsou však pro třetí osobu interní. Relace vzniká mezi dvěma množinami symbolů v sémantické síti uvnitř mysli. Původní mluvčí může mít pouze vlastní domněnky o tom, že jeho úvahy jsou správné. Viz [9, str. 67 – 74].

K ilustraci problému při rozhodnutí, zda někdo něčemu rozumí, pak Rapaport použije následující příklad: Korejský profesor anglické literatury nerozumí anglicky, ale studuje korejské překlady Shakespearových děl a píše o nich v korejštině články. Angličtí odborníci na Shakespeara, kteří čtou překlady těchto článků, pak korejského profesora uznávají jako odborníka na Shakespearova díla, tedy někoho, kdo rozumí Shakespearovým dílům. Obdobně vnímá Rapaport i Searlův argument čínského pokoje. Searle v čínském pokoji ve skutečnosti něčemu rozumí. Provádí totiž sémantickou interpretaci přirozeného jazyka na základě jeho syntaktických pravidel, tedy systém podle Rapaporta rozumí přirozenému jazyku. Dále lze úvahu rozvíjet dvěma způsoby. Protože systém rozumí přirozenému jazyku a tím je čínština, pak systém rozumí čínštině. Ovšem tento systém sdílí jen velmi málo čínské kultury, tedy možná jen rozumí kódu, do kterého mu program překládá čínské znaky. Jak ale ukazuje příklad s korejským profesorem, který rozumí Shakespearovi, ne jen korejským překladům Shakespeara, tedy i systém v čínském pokoji rozumí čínštině, ne jen kódu programovacího jazyka. Rapaport dodává, že systém rozumí čínštině alespoň tak dobře jako každý nerodilý mluvčí. A znovu připomíná, že lidé nikdy nemohou zcela přesně mínit to, co ti druzí, ale to neznamená, že si nemohou navzájem rozumět. Jejich sémantické sítě se během konverzace dostatečně sladí. Viz [8, str. 114 – 116].

Dále se Rapaport ptá, zda počítač může rozumět tomu, že rozumí, respektive, zda může rozumět tomu, že rozumí přirozenému jazyku. Rapaport říká, že program nemusí rozumět

tomu, že rozumí přirozenému jazyku, aby opravdu rozuměl přirozenému jazyku. Ono rozumět přirozenému jazyku je podle Rapaporta pouze nálepka pro danou činnost. Pro ilustraci uvádí případ studenta vykonávajícího program Turingova stroje. Tento student nemusí vědět, že ve skutečnosti počítá největšího společného dělitele. Ví například jen, že vykonává program Turingova stroje. Pokud by ale Rapaport chtěl znát největšího společného dělitele, mohl by nechat studenta vykonat daný program. Student tedy podle Rapaporta rozumí výpočtu největšího společného dělitele, ačkoliv nerozumí tomu, že tomu rozumí. Rozumí tomu pod jiným názvem. Rapaport dále tvrdí, že pokud studentovi řekne, co vlastně počítá, bude student rozumět tomu, že rozumí výpočtu největšího společného dělitele. A obdobně je to pak pro Searla v čínském pokoji. Pro případ počítače Rapaport dodává, že je potřeba, aby program dokázal alespoň nepřímo spojit uzly příslušící sdělení, že rozumí přirozenému jazyku, s tomu odpovídajícími svými aktivitami. Viz [8, str. 116 – 119].

U experimentu s čínským pokojem považuje Rapaport právě úhel pohledu za nejdůležitější hledisko celého sporu. Ze svého pohledu Searle nerozumí čínsky, z pohledu rodilého Číňana však čínsky rozumí. Rapaport se ptá, který úhel pohledu v tomto případě převáží. Podle Rapaporta je to právě úhel pohledu rodilého Číňana, který je rozhodující. Právě on totiž má vzhledem k rozumění čínštině nejlepší předpoklady ke správnému rozhodnutí, co je to rozumět čínsky. Searlův úhel pohledu pak podle Rapaporta není ani úhlem pohledu nerodilého mluvčího, ale pouze části systému, jehož kognitivní schopnosti nejsou redukovatelné na každou jeho část. Searle v čínském pokoji je tak podle Rapaporta buď v roli počítače jako hardwaru, nebo dokonce jen v roli procesoru – a z tohoto úhlu pohledu pak Searle není schopen určit, zda systém jako celek – tedy obohacený o program, který je navíc Searlem vykonáván, skutečně rozumí nebo nerozumí čínsky. To, že Searle za sebe říká, že nerozumí čínsky, znamená jen tolik, že počítač jako hardware – nebo jeho procesor – nerozumí čínsky, což však Rapaport nerozporuje. Viz [10, str. 476 – 481].

7 Závěr

Otázkou lidského uvažování a s ní spjatou otázkou, zda i někdo jiný než člověk může myslet, se zabývá už na počátku 17. století francouzský filosof René Descartes. Jeho systém předpokládá značné odlišení funkcí mysli a funkcí těla. Descartes tedy v našem problému zanechává dualizmus předchozích dob. Funkce těla jsou u Descarta dány uzpůsobením těla, které zároveň Descartes přirovnává ke stroji stvořenému – jak jinak – „Bohem“. Funkce lidské mysli jsou dány její podstatou – duší. Tato duše je vlastní jen člověku a je mu dána opět oním „Bohem“.

To, co dělá člověka člověkem, je u Descarta duše. Její projevy, podle nichž můžeme člověka poznat, pak jsou tyto: schopnost vést rozumnou řeč a schopnost záměrně jednat. V obojím se podle Descarta ukazuje lidský rozum a především jeho univerzálnost. Člověkem vytvořený stroj pak podle Descarta nebude mít ani jednu z těchto vlastností, protože nebude mít danou duši.

Za nejdůležitější Descartův přínos pro rozebírané téma lze pokládat to, co pracovně nazvu „Descartovým testem“. Nemyslící stroj poznáme od myslícího člověka tak, že ověříme jeho schopnost vést rozumnou řeč stejně jako jeho schopnost záměrně jednat. Zde je potřeba pro srovnání s dalšími si všimnout, jak Descartes svou argumentaci staví: stroj není bytostí srovnatelnou svou podstatou s člověkem, protože není schopen vést rozumnou řeč a není schopen záměrně jednat.

K problému, zda stroj může myslet, se v polovině 20. století vrátil anglický matematik a informatik Alan Turing. Turing celý problém otáčí do podoby imitační hry, známé jako Turingův test, tedy klade otázku, zda dokáže stroj napodobit k nerozeznání člověka. Turingova implikace zní: Obstojí-li stroj v imitační hře, myslí. Zde se ale Turing dopouští omylu – jak ukáže

Searle, Turingovým testem projde i systém, který zpracovávanému obsahu nerozumí a o němž tedy jen stěží lze říci, že myslí.

Turing očekává, že digitální počítač se správným programem bude schopen uspět v jeho testu a jako jeden ze způsobů tvorby tohoto programu navrhuje strojové učení.

Nejsilnější argument pro Turingův test spočívá v tom, že obdobná metoda se používá pro ověření, zda student látku pochopil, nebo zda se ji naučil nazpaměť. Platnost a oprávněnost takové metody ale spočívá v její aplikaci na člověka, tedy někoho, o kom jistě víme, že má intencionalitu a kognitivní stavy.

Myšlenkový experiment s čínským pokojem, se kterým přišel v 80. letech 20. století americky filosof John Searle, je snahou podívat se problém stroje a myšlení zevnitř. Searlovi se v čínském pokoji jeví, že ač systém může zvenčí vypadat, že má intencionalitu, nemusí ve skutečnosti mít vůbec rozumění. Intencionalitu do systému mylně vkládá pozorovatel na základě vnějších znaků chování sytému. Searle tedy upřesňuje implikaci Turingova testu: Má-li zkoumaný systém intencionalitu, projde Turingovým testem. Ovšem i systém bez intencionality, tedy nemyslící, může testem projít. Z úspěchu v Turingově testu tedy nelze vyvodit myšlení.

Podle Searla je lidské tělo i mozek strojem. Searle však odmítá přísnou dualitu mysl – tělo, podle něj je mysl možná díky specifické biochemické struktuře lidského mozku. Tedy jsou to kauzální schopnosti mozku, které umožňují myšlení. Lidský mozek je však jen specifickým strojem, tedy i stroje se specifickými vlastnostmi lidského mozku mohou myslet, ať už jsou vytvořeny člověkem nebo ne.

Digitální počítače se svými formálními programy však provádějí pouze formální manipulace s neinterpretovanými symboly, což jak ukazuje Searlův čínský pokoj, do systému žádnou intencionalitu nevnáší. Sama formální manipulace se symboly nestačí k rozumění. Toto rozumění nelze podle Searla dodat dalšími symboly, které by nesly nějaké informace o informacích – pro stroj to budou jen další formální znaky bez významu. Mentální stavy jsou definovány kontextem nikoliv formálně. Digitální počítače mohou myšlení simulovat, avšak simulace není duplikací.

Searle na svém případu čínského pokoje ukazuje, že schopnost vést rozumnou řeč nestačí k tomu, aby se dalo usuzovat, že stroj myslí. Podle Searla je pro fenomén myšlení důležité, zda stroj zpracovávanému skutečně rozumí, čistě z vnějších projevů něco takového usuzovat nelze.

Kanadští neurofilosofové manželé Churchlandovi upozorňují na neporovnatelnost úrovní reality v Searlově experimentu s čínským pokojem. Rychlost, s jakou čínský pokoj provádí manipulace se symboly, je o mnoho řádů nižší než, jak je tomu u počítače. Systémy se tedy tak zásadně kvantitativně liší ve své esenciální vlastnosti, že jsou naprosto neporovnatelné. Searle tak nemůže zachytit svými smysly hledaný fenomén, protože ten – je-li – je zcela mimo jejich rozlišovací schopnost. Searlův myšlenkový experiment simulující počítač tak není adekvátní simulací počítače, protože ignoruje jednu z jeho esenciálních vlastností.

Další kritiku směřují Churchlandovi k tomu, že Searlova argumentace z velké části stojí na pojetí výroku: „Syntaxe nezakládá ani nepostačuje pro sémantiku“ jako axiomu. Toto tvrzení si podle nich žádá důkaz, který čínský pokoj nepodává.

Churchlandovi se však se Searlem shodují v kritice přístupů klasické umělé inteligence. Zaměření pouze na vstup a výstup je i podle nich nedostatečné. Na rozdíl od Searla se na problém dívají odspodu a postulují nutnost ukotvit teorii významů v neuronové síti. Vhodnou cestu pro umělou inteligenci vidí v modelech inspirovaných právě architekturou mozku, která se ukázala jako výhodná pro úkoly, jež mozek provádí. Nevidí pak žádné apriorní důvody, proč by systém využívající masivní paralelizmus umělých neuronových sítí nemohl myslet. I Churchlandovi tedy chápou mozek jako počítač.

Americký filosof a lingvista William Rapaport ukázal ve svých článcích problémy Searlova

přístupu. Jeho protiargument implementace chápe kognici jako – v termínu informační vědy – abstraktní datový typ, jehož jednou možnou implementací je kognice, kterou provádí člověk, a jinou ta, kterou provádí počítač. Člověk a počítač jsou v jeho pojetí různá média, ve kterých je možná implementace abstraktní kognice.

Rapaport vyvrací Searlovo tvrzení, že syntaxe nezakládá sémantiku. Ve svém pojetí syntaktické sémantiky definuje sémantiku jako rekurzivní binární vztah symbolu a významu. Syntaxe je pak n-ární vztah symbolů a symbolizovaného významu. Význam je tedy o vztazích mezi symboly. Tím je podle Rapaporta internalizovaný v systému symbolů, což umožňuje přechod od syntaxe k sémantice. Význam je tak daný sítí vztahů mezi jednolitvými symboly. Je-li ale význam symbolu určen dalšími symboly, vyvstává problém, jak je určen význam prvních symbolů, které se systém učí. Rapaport zde rozlišuje dva druhy symbolů: Jedněmi jsou slova, která chceme definovat. Druhými pak jsou vnitřní reprezentace vjemů našeho okolí. První soubor slov je pak definován svým propojením s vnitřními reprezentacemi vjemů. Až další slova jsou pak definována jen jinými slovy. Takto Rapaport řeší problém ukotvení sybolů ve své syntaktické sémantice.

Rozpracováním systémové námitky [21] k Searlově čínskému pokoji do podoby argumentu úhlů pohledu ukazuje Rapaport, že Searle v čínském pokoji je v roli části systému nesoucí schopnost realizovat znalosti obsažené v programu. Searle tedy ve svém myšlenkovém experimentu nezaujímá pozici počítače ale pozici procesoru. Jeho úhel pohledu tedy není pohledem počítače a nemůže tedy převážit nad úhlem pohledu vnějšího pozorovatele. V tomto kontextu je Searlova odpověď na systémovou námitku nedostatečná – co je vlastně onou Searlovou internalizací systému? A bude po ní skutečně nadále nerozumět čínsky?

To, že stroje nemohou myslet, Descartes přímo nezmiňuje, avšak vyplývá to z jeho definice lidské podstaty závislé na duši a dané „Bohem“. Turing také přímo neříká, jak je to se stroji a myšlením, optimismus je však v jeho článku patrný. Turing se především snaží najít způsob, jak ověřit, zda stroj myslí. Navrhuje také několik cest, které by mohly vést k vytvoření stroje, jenž by obstál v jeho testu. Searle pak značně umírňuje závěry, které je možné vyvodit z úspěchu v Turingově testu. Dále ukazuje, že k duplikaci mysli jen formalismy nestačí, ale že jsou potřeba specifické kauzální funkce, jakých je schopen lidský mozek. Churchlandovi upozorňují na nepřiměřenost Searlova experimentu, ale oceňují jeho snahu podívat se na problém „zevnitř“. Sami pak navrhují umělé systémy založené na principech neuronových sítí a zdůrazňují, že míra komplexity takového systému hraje pro jeho schopnosti podstatnou roli. Připouštějí tak vlastně emergenci mysli jako vlastnoti, která není patrná u jednotlivých neuronů ani u samotného faktu jejich množství, ale která je možná díky kombinaci obojího. Takovému systému pak podle Churchlandových nic v myšlení apriorně nebrání. Rapaport se snaží ukázat, jak může samotná manipulace a usouvstažnění symbolů přinést význam. Dále se soustředí především na podmínky strojového porozumění jazyku. To je podle něj totiž klíčové pro schopnost myslet. Rapaport pak rozlišuje lidské a strojové myšlení jako implementaci obecného abstraktního myšlení v různých médiích.

Přístupy ke zjišťování, zda stroje mohou či nemohou myslet, lze redukovat na povahu vztahu mezi myšlením a vnějšími projevy myšlení. Nikoliv však aktuálním myšlením a aktuálními projevy myšlení, ale schopností tyto mít. Tedy hlavní otázka zní, jaký typ vztahu je mezi schopností myslet a schopností navenek své myšlení projevovat.

Obhájci Turingova testu zjevně předpokládají, že tento vztah má povahu ekvivalence. V tom případě Turingův test platí. Pokud však tomuto vztahu přisoudíme pouze podobu implikace: Když mám schopnost myslet, mám schopnost své myšlení projevit, pozbude Turingův test relevantnosti. Neboť implikaci nelze vždy beztrestně obrátit.

[21] Fenomén myšlení není redukovatelný na konkrétní části systému, ale vzniká díky spojení všech částí systému: výpočetních schopností počítače, instrukcí programu a faktu, že program je počítačem vykonáván.

Ze závěru lze korektně uvažovat předpoklad, pokud je vztahem mezi předpokladem a závěrem ekvivalence. U implikace toto korektní není – při platnosti předpokladu i závěru však nastává něco, co lze nazvat pastí implikace. Platí-li totiž předpoklad i závěr – jako je tomu zjevně v případě tvrzení: člověk má schopnost myslet, člověk má schopnost své myšlení projevit – lze beztrestně tyto výroky spojit téměř jakýmkoliv logickým operátorem. Tedy – pro potřebu tohoto argumentu relevantní – tvrzení,

1. člověk má schopnost myslet ⇒ člověk má schopnost své myšlení projevit,
2. člověk má schopnost své myšlení projevit ⇒ člověk má schopnost myslet,

jsou obě platná.[22] Logicky korektní zobecnění pro případy všech možných kombinací pravdivostních hodnot je však možné pouze pro případ 1.[23] Přijmeme-li předpoklad, že skutečný vztah mezi schopností myslet a schopností své myšlení projevovat je implikace, nelze z úspěchu v Turingově testu vyvodit schopnost myslet.[24]

Úvaha o podstatě vztahu mezi schopností myslet a schopností myšlení projevovat však nemá dopady pouze na Turingův test. Poznání pravé podstaty tohoto vztahu určí validnost aplikace Rapaportova principu abstrakce a implementace na problém strojového myšlení. Tento princip se zdá být v podstatě dobrý, ovšem to, zda implementace řeči – jakožto projevu myšlení – do počítačového systému zakládá myšlení onoho systému, není samo o sobě zřejmé.

Pokusím se dále rozvést a ukázat argument Churchlandových týkající se neporovnatelnosti úrovní. Searle se v čínském pokoji snaží simulovat počítač. Esenciálními vlastnostmi takového počítače je to, že provádí relativně jednoduché operace (de facto manipulace se symboly) velice rychle po sobě – tedy s vysokou frekvencí. K tomu má nějaký hardware a program, který určuje, jak se bude vstup transformovat na výstup. Čínský pokoj toto vše adekvátně simuluje až na frekvenci operací. Searle sám pak frekvenci, s jakou provádí operace po sobě, zanedbává a říká, že i pomalý myslitel myslí.

Předvedu zde analogii s filmem. Co je to film? Divák odpoví, že pohybující se obraz. Filmový pásek je ale pokryt množstvím statických fotografií, které jsou promítány na plátno. Kvalita pohybu není obsažena ani v jedné z těch fotografií, ani v jejich počtu, ani v tom, že jsou promítány. Budou-li se ale střídat statické obrázky po sobě s dost vysokou frekvencí, tedy dost „rychle“, divák bude vnímat na plátně pohyb. Sníží-li se výrazně frekvence promítání, pohyb zcela ustane a divák uvidí jen jednu fotografii.

K čemu ve zde uvedené analogii došlo? Zdá se, že frekvence promítání statický obrázků je pro film klíčová vlastnost. Je-li tato frekvence příliš nízká, není před námi vlastně film ale jen jednotlivé fotografie. Od určité frekvence však před námi už nejsou jednotlivé fotografie

[22] V dané dílčí situaci jsou platná i další tvrzení jako:

- člověk má schopnost myslet ⇔ člověk má schopnost své myšlení projevit,
- člověk má schopnost myslet ∧ člověk má schopnost své myšlení projevit,
- člověk má schopnost myslet ∨ člověk má schopnost své myšlení projevit.

Za předpokladu skutečného vztahu implikace však nemá smysl se jimi zabývat a pro ukázání chyby ve vyvození Turingova testu nejsou potřeba.

[23] Pro ještě větší zjevnost tohoto argumentu přidávám mnemotechnickou pomůcku pro vyhodnocení implikace: „Když prší, je bláto – když neprší, není bláto – když neprší, je bláto (mohlo pršet včera) – ale není pravda, že když prší, není bláto.“ Z toho, zda je nyní bláto, tedy nelze zjevně usuzovat na to, že nyní prší.

[24] A to ještě za předpokladu, že Turingovým testem ověříme schopnost mít projevy myšlení, koncepce testu spíše zkoumá aktuální projevy myšlení. Negativní dopad myslícího subjektu, který své myšlení neprojeví, se snaží minimalizovat předpokladem spolupráce. Zde ale nesmíme sklouznout k přílišnému uvažování nad tím, co je možné vyvodit z vnějších projevů o vnitřních dějích. To by vedlo do „pekel“, ve kterých lze snad říct akorát tolik, že o vnitřku toho vnějšího nelze jistě říct nic, a tedy v závislosti na míře našeho optimizmu je pak buď vše možné nebo nemožné.

ale film. Kvantita – množství promítnutých obrázků v daném čase – umožnila vznik kvalitě – pohybu. V situaci, kdy mezi promítnutím dvou obrázků uplyne příliš dlouhý čas, žádný pohyb na plátně neuvidíme, ani nemůžeme. Nejsou splněny všechny předpoklady pro film. To je chyba Searlova čínského pokoje. Ať už počítač myslí, nebo nemyslí, Searle to zjistit nemůže, protože nevykonává operace ani zdaleka tak rychle jako počítač, a jeho simulace počítače je v tomto ohledu neadekvátní.

Tvrzení, že pomalý myslitel také myslí, zde neobstojí. Člověk má nějaké rozlišovací schopnosti, dokáže tedy poznat myšlení i „pomaleji" myslícího myslitele ale jen do určité míry pomalosti. Jako člověk nevidí trávu růst, tak nerozliší Searle, zda jeho čínský pokoj myslí opravdu velice pomalu, či vůbec. Rozdíl frekvencí prováděných operací je totiž neporovnatelný.

Podobnou úvahu pak lze provést i o neuronové síti. Pro ni je klíčové množství neuronů, kterými disponuje. Přesáhne-li toto množství určitou mez, projeví se vlastnosti, kterými jeden neuron nedisponuje. Podobná situace nastává v dostatečně veliké lidské společnosti, kde vzniká spontánní tržní řád. Pro simulaci takové situace, je ale klíčové toto množství. Searlovův protipříklad s mužem obsluhujícím potrubí nebo čínskou tělocvičnou je co do počtu prvků neporovnatelný a tudíž opět neadekvátní.

Přijmeme-li Rapaportovo pojetí vztahů abstrakce a implementace, pak musíme ovšem přiznat správnost i Searlovu pojetí simulace, která není duplikací. Simulace lidského myšlení strojovým myšlením skutečně není duplikací myšlení lidského. Vztah simulace je vztah různých implementací. Jedna tato implementace je schopna simulovat druhou zastoupit ji, ale nestává se tou druhou, tedy ji neduplikuje. Máme však dvě implementace abstraktního myšlení, tedy vlastně máme duplikované to, co je onomu abstraktnímu myšlení vlastní. Došlo tedy k duplikaci realizací abstraktního myšlení, tato realizace však je pouze částí konkrétní implementace. Celá konkrétní implementace však není duplikací jiné implementace, a tedy nemůže platit, že simulace jedné implementace je její duplikací.

Zde prezentované pohledy na ústřední otázku této práce se jeví jako protikladné. Turingova klasická umělá inteligence se na problematiku dívá zvenčí a akcentuje především shodu ve vnějších znacích tedy vstupech a výstupech systému. Searlův přístup oproti tomu zaujímá filosofický pohled zevnitř. Akcentuje tak souvislost rozumění a myšlení. Rapaport pak na poli rozumění přirozenému jazyku sice jakoby drží především vnější pohled na problém, ale jeho výklad syntaktické sémantiky je výkladem o tom, co se děje uvnitř systému. Podobně i snaha Churchlandových o prosazení nových přístupů v umělé inteligenci ukazuje, že mají na zřeteli jak shodu vnější – na úrovni vstupů a výstupů – tak i shodu vnitřní – zde na úrovni struktury neuronové sítě. Přístup Rapaportův i manželů Churchlandových je tak syntézou klasické Turingovské umělé inteligence tak i Searlovy kritiky tohoto přístupu.

Reference

[1] DESCARTES, R.: *Rozprava o metodě.* 3. vyd. Praha, Svoboda 1992. 67 str.

[2] TURING, A. M.: *Computing machinery and intelligence.* Mind, 59, str. 433 – 460. 1950. [cit. 2010-04-16]. Dostupný také z WWW: <http://loebner.net/Prizef/TuringArticle.html>.

[3] SEARLE, J. R.: *Minds, Brains, and Programs.* The Behavioral and Brain Sciences, 3, str. 417 – 424. 1980. [cit. 2010-04-16]. Dostupný také z WWW: <http://web.archive.org/web/20071210043312/http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html>.

[4] JACQUETTE, D.: *Adventures in Chinese Room.* Philosophy and Phenomenological Research, 49-4, str. 605 – 623. 1989. [cit. 2010-07-31]. Dostupný také z WWW: <http://www.jstor.org/stable/2107850>.

[5] SEARLE, J. R.: *Reply to Jacquette.* Philosophy and Phenomenological Research, 49-4, str. 701 – 708. 1989. [cit. 2010-07-31]. Dostupný také z WWW: <http://www.jstor.org/stable/2107856>.

[6] CHURCHLAND, P. M. – CHURCHLAND, P. S.: *Could a Machine Think? Classical AI is unlikely to yield conscious machines; systems that mimic the brain might.* Scientific American, str. 32 – 37. Leden 1990.

[7] RAPAPORT, W. J.: *Machine Understanding and Data Abstraction in Searle's Chinese Room.* Proceedings of the 7th Annual Conference of the Cognitive Science Society (University of California at Irvine) (Hillsdale, NJ: Lawrence Erlbaum Associates): str 341 – 345. 1985. [cit 2010-08-05]. Dostupný také z WWW: <http://www.cse.buffalo.edu/~rapaport/Papers/rapaport85-cogscisoc.pdf>.

[8] RAPAPORT, W. J.: *Syntactic Semantics: Foundations of Computational Natural-Language Understanding.* Aspects of Artificial Intelligence (Dordrecht, The Netherlands: Kluwer Academic Publishers), str. 81 – 131. 1988. [cit. 2010-08-19]. Dostupný také z WWW: <http://www.cse.buffalo.edu/~rapaport/Papers/synsem.pdf>.

[9] RAPAPORT, W. J.: *Understanding Understanding: Syntactic Semantics and Computational Cognition.* Philosophical Perspectives, 9, AI, Connectionism and Philosophical Psychology, str. 49 – 88. 1995. [cit. 2010-08-30]. Dostupný také z WWW: <http://www.cse.buffalo.edu/~rapaport/Papers/rapaport95-uu.pdf>.

[10] RAPAPORT, W. J.: *How to Pass a Turing Test: Syntactic Semantics, Natural-Language Understanding, and First-Person Cognition.* Journal of Logic, Language, and Information, 9-4, AI, Special Issue on Alan Turing and Artificial Intelligence, str. 467 – 490. 2000. [cit. 2010-09-09]. Dostupný také z WWW: <http://www.jstor.org/stable/40180238?origin=JSTOR-pdf>.

E–LOGOS
ELECTRONIC JOURNAL FOR PHILOSOPHY

Ročník/Year: 2011 (vychází průběžně/ published continuously)
Místo vydání/Place of edition: Praha
ISSN 1211-0442

Vydává/Publisher:
Vysoká škola ekonomická v Praze / University of Economics, Prague
nám. W. Churchilla 4
Czech Republic
130 67 Praha 3
IČ: 61384399

Web: http://e-logos.vse.cz

Redakce a technické informace/Editorial staff and technical information:
Miroslav Vacura
vacuram@vse.cz