

**UNIVERZITA PALACKÉHO V OLOMOUCI**

**Filozofická fakulta**

**Katedra filozofie**

**John R. Searle: Reakce na silnou umělou inteligenci a jeho pojetí  
intencionality**

John R. Searle: Reaction to the Strong Artificial Intelligence and His  
Conception of Intentionality

**Diplomová práce**

Vypracovala: Tereza Zbíralová, Filozofie – Anglická filologie

Vedoucí práce: Mgr. Filip Tvrdý

**Olomouc 2010**

Prohlašuji, že jsem tuto magisterskou práci vypracovala samostatně a uvedla úplný seznam citované a použité literatury.

V Olomouci dne 16. 8. 2010

Tereza Zbiralová

Tímto bych chtěla poděkovat Mgr. Filipu Tvrdému za velmi vstřícnou a obětavou metodickou pomoc při vedení této práce a rovněž za poskytnutí nepřehledného množství rad a odkazů k relevantní literatuře.

# OBSAH

<b>Úvod .....</b>	<b>5</b>
<b>1. Searlova kritika umělé inteligence .....</b>	<b>8</b>
1.1 Argument čínského pokoje a námitky .....	10
1.2 Turingův test a imitační hra .....	18
1.3 Je mozek digitalní počítač? .....	24
<b>2. Vědomí a intencionalita .....</b>	<b>30</b>
2.1 Vymezení problému vědomí .....	31
2.2 Definování a vlastnosti vědomí .....	36
2.3 Obecná charakteristika vědomí.....	40
2.4 Jak zkoumat vědomí? .....	42
<b>3. Proč Searle není fenomenolog ani karteziánista .....</b>	<b>46</b>
3.1 Fenomenologická interpretace Searlovy filosofie mysli .....	46
3.2 Kartezianismus.....	50
<b>4. Svoboda vůle a intencionalita .....</b>	<b>55</b>
4.1 Determinismus, kompatibilismus a kauzální vysvětlení jednání .....	55
4.2 Neurobiologická rovina problému svobodné vůle .....	60
<b>Závěr.....</b>	<b>62</b>
<b>Resumé .....</b>	<b>68</b>
<b>Summary .....</b>	<b>70</b>
<b>Anotace .....</b>	<b>72</b>
<b>Annotation .....</b>	<b>73</b>
<b>Seznam citované a použité literatury .....</b>	<b>74</b>

## Úvod

John R. Searle (1932-) se v posledních čtyřiceti letech zabývá filosofickými problémy ve třech hlavních oblastech. První a zároveň základní oblast představuje problematika mysli, další je rovina jazyka a třetí je oblast společnosti. Searlova zkoumání jsou systematická a vykazují jistý vzorec – výchozím bodem je vždy analýza faktů určité úrovně popisu, na jejímž základě tato fakta buď způsobují nebo konstituují další fakta vyšší úrovně nebo na níž jsou samy způsobovány – kterým se Searle řídí v rámci všech tří oblastí, na něž soustřeďuje svou pozornost.<sup>1</sup>

Způsob, jímž Searle postupuje při řešení filosofických problémů vychází z předpokladu, že filosofické problémy je třeba nejprve analyzovat a poté je zrekonstruovat, to vše ještě než se pokusíme o jejich vyřešení. Searle v tomto směru navazuje na metodologii filosofie jazyka a zkoumání podrobuje již samotné otázky. Analytický postup Searlovi umožňuje zjistit, zda určitá otázka a tedy i problém nejsou založeny na chybném předpokladu nebo na systematickém používání vágních pojmů (Searle 2002, p. 2). Cílem Searlovy metody je vyjasnění a osvětlení filosofického problému, což následně umožňuje přesunout zbylý problém do oblasti některého z vědních oborů, například neurobiologie.

Filosofii Johna R. Searla můžeme chápat jako snahu o nalezení odpovědi na otázku po možnosti koexistence koncepce sebe sama jako myslících a racionálně jednajících svobodných bytostí a zcela protichůdné koncepce přírodního světa:

„Jak je možné mít jednotné a teoreticky dostačující pojetí sebe sama a vztahů k ostatním lidem a k přírodnímu světu? Jak můžeme smířit naši na zdravém rozumu založenou představu sebe sama jako vědomých, svobodných bytostí, které mají mysl a provádějí mluvní akty a jsou

---

<sup>1</sup> Srov. „Over the past forty years, Searle has been working on a selection of problems in philosophy at three levels of description: mind (the basic level), language (the middle level), and society (the highest level). In each case Searle can be seen as following a certain pattern: he proposes analyses of facts at one level of description in which they cause, are realized in, or constitute, facts at another higher level.“ Borchert (2006), Vol. 8, p. 705

racionální, se světem, o kterém se domníváme, že se zcela skládá ze surových, nevědomých, nemyslicích, nesmyslných a němých fyzikálních částic nacházející se v silových polích? Jak můžeme, jednoduše řečeno, učinit pojetí sebe sama zcela konzistentním a koherentním s představou světa, kterou jsme získali pomocí přírodních věd, zejména v rámci fyziky, chemie a biologie?“<sup>2</sup>

Tato otázka nebo spíše několik dílčích otázek podle Searla představují nejzásadnější filosofický problém současnosti (Searle 2002, p. 1). Searlův filosofický projekt se nezaměřuje na odvrácení skepticismu v otázce poznání, protože epistemologie<sup>3</sup> a s ní i potřeba fundace vědění se v současnosti již nenachází v centru zájmu filosofie, a to z jednoho prostého důvodu, který zmiňuje Searle: „Víme příliš mnoho“<sup>4</sup>. Podle Searla si můžeme být dnes jisti tím, že objem vědění neustále roste, což následně mění i charakter filosofie v tom smyslu, že epistemologie je odsunuta na okraj filosofického zájmu a s ní i skepticismus.<sup>5</sup>

Fakt, že se objem vědění neustále zvětšuje však neznamená, že dosud uznávané teorie nebudou v budoucnosti zahrnuty do vyšších teorií, které ty dílčí obsáhnou stejně jako v případě newtonovské fyziky a einsteinovské revoluce. Teorie, které dnes uznáváme jako pravdivé primárně na základě velkého množství důkazů, nevylučují možnost, že je jednou budeme nuceni opustit v případě, že se objeví nějaké nové skutečnosti (Searle 2008, pp. 7-8).

---

<sup>2</sup> Srov. „How can we have a unified and theoretically satisfactory account of ourselves and of our relations to other people and to the natural world? How can we reconcile our common-sense conception of ourselves as conscious, free, mindful, speechact performing, rational agents in a world that we believe consists entirely of brute, unconscious, mindless, meaningless, mute physical particles in fields of force? How, in short, can we make our conception of ourselves fully consistent and coherent with the account of the world that we have acquired from the natural sciences, especially physics, chemistry, and biology?“ Searle (2002), p. 1

<sup>3</sup> Když Searle mluví v tomto kontextu o epistemologii, má tím na mysli de facto skepticismus, respektive profesionální epistemologické paradoxy (professional epistemological paradoxes). Viz Searle (2008), p. 6

<sup>4</sup> Srov. „We know too much“. Searle (2008), p. 5

<sup>5</sup> Srov. „The central intellectual fact of the present era is that knowledge grows. It grows daily and cumulatively. We know more than our grandparents did; our children will know more than we do. [...]The modern era in philosophy, begun by Descartes, Bacon, and others in the seventeenth century, was based on a premise which has now become obsolete. The premise was that the very existence of knowledge was in question and that therefore the main task of the philosopher was to cope with the problem of skepticism.Descartes saw his task as providing a secure foundation for knowledge [...]“ Searle (2008), p. 4

Searle se na poli filosofie mysli vyrovnává s materialistickou a dualistickou tradicí, která na jedné straně bránila adekvátnímu zkoumání vědomí a na druhé straně dala ve dvacátém století vznik koncepci umělé inteligence a kognitivismu.

Následující práce se bude primárně zabývat Searlovou filosofií mysli. Představíme Searlovu kritiku koncepce silné umělé inteligence, program znovuoživení vědomí a intencionality pro objektivní zkoumání přírodních věd, kritické odmítnutí tradice karteziánského dualismu a v neposlední řadě také Searlův pohled na problematiku svobodné vůle. Práce si klade za cíl pokusit se zodpovědět, do jaké míry je Searlovo pojetí filosofie mysli přijatelné a udržitelné a případně nastínit, v čem je jeho pohled na problematiku mysli užitečný a inovativní.

V první kapitole představíme Searlovu kritiku koncepce silné umělé inteligence založenou na metafoře a argumentu čínského pokoje a vyvrácení adekvátnosti Turingova testu neboli imitační hry.

Druhá kapitola práce se bude zabývat detailnějším pohledem na Searlův projekt znovuoživení mysli pro objektivní zkoumání přírodních věd. Představíme zde Searlovu definici a vlastnosti vědomí, z nichž mezi nejpodstatnější řadí Searle subjektivitu a intencionalitu.

Třetí kapitola podá vhled do Searlovy kritiky tradice karteziánského dualismu spojené s řešením vztahu mysli a těla. Součástí této kapitoly bude také snaha vymezit se proti dualistickým a fenomenologickým interpretacím Searlovy filosofie mysli.

Poslední, v pořadí čtvrtá, kapitola se pokusí představit Searlův pohled a případný návrh na řešení problému svobodné vůle za pomocí pojmu intencionality a neurobiologie.

## 1. Searlova kritika silné umělé inteligence

První kapitola práce se bude zabývat reakcí Johna R. Searla a jeho vymezením se proti koncepci silné umělé inteligence. Představíme úvod do problematiky silné umělé inteligence, Turingova testu – jeho důsledků a alternativ – a také problematiku počítačích modelů mysli, jež prosazuje kognitivismus.

Searle se domnívá, že současný intelektuální život je do jisté míry ovlivňován třemi tendencemi – dualismem, posedlostí po metodě a odmítnutím pravdy jakožto cíle vědeckého zkoumání. Jejich vliv Searle sleduje a dochází až ke koncepci silné umělé inteligence, v níž nachází všechny tři tendence:

- 1) dualismus vidí v počítačím pojetí mysli, kdy je mysl vyloučena z prostoru ostatních biologických fenoménů a je považována za abstraktní a formální, což vylučuje zájem o biochemické procesy způsobující vědomí a samozřejmě i intencionalitu;
- 2) posedlost metodou je patrná v Turingově testu, který představuje nezvratnou metodu potvrzení přítomnosti mentálních jevů;
- 3) potvrzení teorií Turingovým testem představuje dosažení jediného stupně pravdy, který kdy budeme potřebovat (Searle 2008, p. 77).

Pro Searla je částečně problematický a nevyjasněný již pojem *umělá inteligence* (artificial intelligence), jelikož může znamenat jednak skutečnou inteligenci, která byla uměle vytvořena nebo může označovat něco, co inteligence ve skutečnosti vůbec není – imitaci inteligence. Za problematické považuje také pojmy *inteligence* a *intelligentní*, které je možno užívat doslovně, aniž bychom předpokládali, že zmíněný předmět disponuje vědomím nebo jakoukoli formou myšlení. Searle se zmíněné pojmy snaží vyjasnit na základě rozlišení čtyř významů *umělé inteligence*:



- 1) skutečná mentální inteligence, která je vytvořena uměle – tento význam spojuje se silnou umělou inteligencí, kterou označuje za identickou s komputačním funkcionalismem;
- 2) imitace mentální inteligence – nepředstavuje skutečnou mentální inteligenci;
- 3) nementální inteligence (non-mental) – většinou ji připisujeme knihám a kalkulačkám;
- 4) imitace nementální inteligence.

Searle v důsledku rozlišuje pouze tři typy umělé inteligence, protože třetí a čtvrtý typ umělé inteligence jsou v podstatě identické – pokud totiž dokážeme simulovat nementální inteligenci, pak jsme nementální inteligenci vlastně vytvořili (Searle 2008, pp. 57-58).

Silná umělá inteligence (dále silná UI) představuje podle Searla završení dvoutisícileté tradice materialismu. Silná UI je podivnou směsicí behaviorismu a dualismu – behaviorismu, protože přijímá Turingův test, dualismu, protože odmítá jeden ze Searlových základních předpokladů spočívající v tvrzení, že vědomí a intencionalita představují obyčejné biologické fenomény, jako je například trávení (Searle 2008, p. 74).

Searle rozlišil nejen čtyři základní významy pojmu umělá inteligence, ale postuluje také dva hlavní druhy koncepce umělé inteligence, které se liší především různým přístupem ke zkoumání mysli. Označení „silná umělá inteligence“ (strong artificial intelligence) v protikladu k slabé umělé inteligenci (weak artificial intelligence) Searle formuloval v článku „Minds, Brains, and Programs“ (1990), kde také shrnuje své námitky proti programu silné UI; v knize *Mysl, mozek a věda* (1994)<sup>6</sup> pokračuje v kritice silné UI a zabývá se mimo jiné otázkou, zda mohou počítače myslet, a kognitivismem, který v této kapitole také představíme.

---

<sup>6</sup> Orig. Searle (1984)

Searlovy námitky v „Minds, Brains, and Programs“ se vymezují proti tezi programu silné umělé inteligence, která podle Searla tvrdí, že mysl je k mozku ve stejném vztahu jako program k hardwaru počítače, přičemž správně naprogramovaný počítač představuje mysl – to znamená, že uvedený počítač má porozumění a jiné další kognitivní stavy. Rozdíl mezi silnou UI a slabou UI spočívá v tom, že slabá UI považuje počítač za pouhý nástroj při výzkumu mysli, zatímco v programu silné UI počítač představuje mysl. Searle se vymezuje proti předpokladu silné UI, který spočívá v tvrzení, že správně naprogramovaný počítač doslova má kognitivní stavy a zároveň, že takto naprogramovaný počítač podává vysvětlení procesu lidského myšlení. Proti koncepci slabé UI Searle nic nenamítá (Searle 1990, p. 68).

V reakci na tvrzení, že systém skládající se z počítače a správného programu rozumí a má kognitivní stavy, formuluje Searle argument čínského pokoje<sup>7</sup> (Chinese Room Argument), který se následně pokusím stručně představit, včetně námitek odpůrců argumentu a Searlových reakcí na tyto námitky.

### **1.1 Argument čínského pokoje a námitky**

Searle se v úvodu „Minds, Brains, and Programs“ (Searle 1990, pp. 67-68) soustřeďuje na vyvrácení závěrů Rogera Schanka a jeho kolegů z Yale, jejichž program má simulovat lidskou schopnost porozumění příběhu. Ze Schankova programu pak zastánci silné UI podle Searla vyvozují nepodložené závěry, proti nimž se chce vymezit a přichází proto s myšlenkovým experimentem čínského pokoje. Je nicméně potřeba zdůraznit, že Searle se nevymezuje proti Shankově programu, ale proti tezím, které z Schankova programu vyvozují zastánci silné UI.

Nyní se zaměříme na Searlovu metaforu čínského pokoje. Představme si, že jsme uzamčeni v pokoji, ve kterém se nachází obrovský koš s čínskými znaky. Já – osoba uzamčená v pokoji – neumím čínsky ani trochu, ale mám k dispozici manuál pro manipulaci s čínskými znaky, který je napsán v mé mateřštině. Do pokoje dostanu ještě druhý a třetí koš, ve kterém jsou další čínské znaky a také

---

<sup>7</sup> Argument čínského pokoje také viz Searle (1994), pp. 33-34

instrukce v mém rodném jazyce, které mi umožní spojovat znaky z tohoto koše se znaky z předchozích dvou košů, a zároveň mi tato pravidla předepisují, jakým způsobem mám vracet zpět jisté čínské znaky, a to jako odpověď na ty znaky, které jsem obdržela ve třetím koši. Aniž bych o tom věděla, nazývají lidé, kteří mi dávali koše, první koš „písmo“, druhý koš „příběh“, třetí koš „otázky“ a ty znaky, které jsem dávala nazpět ven, „odpovědi na otázky“ a manuál nazývají „program“. Po jisté době se natolik zlepším já i manuál, že pro vnějšího pozorovatele nejsou mé odpovědi k rozeznání od odpovědí rodilého Číňana, aniž by pozorovatelé tušili, že čínsky vůbec neumím. Je jasné, že odpovědi, které jsem dávala ven z pokoje jsem vytvořila pouze na základě formální manipulace se znaky, aniž bych znala význam těchto znaků – dá se říci, že se chovám jako počítač nebo že vystupuji jako počítačový program (Searle 1990, pp. 68-69).

Argument čínského pokoje Searle zakládá na rozdílu mezi syntaxí a sémantikou<sup>8</sup>, což laicky řečeno znamená, že počítač pracuje s formálními symboly (tedy na bázi syntaxe), a nikoli s jejich významy (semantická rovina) – což je naopak charakteristické pro lidské myšlení.

Pojetí mentálních jevů a programových procesů jako identických troskotá podle Searla právě na ryze formálním neboli syntaktickém definování programů. Searle totiž tvrdí, že na rozdíl od programů mají naše mentální procesy jisté obsahy. Z toho důvodu nebude mít počítačový program nikdy mysl, protože program nemá nic než formální strukturu (Searle 1994, pp. 32-33).

V návaznosti na Schankův program a svůj argument čínského pokoje Searle zmiňuje názory oponentů, kteří namítají ve prospěch porozumění systému, například na základě předpokladu existence různých stupňů porozumění (Searle 1990, p. 71) – Searle různé stupně porozumění nepopírá, ovšem upozorňuje, že existují zároveň případy, kdy je možné jednoznačně určit, zda systém porozumění

---

<sup>8</sup> Searle používá analogii k syntaxi a sémantice viz Searle (1990), p. 83; analogické rozlišení jsem si dovolila použít již předtím, než bychom k němu dospěli až po představení všech šesti námitek zastánců programu silné umělé inteligence.

má nebo nemá. Co se týká přítomnosti nebo nepřítomnosti porozumění v případě naprogramovaného počítače, je Searlův postoj zcela jasný:

„[...] naprogramovaný počítač rozumí asi tolik, co rozumí auto nebo kalkulačka, což znamená, že nerozumí naprosto nic. Porozumění počítače není (totéž, jako mé porozumění němčině) pouze částečné nebo neúplné; rovná se totiž nule“<sup>9</sup> (Searle 1990, p. 72).

Další typické námitky, proti nimž se Searle vymezuje, jsou rovněž obsaženy ve zmíněném článku a je jich celkem šest: systémová námitka; námitka ze strany umělé inteligence; námitka ze strany skutečné simulace mozku rodilého Číňana; námitka ze strany robota; námitka na základě neurčitosti kognice; námitka založená na předpokladu pokroku a redefinování programu silné umělé inteligence. Je třeba podotknout, že námitky uvedené v článku jsou Searlovou rekapitulací nebo shrnutím typických námitek proti adekvátnosti argumentu čínského pokoje a jsou spojovány vždy s některou z amerických univerzit – první námitka s Berkeley, druhá s Yale, třetí s Berkeley a MIT (Massachusetts Institute of Technology), čtvrtá s Berkeley and Stanford, pátá s Yale a poslední opět s Berkeley.

Nyní si dovolím stručně představit výše jmenované námitky vůči Searlově metafoře čínského pokoje a Searlovy reakce na ně.

V pořadí první – systémová námitka – připouští, že osoba uzamčená v pokoji čínštině nerozumí, ovšem zároveň se snaží přisoudit porozumění alespoň systému jako celku, jehož je osoba uzamčená v pokoji pouze součástí. Návrh systémové námitky je však Searlem zavržen, protože podle jeho názoru systém nedisponuje ničím víc, než tím, co má osoba uzavřená v čínském pokoji ve své hlavě – pokud tedy nerozumí jedinec, pak nemůže rozumět ani systém, protože systém je zde pouze částí individua a nikoli opačně, jelikož je to osoba uzamčená v pokoji, kdo vykonává veškeré operace. Alternativní verze systémové námitky zase předpokládá existenci dvou subsystémů, které se nacházejí v hlavě osoby

---

<sup>9</sup> Srov. „[...] the programmed computer understands what the car and the adding machine understand, namely, exactly nothing. The computer understanding is not just (like my understanding of German) partial or incomplete; it is zero.“ Searle (1990), p. 72

v pokoji – jeden ze subsystémů rozumí mateřštině a druhý rozumí čínštině, aniž by oba subsystémy měly mnoho společného. V rámci této verze člověk údajně čínsky rozumí. I v tomto případě však Searle přítomnost porozumění čínštině zcela vylučuje, jelikož zmíněné dva subsystémy si nejsou vůbec podobné – čínský subsystém je součástí mateřského subsystému (který běžně pracuje i s významem znaků, a nikoli pouze s jejich formální reprezentací) a provádí operace s bezvýznamnými znaky na základě pravidel v mateřštině. Motivace k potvrzení, že čínský subsystém má porozumění, je podle Searla přibližně stejná jako ta, co by nás nutila předpokládat, že například i žaludek jistým způsobem rozumí nebo chápe. Searle mezi zpracováním potravy a zpracováním bezvýznamných znaků nevidí žádný rozdíl – ani v jednom z případů nenese systém žádnou informaci (Searle 1990, pp. 72-76).

Program a spolu s ním i zastánci silné umělé inteligence se podle Searla musejí v první řadě naučit rozlišovat mezi mentálními systémy a nementálními systémy, alespoň pokud se program silné UI hodlá soustřeďovat na vysvětlení mentálních jevů. Tím pádem by se pravděpodobně zdržovali pronášení výroků typu, že termostat má také určité přesvědčení<sup>10</sup>. Zkoumání myslí totiž podle Searla „začíná u předpokladu, že lidské bytosti mají přesvědčení, zatímco termostaty, telefony a kalkulačky přesvědčení nemají.“ Pokud má být silná UI vědou o mysli, musí uznat, že termostaty a žaludky nemohou mít takové přesvědčení jako mají lidské bytosti (Searle 1990, p.75).

Druhá námitka – ze strany umělé inteligence – předkládá protipříklad s robotem, kterému do hlavy vložíme počítač, jež bude robota ovládat tak, že se bude zdát, že robot skutečně vnímá, chodí, atd. Počítač v tomto případě bude představovat mozek, který řídí např. kameru určenou ke snímání okolí a také případně ovládá robotovy nohy. Námitka ze strany umělé inteligence již předpokládá, že k porozumění nestačí pouze formální manipulace se symboly, a proto přidává kontakt robota s okolím. Ani tato teorie však pro Searla nepředkládá dostatečné podmínky pro porozumění, protože ani kamera určená k vnímání, ani možnost pohybu po okolí nijak nepřispívají k porozumění – a to primárně kvůli

---

<sup>10</sup> Searle cituje autora výrazu „artificial intelligence“ Johna McCarthyho viz Searle (1994), p. 31 a Searle (1990), p. 72

absenci intencionálních stavů u robota, jelikož robot se pohybuje pouze díky příslušnému programu, kterým je řízen. Ani druhá námitka v důsledku nepředstavuje nic jiného než jen manipulaci s formálními symboly (Searle 1990, pp. 76-77).

Třetí námitka navrhuje simulaci mozku rodilého Číňana – respektive několik programů, které by přesně simulovaly systém neuronů a synapsí v mozku rodilého Číňana přesně v tu dobu, kdy rozumí vyprávění ve své mateřštině a odpovídá na ně. Ovšem ani zde jsme se nevymanili z kruhu čistě formálních operací, jelikož chceme simulovat pouze formální strukturu synapsí a nervových vláken, což není to nejpodstatnější. Intencionální stavy jsou zde to nejdůležitější, s nimi se však v navrhované simulaci nepočítá. Searle vyvrací námitku díky tvrzení, že simulace formální struktury mozku nestačí k porozumění (Searle 1990, pp. 77-78).

V pořadí čtvrtá námitka v podstatě kombinuje předchozí tři protinávry, to znamená, že předkládá robota, který má uvnitř počítač na způsob lidského mozku a program poté disponuje veškerými synapsemi takového mozku. Chování robota je navíc nerozeznatelné od chování člověka. Navrhovaný projekt již není pouze počítačem se vstupy a výstupy, nýbrž jednotným systémem. V tomto případě je Searle poněkud na vážkách, zda systému intencionalitu připsat či nikoli; v případě, že o zmíněném systému nemáme podrobnější informace, připsání intencionality nevyklučuje. Je však důležité poukázat na fakt, že připsání intencionality systému v tomto případě nesouvisí s formálním programem, a tudíž de facto s původní koncepcí programu silné umělé inteligence – původní program se zakládá na tezi, že na základě formálního programu se správnými vstupy a výstupy, lze stroji připsat intencionalitu. V případě navrhovaného řešení čtvrté námitky je však rozhodnutí o připsání intencionality možné pouze za předpokladu, že nemáme tušení, že mozek daného robota stále pracuje na principu formálních operací, v opačném případě bychom systému intencionalitu nepřipsali (Searle 1990, pp. 78-80).

Pátá námitka zpochybňuje připisování kognitivních stavů lidem, respektive se ptá, do jaké míry můžeme vědět, že mají ostatní lidé kognitivní stavy; pokud je připisujeme lidem, pak je musíme připisovat i počítačům. Otázka, kterou

představuje námitka je však irelevantní, jelikož podle Searla jde o to, co připisují lidem, když tvrdím, že mají kognitivní stavy – to co zároveň nemohou být počítačové procesy, protože jejich existence není podmíněna přítomností kognitivních stavů (Searle 1990, p. 80).

Poslední námitka vyhrážející se proti metafoře čínského pokoje apeluje na předpoklad budoucího pokroku v oblasti umělé inteligence a s tím spojeným redefinováním celého programu. Zastánci této námitky se domnívají, že argumenty proti programu silné UI lze aplikovat pouze díky omezenému stavu současné techniky – předpovídají, že v budoucnu bude program silné UI schopen představit stroje, jež budou schopny produkovat kauzální procesy potřebné k získání intencionality, což teprve bude umělá inteligence v pravém slova smyslu. Podobně redefinovaný program UI však není v centru zájmu Searlovy kritiky, jelikož ta se soustřeďuje na vymezení se proti koncepci silné UI, která pokládá mentální procesy za počítačové procesy s formálně definovanými elementy, zatímco budoucí program silné UI ji redefinuje jako cokoli, co kognitivní stavy uměle produkuje a také vysvětluje (Searle 1990, pp. 80-81).

Searlovu kritiku programu silné UI nelze chápat jako úplné odmítnutí možnosti připisování porozumění strojům. Searle se ve své kritice vyslovuje pouze proti připisování porozumění stroji, který by byl řízen na základě počítačových operací s čistě formálně definovanými jednotkami. Vychází přitom z předpokladu, že formální vlastnosti nejsou konstitutivní pro intencionalitu, což znamená, že ani žádný model, který bude založen na čistě formálních principech, nebude dostačovat pro intencionalitu (Searle 1990, p. 82). Searle svou kritikou nepopírá, že by stroje byly schopny myslet, dokonce tvrdí, že my lidé jsme příkladem myslících strojů (Searle 1990, p. 81); „alespoň v jednom významu ‚stroje‘ – totiž ve významu, že stroj je fyzikální systém schopný provádět jisté operace – jsme všichni stroje a přitom můžeme myslet“ (Searle 1994, p. 37).

Je však velký rozdíl mezi duplikací všech příčin, čímž následně získáme i duplikaci účinků, a pouhou simulací příčin. Vědomí, intencionalitu a další lze duplikovat, ale pouze aplikací podobných chemických principů, které fungují u

lidských bytostí. Searle uznává, že i digitální počítač je schopen porozumění a myšlení, ovšem pouhá instalace příslušného programu sama o sobě není dostatečnou podmínkou pro porozumění – proti tomuto předpokladu je namířena Searlova kritika silné UI. Prostřednictvím kritiky silné UI Searle poukazuje na nutnost mít napaměti zásadní rozdíl mezi simulací a duplikací, respektive že simulace se nemůže nikdy rovnat duplikaci. Navíc je třeba si uvědomit rozdílnou povahu mezi vztahem psychických stavů k mozku na jedné straně a vztahem programu k počítači na straně druhé. Na rozdíl od programu jsou psychické stavy produktem mozku (Searle 1990, pp. 83-84).

Počítač je schopen myšlení simulovat, ovšem nikdy nemůže stvořit duplikát (Searle 1994, pp. 38-39). Zásadní rozdíl mezi simulací a duplikováním vysvětluje Searle velmi jasně:

„[...] nikdo nepředpokládá, že při počítačové simulaci bouřky zmokneme nebo že by při počítačové simulaci ohně mohl shořet dům. Proč by si pro všechno na světě někdo, kdo je při smyslech, měl myslet, že počítačová simulace mentálních procesů je skutečně mentálním procesem?“ (Searle 1994, p. 39).

Důvod tohoto přesvědčení Searle vidí v behaviorismu - tezi, že pokud se systém chová, jako by rozuměl čínsky, pak také čínsky rozumí - který se pokusil v rámci argumentu čínského pokoje vyvrátit.

Program silné UI je kvůli přetrvávajícímu behaviorismu a operativismu pro Searla neudržitelný. Silná UI navíc zcela opomíjí zohlednit psychicko-chemické vlastnosti mozku, které při produkci psychických stavů hrají zásadní roli. Největší omyl silné UI spočívá v předpokladu, že pouze stačí mít k dispozici dokonalý program. Pokud by chtěla silná UI trvat na svém, musela by podle Searla uznat oddělitelnost mysli a mozku, tak jak je program oddělitelný od počítače, a to jak empiricky tak konceptuálně (Searle 1990, pp. 85-86).

Problematiku identifikace mysli jako programu Searle komentuje tak, že mozek a dále stroje, které mají stejné kauzální síly jako mozek, jsou schopny myšlení – tímto se však silná UI ze své definice nezabývá, soustřeďuje se pouze



na programy a ty ze své podstaty nejsou dostatečnou podmínkou pro intencionalitu (Searle 1990, pp. 86-87).

Podle Searla zatím nebyla předložena námitka, která by byla schopna vyvrátit základní tezi argumentu čínského pokoje, že čistě formální syntaktické procesy počítačového programu nemohou být samy o sobě dostatečné pro přítomnost mentálních obsahů nebo semántických obsahů, které jsou podle Searla zásadní pro proces myšlení (Searle 2008, p. 67). Chyba, na kterou Searle v rámci svého argumentu poukazuje navíc nesouvisí s technologickým pokrokem, na nějž se odvolává jedna z námitek, protože zmíněná chyba je vlastní koncepci komputace, kterou představil Alonzo Church, Alan Turing, a další. Searle se zmiňuje o nové opozici *člověk – stroj*, která však podle jeho názoru ve skutečnosti žádnou opozici nepředstavuje, jelikož mozek je biologický stroj a není pravda, že by Searle tvrdil, že jej nelze duplikovat (Searle 2008, p. 71). Searle své stanovisko shrnuje do několika bodů, které si v rámci zachování přehlednosti dovolím reprodukovat podle Searlova vzoru v následujících bodech<sup>11</sup>:

- 1) Není pochyb o tom, že stroje mohou myslet, protože mozky lidí a zvířat jsou přesně takové stroje.
- 2) Není pochyb o tom, že by uměle vytvořený stroj principiálně mohl myslet. Není důvod předpokládat, že bychom nemohli vytvořit umělý mozek, je však důležité zdůraznit, že umělý mozek musí duplikovat a nikoliv pouze simulovat kauzální síly původního mozku, stejná podmínka platí i pro vědomí.
- 3) Není možné rozhodnout, zda bude možné vytvořit vědomí za použití jiných, než na uhlíku založených organických materiálech, protože nevíme jaké chemické ustrojí je pro produkci vědomí nutné.
- 4) Komputace je nedostatečná v tom smyslu, že je definována čistě formálně nebo abstraktně na základě použití počítačového algoritmu, a nikoli na základě přesouvání energie (v případě neuronů). Komputace

---

<sup>11</sup> Viz Searle (2008), p. 72

označuje abstraktní matematický proces, který může být využíván stroji, jež se účastní přesunů energie, ale tento přesun již součástí definice komputace není.

Searle dokázal vyvrátit všech šest námitek proti argumentu čínského pokoje na základě tvrzení, že systém – ať už v jakékoliv podobě – nemá porozumění čínštině, protože pracuje s čistě formálně či abstraktně definovanými symboly, a nikoli s významy těchto symbolů.

Závažnou námitku proti jádru Searlova argumentu představuje námitka Davida Chalmerse, kterou zmiňuje Josef Moural v „The Chinese Room Argument“<sup>12</sup>. Námitka argumentuje proti relevanci mentálních stavů části systému provádějící operace – jestli operátor rozumí nebo nerozumí čínštině není podstatné pro otázku, zda příslušný program skutečně produkuje požadované mentální stavy. Námitka tedy upozorňuje, že skutečnou otázkou je, zda má nějaké prorozumění příslušný probíhající program (Smith 2003, p. 229). Moural se domnívá, že pokud Searle předpokládá, že porozumění můžeme nalézt pouze v mysli operátora, pak je celý myšlenkový experiment nepřiměřeně nastaven, jelikož počítač se v tomto ohledu operátorovi rovnat nemůže; protože program nemá vědomí, nemůže mít porozumění (Smith 2003, p. 229).

## 1.2 Turingův stroj a imitační hra

V úvodu „Minds, Brains, and Programs“ Searle vymezuje oblast, na kterou se bude jeho výše představená kritika soustřeďovat – v důsledku na jakoukoli verzi simulace mentálních jevů člověka na základě Turingova stroje.

Nejprve představíme, co Turingův stroj je, kdo je jeho autorem a poté se pokusíme načrtnout problematiku Turingova stroje, jež přímo souvisí s programem silné UI, a tedy i se Searlovou kritikou tohoto programu.

---

<sup>12</sup> In Smith (2003), pp. 214-260

Diskuse o možnosti existence inteligentních strojů má své počátky u komputačních teorií, jejichž systematickou verzi předložil ve svém „Computing Machinery and Intelligence“ (1950) právě Alan Turing.

Turing v tomto článku nahrazuje otázku, zda mohou počítače myslet, novou formulací problému v rámci tzv. imitační hry (Turing 1950, p. 433). Nejprve si představme v čem imitační hra spočívá: imitační hra má tři hráče – hráč A (muž), hráč B (žena) a hráč C (tazatel). Tazatel se nachází v jiné místnosti než zbylí dva hráči a jeho úkolem je rozhodnout, který z hráčů je muž a který je žena. Hráč A má za úkol mást hráče C a přivést jej k nesprávnému rozhodnutí. Úkolem hráče B je pomáhat hráči C. Oba hráči navíc nebudou na tazatelovy otázky odpovídat přímo, ale budou je psát na psacím stroji, tak aby tazatelovo rozhodnutí nebylo usnadněno rozpoznáním tónu hlasu. Otázkou je, co by se stalo, kdybychom hráče A nahradili strojem. Pokud imitační hru takto změníme, rozhodne se tazatel tolikrát špatně, jako kdyby hra probíhala v původním složení? (Turing 1950, pp.433-434)

Turing přesně specifikuje, jaký typ stroje by mohl nahradit hráče A v imitační hře – hráče A by nahradil digitální počítač, který se skládá z paměti (store), procesoru (executive unit) a programu (control). Myšlenka digitálního počítače spočívá podle Turinga v nahrazení lidského operátora provádějícího komputace (human computer). Lidský operátor se řídí stanovenými pravidly, přičemž pravidla jsou pravděpodobně sepsána v nějakém manuálu. Turing digitální počítač označuje také jako *univerzální stroj*, a to z toho důvodu, že dokáže napodobit jakýkoli stroj, což v důsledku znamená, že není potřeba pro každý komputační proces navrhovat zvláštní stroj (Turing 1950, pp. 436-441).

Na základě univerzálního charakteru digitálního počítače Turing vyvozuje vzájemnou ekvivalenci jednotlivých digitálních počítačů a ptá se, zda (za předpokladu, že digitální počítač upravíme tak, aby měl odpovídající paměť, vyšší rychlost prováděných operací a navíc jej vybavíme odpovídajícím programem) bude upravený digitální počítač schopen hrát roli hráče A v imitační hře (Turing 1950, p. 442).

Turing již na začátku článku opouští tradiční otázku „Mohou počítače myslet?“ a nahrazuje ji imitační hrou. K původní otázce se nicméně vrací a v podstatě na ni kladně odpovídá:

„Věřím, že koncem století bude používání slov i všeobecné názory vzdělaných osob natolik jiné, že bude možné hovořit o myslících strojích bez předpokladu, že tomu někdo bude oponovat.“<sup>13</sup>

S původní verzí Turingovy otázky pracuje Searle ve druhé kapitole knihy *Mysl, mozek a věda* (1994), která nese název „Mohou počítače myslet?“. Searle původní otázku taktéž mění, a to na otázku,

„zda digitální počítač, tak jak je definován, může myslet [...] zda příprava či realizace správného programu se správnými vstupy a výstupy je dostatečná, respektive konstitutivní pro naše myšlení.“

Na tuto otázku Searle odpovídá záporně, a to z výše uvedených důvodů založených na rozdílu syntaxe a sémantiky (Searle 1994, p. 38), simulace a duplikace. Podle Searla existují k Turingově testu (imitační hře) tři různé přístupy:

- 1) slabá verze Turingova testu (Weak Turing Test);
- 2) silná verze Turingova testu (Strong Turing Test);
- 3) upravená verze Turingova testu (Modified Strong Turing Test).

Každý přístup interpretuje výsledek Turingova testu jiným způsobem. Slabá verze Turingova testu interpretuje výsledek testu jako úspěšnou či neúspěšnou emulaci lidského chování. Silnou verzi testu Searle charakterizuje jako behaviorismus v rámci umělé inteligence, výsledek testu je v tomto pojetí chápán jako důkaz přítomnosti kognitivních stavů ve stroji, a nikoli jako pouze úspěšná či neúspěšná emulace lidského chování, jak je tomu v prvním přístupu. Třetí přístup zastává stanovisko, že ačkoli existuje rozdíl mezi inteligentním myšlením a vnějším

---

<sup>13</sup> Srov. „I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.“ Turing (1950), p. 442

inteligentním chováním, Turingův test i přesto poskytuje nezvratný důkaz přítomnosti vnitřních myšlenkových procesů (Searle 2008, pp. 53-55).

Pokud se program umělé inteligence zajímá o vytváření strojů, které úspěšně simulují inteligentní lidské chování, pak mu Turingův test poskytuje kritérium, jak posoudit úspěch či neúspěch programu. Jestliže je však test konstruován jako aplikace filosofického nebo logického behaviorismu, vyvozují se z testu závěry, že stroj dokáže simulovat přesně tentýž typ chování, který vykazuje člověk s ohledem na určitý kognitivní fenomén, a tento stroj tedy disponuje i stejným kognitivním jevem v přesně tom samém smyslu jako člověk (Searle 2008, p. 53).

Jak upozorňuje Searle, behavioristé často zapomínají na fakt, že mít mentální stavy je jedna věc, a vyjadřovat tyto stavy chováním je věc druhá. Je totiž možné mít mentální stavy, aniž bychom je museli vyjadřovat chováním, stejně jako je možné vykazovat chování, které neodpovídá mentálnímu stavu. Behavioristé popírají distinkci mezi vnitřními mentálními procesy a chováním. Podle stoupenců behaviorismu existuje pouze inteligentní chování a nic jiného (Searle 2008, p. 55).

Searle pomocí argumentu čínského pokoje vyvrátil základní tezi silné umělé inteligence, zdůrazňuje však potřebu nalézt filosofické vysvětlení, jak se silná UI tohoto omylu dopustila (Searle 2008, p. 59). Pokládá si v tomto duchu zásadní otázku, proč by nuly a jedničky – symboly, s nimiž operuje digitální počítač – měly být dostatečné pro lidské myšlení. Možné spojení vysvětluje funkcí neuronů v mozku, které se buď aktivují nebo neaktivují, což bychom mohli chápat jako určitý typ binárního systému analogického k nulám a jedničkám. Jeden velký rozdíl mezi zmíněnými dvěma systémy spočívá však v tom, že neuronové spoje jsou součástí kauzálního mechanismu, který způsobuje vědomí a myšlení jakožto vlastnost mozku vyšší úrovně. Na rozdíl od nul a jedniček je mozek přístroj, jehož základní procesy se odehrávají na základě přesunu energie. Nuly a jedničky počítačového programu jsou zato čistě abstraktní syntaktické entity a nic podobného jako mozek nezpůsobují (Searle 2008, p. 61).

V důsledku odmítnutí Turingova testu Searle hledá možné alternativy, respektive jiné způsoby, jak prokázat přítomnost mentálních stavů, a zároveň

pochybuje o možnosti, že k tomu bude dostačovat pouze jeden mechanický test. Podle Searla existují různé způsoby, jak zjistit, zda má systém mentální stavy. Searle přitom vychází z předpokladu, že mentální fenomény mají *ontologii první osoby*<sup>14</sup> v tom smyslu, že mohou existovat pouze pokud jsou prožívány člověkem nebo zvířetem – to znamená, že nejlepší způsob, jak zjistit, zda systém vykonává mentální procesy, představuje sám systém. S navrhovaným řešením se však pojí problém, jak vědět o mentálních stavech u jiných lidí a zvířat (Searle 2008, pp. 64-65).

Řešení problému zjištění přítomnosti mentálních stavů u jiných osob a zvířat řeší behaviorismus, který se soustřeďuje pouze na chování; behaviorismus však Searle odmítá s tím, že se nejedná o správné řešení. Své stanovisko Searle vysvětluje na příkladu svého psa, o kterém se domnívá, že má mentální stavy, ale nikoliv na základě toho, jak se pes chová, ale na základě fyziologické podobnosti – má oči, uši, ústa, atd. – a také protože u psa pozoruje jistou podobnost kauzálních mechanismů zprostředkovávající vztah mezi stimulem a odezvou na stimul. Searle se tedy domnívá, že jeho pes má mentální stavy na jiném základě než: stejné chování – stejné mentální stavy (Searle 2008, p. 65), jak proklamuje behaviorismus.

Searle vysvětluje, proč se behaviorismus tak dobře slučuje s formální a abstraktní představou mysli – behavioristický pohled na mentální jevy totiž vylučuje popis mentálna jako něčeho substantivního a biologického – a formální a abstraktní popis mysli nás přivádí až k počítačovým vědám, které produkují formální a abstraktní věci (Searle 2008, p. 82).

Searlovo stanovisko ve věci přítomnosti mentálních stavů je příznačné vědeckým realismem, což v Searlově filosofii znamená, že bychom se měli zajímat o skutečnost (fact) vnitřních mentálních stavů, a nikoli pouze o vnější projevy (appearance). Tvrzení musejí podle Searla odpovídat skutečností ve světě; to, že naše tvrzení projdou Turingovým testem není dostačující (Searle 2008, p. 77). Zajímavý bod Searlova odmítnutí Turingova testu spočívá v rozlišení na observační závislost a observační nezávislost, respektive na Searlově tvrzení, že

---

<sup>14</sup> Vysvětlením ontologie první osoby se práce detailněji zabývá v druhé kapitole.

úspěch či neúspěch (v Turingově testu) je observačně-závislý – to znamená, že nepředstavuje žádnou přírodní kategorii. Podle Searla je chybou domnívat se, že příroda funguje na bázi úspěchu a neúspěchu, zdůrazňuje naopak, že v přírodě se věci pouze dějí. Uvažování o zmíněných kategoriích je závislé na našem vědomí a našich zájmech. Searlově pohledu je možno namítat například přirozeným výběrem, ovšem i úspěch či neúspěch určitého živočišného rodu v boji o přežití je pouze způsob, jakým my lidé interpretujeme to, co se v přírodě pouze děje. Zdánlivá intencionalita boje o přežití je zcela observačně-závislá. Searle uznává užitečnost zmíněných kategorií v rámci evoluční biologie, za chybu ovšem nadále považuje pohled na tyto kategorie jako na něco, co je přírodě vlastní (Searle 2008, pp. 79-80).

Na základě Searlova stanoviska můžeme vyvozovat, že kritikou programu silné umělé inteligence, jejíž součástí je i kritika Turingova testu, Searle sleduje snahu prokázat, že koncepce silné umělé inteligence nepatří mezi přírodní vědy a jejich zkoumání, které se, jak Searle vysvětluje<sup>15</sup>, zaměřuje na přírodu, a ta kategorie úspěchu či neúspěchu nezná. Silná UI dělá vědu tak, že místo aby se soustřeďovala na zkoumání mentální skutečnosti, hovoří o elektronických systémech a programech – v důsledku předpokladu, že není možné provozovat epistemicky objektivní vědu o ontologicky subjektivních jevech. Searle brojí právě proti zmíněnému chybnému předpokladu a upřednostňuje epistemicky objektivní zkoumání mysli, které již nebude ignorovat vědomí a intencionalitu<sup>16</sup>. Searle zamítá behaviorismus i počítačnické modely, protože chování není samo o sobě dostatečné pro vědomí a stejně tak pro něj nejsou dostatečné ani počítačnické modely (Searle 2002, p. 16).

V Searlově pojetí zkoumání mysli je i místo pro program slabé umělé inteligence – počítače zde budou fungovat jako užitečný nástroj pro simulování

---

<sup>15</sup> Srov. Searle (2008), p. 84: „They investigate nature, and nature as such knows nothing of success and failure.“

<sup>16</sup> Searle se tímto problémem zabývá také v *Consciousness and Language*, kde uvádí, že mnozí lidé se chybně domnívají, že podstata vědomí spočívá v nějakém kontrolním mechanismu, dispozicích k chování nebo počítačovém programu – znovu zdůrazňuje, že vědomí nelze vysvětlovat behavioristicky nebo na základě komputace. Turingův test vede k hypotéze, že k tomu aby byl systém při vědomí, je zároveň nutné a dostačující mít správný počítačový program nebo několik programů se správnými vstupy a výstupy. Searle (2002), pp. 15-16

přírodních jevů (Searle 2008, pp. 84-85); program silné umělé inteligence, tak jak je definován dnes, ale podrobuje zásadní kritice.

### 1.3 Je mozek digitální počítač?

Searle se zabývá celkem třemi spolu souvisejícími otázkami: Je mozek digitální počítač? Je mysl počítačový program? Mohou být operace mozku simulovány digitálním počítačem? (Searle 2008, p. 86). Pouze na první otázku se však soustřeďuje ve *Philosophy in a New Century* (2008). Co se týká druhé a třetí otázky, zaujímá Searle jednoznačné stanovisko – na druhou otázku odpovídá záporně na základě argumentu čínského pokoje; odpověď na třetí otázku je jednoznačně kladná, jelikož v podstatě všechno lze simulovat digitálním počítačem.

Otázka „Je mozek digitální počítač?“ se podle Searla rovná otázce „Jsou mozkové procesy počítačové?“ (Searle 2008, p. 87), oběma otázkám je společný jeden směr označovaný jako kognitivismus.

Searle se ve zkoumání kognitivismu vrací k Alonzu Churchovi a Alanu Turingovi a představuje *primal story*<sup>17</sup>, který zakládá na Church-Turingově tezi (což je kombinace Churchovy teze a Turingova teorému) – kdy pro každý algoritmus existuje nějaký Turingův stroj, který je schopen daný algoritmus realizovat. Turingův teorém spočívá v tvrzení, že existuje univerzální Turingův stroj, který je schopen simulovat jakýkoli Turingův stroj, z čehož vyvodíme, že univerzální Turingův stroj je schopen realizovat jakýkoli algoritmus. Zastánci umělé inteligence se následně soustředili na hypotézu, že právě mozek je zmíněným univerzálním Turingovým strojem, která stojí na předpokladu, že alespoň některé z lidských mentálních schopností jsou založené na algoritmu – a tyto schopnosti může provádět univerzální Turingův stroj. Člověk, lidský operátor, provádějící komputace a mechanický počítač tedy používají stejný algoritmus, jediný rozdíl je v tom, že člověk to dělá vědomě a mechanický počítač nevědomě.

---

<sup>17</sup> Searlův termín můžeme překládat jako „prvotní“ nebo také „základní“ příběh, termín nicméně ponechávám v originálním znění.



Můžeme také předpokládat, že existuje spousta dalších mentálních procesů probíhajících nevědomě, které jsou také komputační povahy. Tím pádem k pochopení mozku stačí simulovat zmíněné procesy za pomoci digitálního počítače (Searle 2008, pp. 88-89).

Prostřednictvím *primal story* Searle ukazuje, jak vznikl výzkumný program, který usiluje o objevení programů, které jsou užívány v mozku, a to tím, že programuje počítače tak, aby používaly tytéž programy. Pokud se někdo rozhodne nepohlížet na mozek jako na digitální počítač, vystavuje se odmítnutí vědeckou komunitou, jelikož jedinou další alternativou je dualismus, což je nepřípustné (Searle 2008, pp. 89-90).

Kognitivistický výzkumný program podle Searla zní lákavě hlavně díky tvrzení, že k vysvětlení procesu myšlení nepotřebujeme znát, jak přesně mozek funguje, protože procesy probíhající v mozku představují pouze hardwarové provedení kognitivních programů, přičemž skutečné vysvětlení myšlení je podáváno právě na úrovni programu:

„Teorie předpokládá, že se v mozku manipuluje se spoustou symbolů – nuly a jedničky rychlostí blesku létají mozem a spatřit je nelze ani zrakem, ani tím nejlepším elektronovým mikroskopem – a právě tyto symboly údajně způsobují myšlení.“<sup>18</sup>

Podle Searla však nuly a jedničky nedisponují žádnými kauzálními silami; Searle popírá fyzikální existenci nul, jedniček a domnělé programové úrovně, na níž má probíhat vysvětlení myšlení – program je v podstatě pouze instalované medium, které jinak nemá žádnou ontologii, neexistuje (Searle 2008, p. 98).

Searle se k dualismu nepřiklání a nevnímá jej jako jedinou alternativu k představě mozku jako digitálního počítače, a proto se zaměřuje na definici pojmů, se kterými interpretace mozku jako digitálního počítače pracuje - ptá se, co

---

<sup>18</sup> Srov. „The thesis is that there is a whole lot of symbols being manipulated in the brain, 0s and 1s flashing through the brain at lightning speed and invisible not only to the naked eye but even to the most powerful electron microscope, and it is these which cause cognition.“ Searle (2008), p. 98

přesně je digitální počítač, co přesně je symbol, co přesně je komputační proces a za jakých fyzikálních podmínek realizují (implement) dva systémy tentýž program (Searle 2008, pp. 91). Abychom poznali, zda jde o digitální počítač, nemusíme nutně pátrat po tradičních nulách a jedničkách, ale spíše po něčem, co by mohlo nuly a jedničky zastupovat – z čehož vyplývá, že hardwarová realizace není pro komputační proces podstatná. Pokud bereme vážně předpoklad, že mozek je digitální počítač, pak se nelze podle Searla bránit dojmu, že lze systém, který bude dělat to, co dělá mozek, vyrobit v podstatě z čehokoliv – při zachování komputační ekvivalence obou systémů, respektive na základě, že budou relevantní vlastnosti čistě syntaktické povahy (Searle 2008, pp. 91-92).

V případě, že definujeme komputaci v rámci připisování syntaxe, lze všechno na světě považovat za digitální počítač, protože syntaktické vlastnosti lze připsat čemukoliv – cokoliv můžeme popsat na bázi nul a jedniček. Searle navíc upozorňuje, že syntax není vlastní fyzice, což znamená, že připisování syntaktických kvalit je vždy závislé na pozorovateli, který ten který jev interpretuje jako syntaktický. Komputace, algoritmus a program nepojmenovávají inherentní fyzikální vlastnosti systému – komputační stavy jsou fyzice pouze připisovány, nebyly v ní objeveny (Searle 2008, pp. 93 – 94).

Tak jako argument čínského pokoje ukázal, že sémantika není vlastní syntaxi, zde Searle ukazuje, že syntax není vlastní fyzice – v tomto bodě tedy problematizuje syntaktickou charakteristiku počítače, kterou dříve<sup>19</sup> považoval za neproblémovou. Jak Searle uvádí, není možné určit, zda je něco skutečně digitální počítač, protože je to vždy pozorovatel, který připisuje ryze fyzikálním vlastnostem systému syntaktickou interpretaci – připisovaná psychologická interpretace existuje pouze v očích pozorovatele, jedná se o observačně-závislou interpretaci. Jestliže definujeme komputaci syntakticky, pak na základě fyzikálních vlastností není nic digitální počítač. Kognitivisté se Searlem identifikovaný problém snaží řešit postulováním nějakého konatele - homunkula<sup>20</sup>, který jakoby sedí v mozku a používá jej pro komputaci. Pomocí rekurzivní dekompozice, kdy je do mozku dosazován stále hloupější a hloupější konatel, až není vůbec zapotřebí, se

---

<sup>19</sup> Viz Searle (1990)

<sup>20</sup> Tzv. homunculus fallacy

následně snaží konatele zbavit. Bez homunkula však podle Searla nedokáže kognitivismus udržet syntax v rámci fyziky. Problém s homunkulem odpadá pouze u běžných uživatelských počítačů, kdy je homunkulem sám uživatel. V případě, že však chceme i nadále tvrdit, že je mozek digitální počítač, nelze se ubránit otázce, kdo je potom jeho uživatelem? (Searle 2008, pp. 82, 95-97)

Kognitivismus tvrdí, že mozek funguje stejně jako běžný uživatelský počítač, a tak způsobuje myšlení. Searle ale namítá, že bez přítomnosti homunkula zbývají v mozku i v počítači pouze vzorce, které nemají žádné kauzální síly, z čehož vyvozuje, že kognitivismus není schopen předložit kauzální vysvětlení kognitivních procesů. Vysvětlení mozku na základě použití určitého programu je irelevantní, protože pokud dokážeme něco vysvětlit na neurologické bázi, nebudeme ochotni přijmout vysvětlení v rovině programu. Hlavní problém kognitivismu Searle vidí v neschopnosti počítačové simulace zajistit kauzální vysvětlení simulovaného jevu. Podle Searla nelze fyzikální systém vysvětlovat tím, že určíme formální vzorec, který sdílí se svou počítačovou simulací – protože existence takového vzorce nevysvětluje, jak daný systém ve skutečnosti funguje jako fyzikální systém. Kognitivistický předpoklad, že mozek používá při myšlení algoritmy, se ukázal jako zcela mylný. Jediné, co bychom mohli v mozku nalézt, jsou vzorce událostí, jež lze považovat za formálně podobné programu mechanického počítače. V tomto bodě možné srovnání ovšem končí, jelikož program mechanického počítače nedisponuje žádnými kauzálními silami, tudíž nemůže nic vysvětlovat (Searle 2008, pp. 99-101).

Explanační modely kognitivních věd nejsou schopny splnit podmínku kauzální reality (causal reality constraint). Neadekvátnost explanačních modelů ukazuje Searle na příkladu vysvětlení na základě Marrovy verze explanačního modelu zpracování informací; k prokázání nesmyslnosti explanačního modelu používá opět příklad se psem, který se pokusím ve zkratce ilustrovat:

Searlův pes Ludvík chytá od stěny odražené tenisové míčky; tento explanační model to vysvětluje tak, že pes provádí složité zpracování informací zahrnující přijetí informace na svou 2D rohovku, vyhodnotí ji za pomoci svého vizuálního systému, jehož výsledkem je 3D reprezentace vnějšího světa, kterou

převeďte do motorického aparátu, přičemž se nevědomky řídí pravidly, jak přesně má vyskočit, aby chytil míček atd. Nevědomky totiž provádí spoustu složitých matematických operací. Na uvedeném explanačním modelu není podivné jen to, že předpokládá rozsáhlou nevědomou znalost matematiky u psa Ludvíka, ale hlavně, že zapomíná na fakt, že Ludvík je vědomý racionální činitel, který se o něco snaží. Součástí komputační úrovně tohoto explanačního modelu je i rekurzivní dekompozice; znamená to, že složité úrovně se rozkládají na jednodušší úrovně tak dlouho, dokud není dosaženo nejnižší úrovně, kde se jedná o nuly a jedničky nebo jakékoli jiné binární opozice – takže pes Ludvík nakonec pouze manipuluje s nulami a jedničkami a všechno ostatní je pouhé zdání. Otázka zní: Můžeme připsat nějaký empirický smysl (sense) základním konceptům modelu zpracování informací, jež by učinil z této verze kognitivní vědy legitimní empirickou disciplínu? (Searle 2002, pp. 106 -109).

Searle představenému explanačnímu modelu namítá, že libovolný systém s libovolnou složitostí lze podrobit analýze na základě zpracování informací, např. i kámen padající k zemi. Na tuto námitku odpůrci odpovídají, že je třeba rozlišovat mezi systémy, jimž lze připsat komputaci, a těmi, které ji skutečně provádí (Searle 2002, pp. 110-111). Jak uvidíme dále, ani tuto námitku Searle nepovažuje za adekvátní.

Komputační popisy podle Searla nepředstavují samostatnou kauzální rovinu, která by se odlišovala od fyzikální struktury organismu. Problém komputačních vysvětlení a popisů na základě zpracování informací spočívá v tom, že nemají v důsledku žádnou kauzální realitu. Představa mechanického vysvětlování, které nepracuje s intencionalitou, je pro Searla značně nepochopitelná; zastává totiž názor, že základní forma vysvětlení je založena právě na intencionalitě – tedy na naturalistickém vysvětlení. Stejně tak ale považuje Searle za chybné připsování intencionality<sup>21</sup> věcem, které jsou lidskými výtvary a je zcela jasné, že intencionalitu nemají a nelze je tedy na základě intencionality popisovat. Připsování intencionality termostatu či kalkulačce je vždy připsání observačně-závislé vlastnosti, což znamená, že intencionalitu systému

---

<sup>21</sup> Searlovým rozlišením základních druhů intencionality se zabývám podrobněji v druhé kapitole této práce.

připisujeme jen z toho důvodu, že jsme předmět užívali apod., a nikoli na proto, že by systém sám o sobě disponoval intencionalitou (Searle 2002, p. 127), v takových případech systému připisujeme pouze intencionalitu metaforickou, která de facto intencionalitu ve skutečnosti vůbec nepředstavuje. Bez existence pozorovatele, jež intencionalitu připisuje, systém žádnou vlastní intencionalitu nemá.

Searle se proti komputaci a zpracování informací nevymezuje absolutně, pouze je musíme chápat jako vlastnosti observačně-závislé, respektive jako vlastnosti, které by bez pozorovatele připisujícího tyto vlastnosti neexistovaly. Z fyzikálního hlediska pocházejí počítačnické popisy a také popisy na základě zpracování informací od lidí – pozorovatelů. Oba popisy se však chybně zaměřují za kauzální vysvětlení (Searle 2002, p. 128) – proti tomu je Searlova kritika namířena.

Totéž, co platí pro popisy na základě komputace nebo zpracování informací, platí v rámci Searlovy kritiky také pro vysvětlení na základě pravidel. V případě, kdy se pokoušíme o intencionalistická vysvětlení založená na následování určitého pravidla (příklad chování psa Ludvíka) vlastně nahrazujeme dvě explanační roviny jednou. Searle tento fakt ilustruje na příkladu rádobý darwinovského vysvětlení evoluční výhody jistého tvaru ryby v jistém typu prostředí, které formulujeme tak, že ryba má jistý tvar, proto aby přežila ve vodě – místo toho bychom však měli užívat vysvětlení na dvou rovinách, tedy, že ryba má jistý tvar díky genetické struktuře a dále, že ryby tohoto tvaru mají lepší šanci na přežití než ryby, které takový tvar nemají. Stejně tak bychom měli vysvětlovat i pohyb lidského oka, které se nepohybuje kvůli tomu, že by se řídilo pravidlem vestibulookulárního reflexu (VOR), ale kvůli struktuře vizuálního systému. Vestibulookulární reflex pouze *popisuje*, jak se oko chová (Searle 2002, p.128).

## 2. Vědomí a intencionalita

V druhé kapitole se pokusím představit Searlův pohled na problematiku vědomí a následně i intencionality. Vědomí je totiž předstupněm intencionality, a tudíž spolu oba pojmy úzce souvisí. V úvodu *Rediscovery of the Mind* (1992) Searle představuje program znovuobjevení mysli, který souvisí se snahou jednou pro vždy skoncovat s pokusy „naturalizovat“ neboli redukovat intencionalitu a mysl; tyto pokusy v důsledku odmítají přijmout, že se ve skutečnosti nejedná o nic jiného než o biologické jevy způsobované neuronovými procesy nižší úrovně odehrávající se v mozku. Redukcionistický trend je podle Searla, patrný v rámci různých odnoží materialismu a dualismu, ale i v kognitivních vědách, jejichž předmětem výzkumu nejsou vědomé subjektivní mentální procesy, nýbrž výlučně objektivně pozorovatelné fenomény – z čehož můžeme vědomí vyloučit, jelikož se vyznačuje již zmíněnou subjektivitou. Searle na rozdíl od materialismu a kognitivních věd přichází s novým přístupem a navrhuje zkoumat vědomí v rámci jeho vlastních podmínek a termínů, a snaží se tak překonat jednostrannost vědy spočívající ve snaze zabývat se pouze zkoumáním objektivních jevů, zatímco subjektivní jevy ponechává stranou vědeckého bádání.

Tím, že se Searle rozhodl zkoumat jevy ryze subjektivní, přináší do zkoumání nové hledisko – ontologii první osoby, již nelze opomenout, jelikož představuje hlavní charakteristiku vědomí (Searle 2008, p.11). Searle však nechce zkoumat povahu vědomí bez toho, aby se snažil porozumět také tomu, jak procesy probíhající v mozku vědomí způsobují a uskutečňují (Searle 2002, p. 58).

Searlův postoj k vědomí je unikátní nejen v tom, že se pokouší vědomí zařadit do zbytku univerza a vytvořit tak jakousi jednotnou teorii, ale i z hlediska, které zmiňuje Barry Smith ve svém *John Searle in Focus* (2003, p. 2). Searle podle Smithe představuje odmítnutí velmi dlouhé filosofické tradice, která vnímala již samotnou existenci poznání (knowledge) jako problematickou, jejíž počátky sahají až k Descartovi. Tvrdí, že Searle pohlíží na svět popisovaný vědou s velkou dávkou jistoty. Searle podle Smithe zastává názor, že se suma poznání neustále zvětšuje, což v důsledku umožňuje filosofům formulování velkých sjednocujících teorií, mezi něž podle mého názoru jistě můžeme řadit i Searlovu snahu vrátit

vědomí zpět do zbytku přírody a snažit se jej vysvětlit v rámci teorie, skrze kterou dnes vnímáme svět.

Cílem této kapitoly je také představení Searlových pojmů, a to z toho důvodu, že právě ty nejzásadnější pojmy, které Searle v souvislosti s problematikou vědomí užívá, bývají mnohdy chybně interpretovány.

## 2.1 Vymezení problému vědomí

Hned v úvodu třetí kapitoly *Consciousness and Language* (2002) Searle nahlíží do období uplynulých dvaceti let a konstatuje velmi malý zájem o zkoumání vědomí mezi vědci. Za jeden z možných důvodů odporu ke zkoumání vědomí považuje mnohými přijímaný předpoklad, že vědomí nepředstavuje pro neurovědy vhodný předmět zkoumání. Dalším důvodem mohl být přístup, který sice nevyklučoval vědomí z oblasti zkoumání věd, na druhou stranu však vyjadřoval obavy, že na zkoumání takového jevu ještě nejsme připraveni. Searle ale uzavírá, že i přes nepříznivou historii se dnes již vědci pokoušejí problém vědomí zkoumat. Searlovo zkoumání se ovšem nezaměřuje na publikace z oblasti neurověd, nýbrž se pokouší charakterizovat jinak neurobiologický problém z filosofického úhlu pohledu. Pokládá si zásadní otázku: Jak přesně způsobují procesy probíhající v mozku vědomé stavy a jak jsou zmíněné stavy realizovány v rámci struktury mozku?<sup>22</sup>(Searle 2002, pp. 36-37) Problém podle Searla spočívá ve vysvětlení, jak procesy probíhající v mozku, jež jsou objektivní, biologické, chemické a elektrické, produkují subjektivní stavy pociťování a myšlení (Searle 2002, p. 42).

---

<sup>22</sup> Searlova ústřední otázka, jež zároveň vymezuje jeho další zkoumání, se rozpadá do mnoha dílčích otázek: „What exactly are the neurobiological correlates of conscious states (NCC), and which of those correlates are actually causally responsible for the production of consciousness? What are the principles according to which biological phenomena such as neuron firings can bring about subjective states of sentience or awareness? How do those principles relate to the already well understood principles of biology? Can we explain consciousness with the existing theoretical apparatus or do we need some revolutionary new theoretical concepts to explain it? Is consciousness localized in certain regions of the brain or is it a global phenomenon? If it is confined to certain regions, which ones? Is it correlated with specific anatomical features, such as specific types of neurons, or is it to be explained functionally with a variety of anatomical correlates? What is the right level for explaining consciousness?...“ Searle (2002), pp. 37-38

V úvodu *Freedom and Neurobiology* (2004) Searle vymezuje problém zkoumání vědomí z hlediska tzv. základních faktů o skutečnosti (basic facts), z nichž ta opravdu nejzákladnější jsou pro účel Searlových zkoumání formulována v rámci atomární teorie hmoty a evoluční teorie v biologii (Searle 2004, p. 4). Jeden z problémů obklopující vědomí je také jeho skloubení se zmíněnými základními fakty, jinak řečeno: Jak subjektivní vědomí zapadá do objektivní koncepce světa, kterou představují základní fakta? Searle však hned v zápětí odpovídá, že vědomí zapadá do základních faktů díky tomu, že vědomé stavy jsou zcela způsobovány neuronovými procesy v mozku a přesně tamtéž se také odehrávají. Otázka zasazení vědomí do zbytku univerza v sobě obsahuje i problém skloubení mentalistické představy, kterou máme sami o sobě, s fyzikalistickým – a tedy na první pohled neslučitelným – pojetím univerza (Searle 1994, p. 8), což v důsledku, jak zjistíme dále, souvisí s tradičním problémem řešení vztahu mysli a těla, jehož kořeny sahají až do sedmnáctého století.

Než dospějeme k vlastní definici vědomí, je třeba zmínit několik přístupů k vědomí (Searle 2002, pp. 21-35) proti nimž se Searle vymezuje.

První přístup se domnívá, že nemáme dostatečně pevnou definici vědomí, a tudíž celý pojem vědomí je záhadou. Searle však namítá, že definice vědomí založená na zdravém rozumu – jedná se o vnitřní, kvalitativní, subjektivní stav, který se vyskytuje u lidských bytostí a vyšších živočichů – je dostačující pro určení cíle našeho zkoumání, jelikož teprve na začátku zkoumání není potřeba snažit se zajistit analýzu pojmu. V počáteční fázi podle Searla nemá smysl zabývat se otázkami typu, zda mají slimáci vědomí a podobně. Na druhou stranu však Searle v zápětí považuje za nutné vysvětlit, čemu se při definování vědomí vyvarovat – zejména zaměňování vědomí za sebevědomí anebo za pozornost; argumentuje tím, že sebevědomí představuje pouze jakousi zvláštní formu vědomí, jež je typická pro lidské bytosti a živočichy, a obsahem vědomí mohou být věci i bez toho, abychom na ně upírali pozornost. Navíc například pociťování bolesti podle Searla nemusí předpokládat vědomí sebe sama. Vyjadřuje se také proti námitce založené na nemožnosti pozorovat vědomí ostatních osob, která podle něj nedokazuje, že není možné vědomí vědecky popsat (stejně jako je tomu v případě elektronů, černých děr a „velkého třesku“).



Námitka číslo dvě vychází z předpokladu, že věda má být objektivní, přičemž Searle o vědomí prohlašuje, že je subjektivní, a z toho je vyvozen důsledek, že věda o vědomí jednoduše není možná. Na druhou námitku Searle odpovídá představením dvou odlišných koncepcí protikladných termínů subjektivní – objektivní (také Searle 1992, p. 94). Searle upozorňuje, že někdy mezi sedmnáctým stoletím a současností došlo k převrácení výše zmíněných termínů. V rámci vysvětlení nastalého zmatení poukazuje na nutnost rozlišovat mezi epistemickou rovinou termínů objektivní (např.: Rembrandt se narodil v roce 1606 – tvrzení, jež je možné objektivně ověřit v tom smyslu, že toto tvrzení je objektivně považováno za pravdivé či nepravdivé nezávisle na rozličných názorech nebo předsudcích lidských subjektů) a subjektivní (např.: Rembrandt byl lepší malíř než Rubens – toto tvrzení již nelze objektivně ověřit, jelikož spadá do roviny subjektivních názorů), a rovinou ontologickou. Ontologická subjektivita znamená, že některé entity mají subjektivní způsob existence – např. Searlův pocit bolesti v kříži, který je ontologicky subjektivní v tom smyslu, že je prožíván jím, může existovat pouze tak, že je zakoušen Searlem. Všechny mentální stavy jsou ontologicky subjektivní, protože musejí být zakoušeny lidskou bytostí nebo živočichem, aby mohly existovat. Ostatní věci, jako například vodopády a atomy vodíku jsou pak ontologicky objektivní, jelikož k tomu, aby mohly existovat, nepotřebují být zakoušeny člověkem nebo zvířetem. Searle tímto rozlišením vyvrátil druhou námitku a zároveň jím upozorňuje na fakt, že věda sice hledá objektivitu v epistemickém smyslu, to však nesmí odvracet od existence ontologicky subjektivních entit, které patří do vědeckého zkoumání stejně jako kterékoliv jiné biologické jevy. Podle Searla je možné mít epistemicky objektivní vědění o oblastech, které jsou ontologicky subjektivní.

V pořadí třetí námitka vychází opět z domněle nepřekonatelného rozdílu mezi subjektivním a objektivním a vyjadřuje nemožnost nalezení či formulování teorie, která by objasňovala, jak je možné, že něco subjektivního a kvalitativního je způsobováno něčím objektivním a kvantitativním – neurobiologickými jevy. Námitka zároveň neuznává možnost nalezení jakéhokoli spojení mezi výše zmíněnými stranami. Searle však odpovídá, že dnes je již všeobecně známo, že procesy v mozku způsobují vědomí, to znamená, že spojení zde existuje; problém podle něj spočívá ve formulování teorie, která by vysvětlila, jak je to možné.

Podle Searla nám však k potřebě dalšího zkoumání zmíněného jevu stačí vědomí, že k němu opravdu dochází, a neměli bychom pochybovat, zda se to vůbec děje. Problém vztahu mozku a vědomí vnímá Searle jako další zdánlivě nevyřešitelný problém v historii, což by rozhodně nemělo brzdit jeho další zkoumání; je to jen další záhada, kterou je nutno vyřešit.

Námitka číslo čtyři představuje tezi, jež odděluje vědomí a problém qualí, která jsou odsunuta stranou. Searle samozřejmě s oddělováním problematiky qualí od vědomí nesouhlasí, protože zmíněné dvě oblasti považuje za identické – vědomé stavy představují kvalitativní stavy – a pokud vynecháme qualia, nic nezbyde.

Pátá teze hovoří o kauzální irelevanci vědomí a navíc vědomí přisuzuje pouze epifenomenální charakter. Searle domnělý epifenomenalismus odmítá, jelikož neexistuje důvod, proč považovat vědomí za epifenomenální. Jak uvádí, svět se zakládá na tom, že vyšší úrovně nebo makrostruktury jsou obsaženy v nižších úrovních – mikrostrukturách; to však ještě nepředstavuje důvod, proč je považovat za epifenomenální. Searle si pokládá rovněž otázku, proč o vědomí tvrdíme, že je epifenomenální, když o ostatních makrostrukturách (jejichž vlastnosti se zakládají na mikrostrukturách) se to nepokoušíme tvrdit. Odpověď Searle hledá v pozůstatcích dualismu, které přetrvávají v předpokladu, že vědomí díky své ne-fyzikální povaze nemůže působit fyzikálně – nemůže mít fyzikální účinky. Dualismus pak Searle vidí konkrétně v tom, že zastánci této teze nejsou ochotni spojovat ontologicky subjektivní a epistémicky biologickou charakteristiku vědomí. Searle neodmítá epifenomenalismus jako možnost, upozorňuje pouze, že v současném stavu pro takový pohled jednoduše neexistují dostatečné argumenty, a proto jej v souvislosti s vědomím odmítá.

Další námitka, v pořadí šestá, se ptá po evoluční funkci vědomí a namítá, že ještě nebyla podána žádná uspokojivá odpověď. Searle se opět snaží nalézt jádro námitky, které podle něj spočívá znovu v dualismu, tedy v naší neschopnosti vnímat vědomí jakožto součást fyzikálního světa; naopak jej vnímáme jako něco, co do fyzikálního světa nezapadá, a odtud pramení otázky po jeho evolučním významu. Searle zdůrazňuje, že vědomí umožňuje všechny další hlavní aspekty

lidského života. Skeptický pohled na neužitečnost vědomí v rámci evoluce nemá podle Searla opodstatněný základ.

Sedmá námitka se zabývá kauzálním vysvětlením vědomí. Jakýkoli pokus o postulování kauzálního vysvětlení vědomí, jež by se zakládalo na procesech probíhajících v mozku, vede podle této teze k dualismu; na vědomí tedy pohlíží jako na pouhé vzorce neuronových vláken. Searle na sedmou námitku odpovídá skrze svou definici vědomí jakožto kauzálně emergentní vlastnosti – emergentní vlastnosti chování mikročástic systému, který se z těchto mikročástic skládá. Searlův pohled na vědomí, jež řeší otázku jeho kauzálního vysvětlení, však vede k řešení dalších potíží s redukcionismem, se kterým operuje i následující osmá námitka. Podle osmé teze je věda svou povahou redukcionistická, a proto i vysvětlení vědomí bude redukcionistické a ukáže, že vědomí není nic víc, než jen chování neuronů. V rámci odpovědi na osmou námitku Searle představuje dva základní druhy redukcionismu. Eliminativní redukcionismus se snaží úplně zbavit jevu, který byl původně redukován. Další forma redukcionismu ořezává subjektivní zkušenost od redukovaného jevu; protože však podle Searla představuje subjektivita samotnou podstatu vědomí, nelze ji oddělovat:

„[...] v případě že daný jev představuje subjektivní zkušenost samu, od něj subjektivitu nelze oddělit a zároveň tak vytvořit novou definici pojmu na základě jeho příčin; ne bez toho, abychom přišli o důvod, proč jsme daný pojem vůbec používali.“<sup>23</sup>

Při zkoumání vědomí je podle Searla potřeba na redukcionismus zapomenout a soustředit se na jeho kauzální vysvětlení.

Závěrečná námitka, proti níž se Searle vymezuje, vysvětluje vědomí pomocí komputace a chápe jej jako sled operací zpracovávající informace. Důvodem, proč nelze vědomí analyzovat na pojmy zpracování informací a manipulace se symboly je fakt, že vědomí je vlastní (nezávislé na pozorovateli) biologii nervové soustavy, zatímco zpracování informací a manipulace se symboly jsou observačně-relativní (závislé na pozorovateli).

---

<sup>23</sup> Srov. „[...] where the phenomenon that we are discussing is the subjective experience itself, you cannot carve off the subjective experience and redefine the notion in terms of its causes, without losing the whole point of having the concept in the first place.“ Searle (2002), p. 33

## 2.2 Definování a vlastnosti vědomí

Ve čtvrté kapitole *The Rediscovery of the Mind* zasazuje Searle existenci vědomí do zbytku světonázoru, který je v současnosti interpretován na základě atomární teorie. Svět se skládá z částic a tyto částice tvoří systémy, z nichž některé jsou živé a prošly velmi dlouhým obdobím vývoje. U některých živých systémů se vyvinul právě mozek, který dokáže způsobovat a udržovat vědomí. Původ vědomí je tedy pro Searla jasný – jedná se o biologickou vlastnost některých organismů. Biologickou charakteristikou zde Searle míní, že se jedná o vlastnost srovnatelnou se zažíváním nebo fotosyntézou. Vědomí tak podle Searla představuje přírodní biologický jev, který byl chybně vyčleněn ze zbytku přírody v sedmnáctém století pro potřeby tehdejší vědy, pro niž bylo charakteristické zkoumání měřitelných a objektivních jevů (Searle 1992, pp. 88, 93). Vědomí je také kauzálně emergentní vlastnost chování neuronů, což znamená, že vědomí můžeme redukovat na procesy probíhající v mozku, aniž bychom se zároveň dopouštěli ontologické redukce – „[...] způsob, jímž se ukazuje, že předměty určitého typu se neskládají z ničeho jiného, než z předmětů jiného typu.“<sup>24</sup> (Searle 1992, pp. 113, 115).

Nyní se pokusím ve stručnosti představit nejdůležitější pojmy, které Searle používá při charakteristice a procesu definování vědomí a které jsou s touto problematikou spojené. Patří mezi ně: kvalitativnost, subjektivita, jednota, intencionalita, celostní struktura vědomé zkušenosti, nálada (mood), aspekt obeznámenosti (aspect of familiarity), hraniční podmínky (boundary conditions), pozadí (background) a další.

Kvalitativnost, subjektivitu a jednotu Searle profiluje jako vlastnosti, díky nimž se vědomí odlišuje od ostatních biologických jevů (Searle 2002, p. 39). O první ze zmíněných vlastností Searle říká, že pro každou vědomou zkušenost existuje nějaký pocit, který je spjat právě s touto zkušeností, neboli: „Ke každému vědomému stavu patří i jistý kvalitativně odlišený pocit [...]“<sup>25</sup>. Zmíněnou vlastnost mnozí filosofové charakterizují pojmem *qualia*. Pro Searla však *qualia* představují

---

<sup>24</sup> Srov. „[...] the form in which objects of certain types can be shown to consist in nothing but objects of other types. Searle (1992), p. 113

<sup>25</sup> Srov. „Every conscious state has a certain qualitative feel to it [...]“ Searle (2002), p. 39

totéž co vědomé stavy a odmítá tedy oddělování problematiky qualií od problematiky vědomí. Navíc dodává, že qualia nejsou záležitostí jen percepčního vnímání, ale i myšlení samotného. Od kvalitativnosti se dostáváme k další vlastnosti vědomí – subjektivitě, jelikož subjektivita je v kvalitativnosti implicitně přítomna. Searle tyto dvě vlastnosti dokonce považuje za neoddělitelné, což vysvětluje tím, že aby mohl člověk zažívat kvalitativnost určité zkušenosti, musí existovat nějaký subjekt, který bude uvedenou událost prožívat. To znamená, že subjektivní vědomé stavy mají ontologii první osoby, přičemž ontologií zde míní způsob existence. Ani jednotu, která představuje další zásadní vlastnost vědomí, nechápe Searle odděleně, nýbrž opět jako navazující na subjektivitu a kvalitativnost. Veškerá vědomá zkušenost přichází vždy jakožto součást jednotného vědomého pole, což implikuje zajímavé důsledky pro vědecké zkoumání vědomí, o kterých se Searle později zmiňuje. Jednotu vědomé zkušenosti Searle vysvětluje na analogii s vizuálním vnímáním, kde vizuální systém skládá jednotlivé podněty stimulů do jediného vizuálního vjemu, a stejně tak i celý mozek sjednocuje všemožné stimuly v jedinou a jednotnou vědomou zkušenost (Searle 2002, pp. 40-41).

Subjektivita patří mezi vlastnosti nejzákladnější a zároveň neproblematičtější. Tím, že Searle postuluje subjektivitu jako primární a neredukovatelnou vlastnost vědomí, které se nelze při zkoumání vědomí vyhnout, se vymezuje proti tradici sahající až do sedmnáctého století. Podle Searla odmítáme přijmout subjektivitu jako základní, neredukovatelný přírodní fenomén již od sedmnáctého století, a to na základě domněnky, že věda musí být objektivní, přičemž argumentuje poukazem na rozdílnost mezi epistemickou objektivitou vědeckého zkoumání s ontologickou subjektivitou předmětu zkoumání věd, jako je fyzika nebo chemie. Searle vysvětluje, že z požadavku objektivity aplikovaného na vědu, se vyvodilo, že také zkoumaná skutečnost musí být objektivně existující – nezávisle na zkušenosti lidského individua. Pokud však chceme zkoumat skutečnost objektivně, je třeba epistemicky objektivně zkoumat i jevy ontologicky subjektivní.

Další důležitý pojem představuje intencionalita – „[...] vlastnost mentálních stavů, díky níž jsou zaměřeny na objekty nebo dění ve světě“<sup>26</sup> – která se však běžně užívá jako označení jakéhokoli mentálního jevu, jež má referenční obsah, tedy vlastnosti mnoha našich mentálních stavů, díky které jsou zaměřeny na stav věcí ve světě. Podle Searla je vždy zapotřebí mentálního obsahu, který je zaměřen na utváření reference ke světu. Searle ale upozorňuje, že ne každý vědomý stav je intencionální – například když máme celkově špatnou náladu, přičemž nejsme deprivovaní ničím konkrétním – a ne každá intencionalita je vědomá. Na druhou stranu některé důležité evoluční (vývojové) funkce vědomí intencionální jsou. Mezi vědomím a intencionalitou Searle vidí konceptuální spojení v tom smyslu, že ačkoli je většina našich intencionálních stavů nevědomá, implikuje to podmínku, že aby byl nevědomý intencionální stav skutečně stavem intencionálním, musí být principiálně přístupný vědomí (Searle 2002, pp.12, 44). Searle rozlišuje dva základní druhy intencionality, a to intencionalitu vlastní (intrinsic), kterou mají mentální stavy, a intencionalitu odvozenou, kterou mají například mapy, knihy a tak podobně. Je třeba také zmínit, že vlastní intencionalita představuje observačně-nezávislou vlastnost, zatímco odvozená intencionalita je vlastnost observačně-relativní. Mimo dva základní druhy intencionality pak podle Searla stojí další druh intencionality, kterou připisujeme věcem – metaforická intencionalita neboli intencionalita jakoby (metaphorical – as-if) – která ve skutečnosti intencionalita není. Pokud ji podle Searla nějakému systému připisují, pak pouze říkám, že systém se chová jako by intencionalitu měl, ačkoli ji vlastně nemá. Rozdíl mezi metaforickou a odvozenou intencionalitou spočívá právě v tom, že odvozená intencionalita ve skutečnosti je intencionalita, protože je odvozená od intencionality vlastní, například od mluvčích jazyka (Searle 2002, p. 116). Pokud se má jednat o vlastní intencionalitu, musí mít předmět, jemuž takovou vlastnost připisujeme, skutečně intencionální stav (Searle 1992, p. 78). Tudiž pokud řekneme například o trávniku, že má žízeň, není to totéž, jako když řekneme o nějakém člověku, že má žízeň, protože „vlastní intencionalita představuje jev, jež

---

<sup>26</sup> Srov. „[...] that property of mental states by which they are directed at or about objects and states of affairs in the world.“ Searle (2002), p. 44

je součástí biologické přirozenosti lidí a některých dalších živočichů.<sup>27</sup> Připisování vlastní intencionality se tedy děje na základě jakési konvence, kdy předpokládáme, že lidé a zvířata mají na základě svých biologických charakteristik intencionalitu vlastní, jelikož u nich předpokládáme, opět na základě jisté konvence, přítomnost mentálních stavů, kterými je zmíněná vlastnost podmíněna. Můžeme tedy uznat, že podmínky pro připsání vlastní intencionality jsou jasné a vyloučit tak možnost připsání vlastní intencionality např. stroji. Problémy však začínají vyplouvat na povrch, pokud se zaměříme na případy intencionality *jakoby*, která je zaměňována za intencionalitu *vlastní*, a to v důsledku popírání výše zmíněného rozlišení na intncionalitu *vlastní* a *jakoby*. Podle Searla však rozlišení odmítnout nelze, jelikož se následně dopustíme *reductia ad absurdum*: „Pokud popřete toto rozlišení, pak má najednou všechno v celém univerzu intencionalitu.“<sup>28</sup> Popřením rozlišení docílíme podle Searla pouze toho, že všechno bude najednou mentální, což je absurdní (Searle 1992, p. 81).

Náladu jako jednu z vlastností vědomí Searle vysvětluje tak, že vědomé stavy mají vždy nějakou příchuť nebo tón, pro který nemusíme mít speciální pojmenování. Celostní struktura (*Gestalt structure*) představuje schopnost rozeznat určitý celek pouze na základě minima stimulů: „Schopnost mozku složit degenerované stimuly do určitých strukturovaných forem.“<sup>29</sup> Povědomost (*familiarity*) se dá nejlépe vysvětlit na příkladu toho, že jsme například schopni identifikovat dům, který jsme nikdy předtím neviděli, jako dům na základě kategorií<sup>30</sup> vytvořených na základě zkušenosti. Existují různé stupně obeznámenosti, ale zásadní je, že nepatologické formy vědomí s sebou vždy nesou tento aspekt. Stejně tak vědomé stavy charakterizují i hraniční podmínky spojené se situovaností vědomí, což znamená, že obvykle vím, jaký je den, měsíc, rok; kde

---

<sup>27</sup> Srov. „[...] intrinsic intentionality is a phenomenon that humans and certain other animals have as part of their biological nature.“ Searle (1992), p. 79

<sup>28</sup> Srov. „If you deny the distinction, it turns out that everything in the universe has intentionality.“ Searle (2002), p. 81

<sup>29</sup> Srov. „The disposition of the brain to structure degenerate stimuli into certain structured forms.“ Searle (2002), p. 45

<sup>30</sup> „In order for me to see the objects in front of me as, for example, houses, chairs, people, tables, I have to have a prior possession of the categories of houses, chairs, people, tables. But that means that I will assimilate my experiences into a set of categories which are more or less familiar to me.“ Searle (2002), p. 14

se nacházím a podobně. Vědomé stavy jsou zasazeny do určité situace, i když situační podrobnosti nemusí patřit do obsahu vědomých stavů (Searle 2002, pp. 13-15, 45).

Jeden z hlavních pojmů, který Searle používá při vysvětlování intencionality představuje „pozadí“ (Background). Termín „pozadí“ je souhrným pojmenováním pro několik schopností nebo funkcí, jež přicházejí do hry v souvislosti s intencionálními jevy (význam, porozumnění, interpretace,...), jež jsou na „pozadí“ zcela závislé, aniž by však samo „pozadí“ bylo intencionální. Searle termín opět vysvětluje na několika příkladech v souvislosti s reprezentací, což znamená, že jistý intencionální stav může mít různé podmínky naplnění (conditions of satisfaction) právě v závislosti na různosti daného „pozadí“. Searle odlišuje pozadí od „sítě“ (Network) – Searle vysvětluje, že mít nějaké mínění nebo tužby, znamená mít celou síť dalších mínění a tužeb, protože intencionální stavy v izolaci nejsou schopny rozlišit podmínky naplnění; přičemž zmíněná síť intencionálních stavů se neobejde bez příslušného pozadí (Searle 1992, pp. 175 -177).

### 2.3 Obecná charakteristika vědomí

Vědomí Searle vymezuje jako stav, který trvá po dobu, kdy je člověk vzhůru nebo když se mu zdají sny. Searle rovněž rozlišuje několik stupňů vědomí a také forem<sup>31</sup>, které samozřejmě přicházejí v úvahu pouze za předpokladu, že je člověk při vědomí. Různé formy vědomí umožňují organismům interakci s okolním světem (Searle 1992, pp. 83, 107). Vědomí lze podle Searla připodobnit k povědomí (awareness), které však není ekvivalentní k pojmu vědomí. Vědomí má také vždy nějaký obsah; a dokonce dodává i zautomatizovaným činnostem schopnost rozlišování a přizpůsobivosti, čímž nám vědomí poskytuje evoluční výhodu (Searle 1992, pp. 84, 108-109).

Searle považuje vědomí za hlavní pojem v mentální oblasti, a to mimo jiné proto, že ostatní mentální pojmy je možné chápat jako mentální pouze prostřednictvím jejich vztahu k vědomí (Searle 1992, p. 84).

---

<sup>31</sup> „[...] there are all sorts of forms of consciousness such as vision, hearing, taste, smell, thirst, pains, tickles, itches, and voluntary actions.“ Searle (1992), p.107



V sedmé kapitole *Rediscovery* Searle představuje řešení otázky vztahu vědomí a nevědomých stavů, přičemž představu, kterou máme o nevědomých stavem charakterizuje jako problematickou a nejasnou, a to z toho důvodu, že máme pojem nevědomého stavu pouze jako stavu potenciálně vědomého:

„Náš naivní, předteoretický pojem nevědomého mentálního stavu představuje myšlenka vědomého mentálního stavu minus vědomí.“<sup>32</sup>

Searle chápe koncepci nevědomých stavů jako parazitující na koncepci vědomých mentálních stavů (Searle 1992, pp. 151-153).

Searle také zavádí třetí druh mentálních stavů, který označuje pojmem *nonvědomé stavy* (nonconscious), které vůbec nepovažuje za přímou součást mentálních vlastností, jako příklad uvádí myelinizaci nervových výběžků. Nevědomé stavy tedy představují mentální stavy, které nejsou v ten určitý moment prožívány a nebo byly potlačeny, podmínkou však stále zůstává, že nevědomé stavy musejí být stavy *inherentně* mentální, což v důsledku znamená, že připisování nevědomých mentálních stavů je třeba chápat doslovně (Searle 1992, p. 154-156) – na rozdíl od již zmiňovaného připisování intencionality *jakoby*.

Searle vysvětluje nevědomé stavy na příkladu spícího člověka, aniž by se mu však zdály nějaké sny; mozek takové osoby je tedy podle Searla ve stavu nevědomí, což znamená, že jediné procesy probíhající v jeho mozku jsou sledy neurofyzilogických událostí a procesů. Protože jsou nevědomé mentální stavy potencionálně vědomé (pokud se budeme držet mimo oblast patologických případů), přisuzuje Searle jejich existenci objektivním vlastnostem mozku, které jsou schopny způsobovat subjektivní vědomé myšlenky. V rámci postoje, který Searle sám označuje za naivní mentalismus, neanalyzuje nevědomé mentální jevy jakožto dispozici k chování, ale jako dispozici k vědomým myšlenkám. Searle dále rozlišuje povrchové (shallow) nevědomé stavy – principiálně přístupné vědomí – a hluboké (deep) nevědomé stavy – nepřístupné vědomí, jejichž existenci popírá (Searle 1992, pp.159-162).

---

<sup>32</sup> Srov. „Our naive, pretheoretical notion of an unconscious mental state is the idea of a conscious mental state minus the consciousness.“ Searle (1992), p.152

Vysvětlení toho, jak funguje vědomí, je nezbytné pro vysvětlení většiny vlastností našeho mentálního života, protože všechny do jednoho se tak či onak týkají vědomí. Searle se domnívá, že nemůže porozumět našim nevědomým stavům, dokud nebudeme vědět více o neurobiologii vědomí (Searle 2002, p. 19).

## 2.4 Jak zkoumat vědomí?

Vlastní zkoumání vědomí Searle rozděluje do tří kroků. Prvním krokem ve studiu vědomí je nalezení neurobiologických procesů spojených s vědomím (NCC – neurobiological correlate of consciousness) – podle Searla se nacházíme přesně v této fázi. Druhý krok by měl představovat pokus prokázat, že souvztažnost mezi výše zmíněnými neurobiologickými koreláty vědomí a vědomím je čistě kauzální povahy. Závěrečným krokem by měla být následná formulace teorie, jež by formalizovala tento kauzální vztah (Searle 2002, p. 49).

To, co nás při zkoumání vědomí v první řadě zajímá, je podle Searla odpověď na otázku: Jak je vůbec možné, že jsme při vědomí? Searle se v rámci této otázky snaží vypořádat s blokovou teorií (building block theory), která tvrdí, že vizuální zkušenost představuje pouze jednu část z celého bloku dalších zkušeností, jež jsou konstitutivní pro vědomí, a tak například stačí, když nalezneme neurobiologické koreláty vědomí pouze pro vidění, díky čemuž vysvětlíme vizuální vědomí, přičemž stejně budeme postupovat i u dalších modalit, až nakonec všechny výsledky složíme dohromady, čímž složíme celé vědomé pole (Searle 2002, pp. 50-53).

Searle proti blokové teorii navrhuje teorii jednotného vědomého pole, která již na vizuální zkušenost nepohlíží jako na jednu z mnoha vrstev vědomí, nýbrž jako na modifikaci již existujícího vědomého pole – jako na novou formu, kterou to existující vědomé pole na sebe bere. V návaznosti na teorii jednotného vědomého pole tedy navrhuje zkoumat obecně stav vědomého mozku v protikladu k bezvědomému stavu mozku. Podle Searla bychom měli hledat vědomí jakožto vlastnost mozku, která vychází z činnosti velkých neuronálních uskupení – tedy jakožto vlastnost, kterou nelze vysvětlit na základě aktivity jednotlivých neuronů (Searle 2002, pp. 54-56).

Searle svůj postoj k vědomí charakterizuje jako biologický naturalismus. Odmítá jak dualismus, tak materialismus (Searle 2002, p. 47). Zvláště se Searle vypořádává s behaviorismem a funkcionalismem. Ve své kritice se Searle nesnaží behaviorismus vyvracet, pouze zdůrazňuje fakt, že behaviorismus pohlíží na mentální jevy jako na pouhé vzory chování, a snaží se dobrat toho, co k tak nepřijatelné tezi vůbec vedlo (viz Searle 2002, pp. 204-205).

Se zkoumáním vědomí a intencionality souvisí i problém přisuzování mentálních stavů. Searle však nechce takové podmínky určovat, protože jakmile se po podmínkách pro připsání mentálních stavů začneme ptát, nutně nás to zavede právě k behaviorismu, funkcionalismu či programu silné umělé inteligence.

Při snaze nalézt odpověď na otázku po charakteristice vztahu mezi mozkiem a vědomím se dostáváme až k samotnému jádru problému, které představuje odvěký problém vztahu mysli a těla. V *Consciousness and Language* Searle poukazuje na to, že jeho řešení vztahu mysli a těla je často odmítáno kvůli námitkám zakládajícím se na tvrzení, že nemá ponětí o tom, jak by neurobiologické procesy mohly vědomí způsobovat. Pro Searla to však nepředstavuje ani konceptuální ani logický problém, naopak je to podle něj empirický problém biologický věd – pro Searla nevyvratitelný fakt, že vědomé stavy jsou způsobovány neurobiologickými procesy znamená přenesení problému z oblasti filosofické a metafyzické nemožnosti do sféry teoretického vědeckého zkoumání (Searle 2002, pp. 10-11).

Searle se snaží problém vztahu mysli a těla analyzovat, respektive rozlišuje dva dílčí problémy. První problém spočívá v otázce po obecné povaze vztahů mezi vědomím a dalšími mentálními fenomény na jedné straně a mezi vědomím a mozkiem na straně druhé. Řešení této části problému spočívá ve dvou principech – zaprvé: vědomí a všechny ostatní mentální fenomény jsou způsobovány neurobiologickými procesy mozku na nižší úrovni a zadruhé, že vědomí a ostatní mentální fenomény jsou vlastnosti mozku na vyšší úrovni. Druhý problém, podle Searla daleko obtížnější, představuje popis toho, jak vlastně vědomí v mozku funguje, jehož vyřešení by představovalo nejdůležitější vědecký objev současnosti. Vysvětlení toho, jak přesně vědomí funguje, je totiž nezbytné pro vysvětlení většiny vlastností našeho mentálního života, protože všechny do jedné

se tak či onak vědomí týkají. Searle se domnívá, že nemůžeme porozumět nevědomým stavům, dokud nebudeme vědět více o neurobiologii vědomí (Searle 2002, pp. 18-19).

V *Rediscovery of the Mind* Searle zmiňuje „řešení“ problému dle Thomase Nagela, které se odvolává na nedostatečnost současného pojmového aparátu (Searle 1992, p. 100). Nagel podle Searla však pouze poukázal na nemožnost vymanit se z područí subjektivity, které nám neumožňuje prohlédnout její vztah k materiálnímu základu. Nejsme schopni formulovat vztah nutnosti mezi subjektivitou a neurofyziologickými jevy právě proto, že se v subjektivitě nacházíme (Searle 1992, p. 102) – jednoduše od ní nejsme schopni odhlédnout.

Problém vztahu mysli a těla, potažmo problém vztahu mozku a vědomí nachází v rámci Searleovy filosofie jednoduché řešení reprezentované biologickým naturalismem (Searle 1992, pp. 1-2). Biologický naturalismus však někteří myslitelé vnímají chybně jako formu kartezianismu, Searle zmiňuje materialisty a dualisty (Searle 1992, p. 2), kteří se snaží vědomí spolu s intencionalitou naturalizovat neboli redukovat. Searlův přístup však představuje pouze snahu vrátit vědomí zpět mezi předměty zkoumání přírodních věd (Searle 1992, p. 85) a navrhuje biologickou definici vědomí, v rámci níž považuje vědomí za biologický jev, který je výsledkem biologické evoluce. Vědomí je podle této definice biologická vlastnost, kterou disponují mozky lidí a také mozky některých zvířat, způsobovaná neurofyziologickými procesy. Vědomí představuje biologickou vlastnost stejně jako například fotosyntéza nebo zažívání (Searle 1992, p. 90). Searle dále přirovnává vztah mozku a vědomí ke vztahu vody k její tekutosti, a navíc vylučuje introspekci jako možný přístup ke zkoumání vědomí, a to z toho hlediska, že introspektivní metoda předpokládá rozlišení mezi vlastní introspekci a předmětem, jež introspekci podrobujeme, což v případě vědomí provést nelze. Vědomí má podle Searla také prostorové charakteristiky, a to proto, že se nachází v mozku, ačkoli nejsme v současnosti schopni určit kde přesně, přičemž Searle předpokládá, že je pravděpodobně rozloženo ve velmi značné části mozku (Searle 1992, p. 105). Searle rozlišuje několik forem vědomí: zrak, sluch, chuť, žízeň, bolest,... a volní činnosti; vědomí organismu umožňuje projevit se ve světě, produkovat ve světě jisté účinky. Vědomí podle Searlea představuje velkou

výhodu spočívající v tom, že nám poskytuje možnost daleko většího rozlišování, než by nám mohl případně poskytnout nějaký nevědomý mechanismus (Searle 1992, p. 107).

Otázka po povaze a vlastnostech vědomí se v této souvislosti přesouvá na pole problému připisování inteligence či inteligentního chování, na které se domníváme mít lepší nebo alespoň objektivnější kritéria oproti připisování vědomí, podle Searla však brzy naráží i tento pohled na překážky (Searle 1992, p. 56). Inteligence existuje podle Searla dvojí: *intelligentní chování* lze připisovat termostatům nebo kalkulačkám, *inteligenci* lze přisuzovat lidem. Searle však upozorňuje, že inteligentní chování, tedy ten „typ“ inteligence, který lze připsat strojům, nemá absolutně žádnou psychologickou relevanci (Searle 1992, pp. 56-57).

V rámci myšlenkových experimentů, které najdeme ve třetí kapitole *Rediscovery of the Mind* ukazuje Searle, že mozek je jednak schopen produkovat vědomí a na druhé straně také chování nebo jednání, přičemž takový systém, v našem případě mozek, je schopen vědomí bez chování (respektive vnějších projevů) a nebo se může projevat, aniž by musel mít vědomí. Z uvedeného Searle tedy vyvozuje, že chování nepředstavuje nutnou podmínku pro přítomnost mentálních stavů. Searle tento fakt vyvodil z myšlenkových experimentů a nazval jej „principem vzájemné nezávislosti vědomí a chování“ (Searle 1992, p. 69).

### 3. Proč Searle není fenomenolog ani karteziánista

V následující kapitole se pokusím vyvrátit dvě chybné interpretace Searlovy filosofie mysli. Jednak se pokusím odpovědět na otázku: Proč Searle není fenomenolog? Prvním tématem kapitoly bude fenomenologická interpretace<sup>33</sup> Searlovy filosofie mysli, jež se odvolává na Searlovo tematizování pojmu intencionalita. Druhým hlavním tématem bude Searlův domněle nevyhnutelný příklon ke karteziánskému dualismu, který se rovněž pokusím vyvrátit na základě představení Searlova vlastního pohledu na dualismus a také materialismus – směry, kterým se dostalo v Searlově filosofii společného jmenovatele, jímž je projekt silné umělé inteligence.

#### 3.1 Fenomenologická interpretace Searlovy filosofie mysli

Otázka zní: Můžeme na Searla pohlížet jako na fenomenologa? Nejprve je nutné zmínit důvody, které by nás k takové interpretaci mohly vést. Prvním z důvodů by mohla být Searlova zaměřenost na intencionalitu, jež nás dovede až k Edmundu Husserlovi, tedy k fenomenologii. Jak uvádí Peregrin (2005, p. 249), Husserl stejně jako Searle, kladl důraz na intencionalitu mysli, čímž se oba myslitelé značně odlišují od Fregeho, který je jednoznačně řazen k analytické větvi filosofie.

Na první pohled by se mohlo zdát, že na základě používání a zkoumání pojmu intencionalita, se Searle vydává po stopách Husserla. Pokud však na chvíli odložím předčasné fenomenologické interpretace, dostávám se k analýze pojmu intencionalita, který se u obou myslitelů odlišuje, ačkoli lze u obou nalézt jisté podobnosti, které však, jak Searle sám uvádí, vyplývají jednoduše z faktu, že se oba zabývají tímtež jevem – intencionalitou (Searle 2008, p. 112).

---

<sup>33</sup> Podnětem a inspirací k napsání reakce na fenomenologickou interpretaci Searlovy filosofie mysli mi byla seminární práce kolegy Otakara Horáka s názvem „Moje přesvědčení o Searlovi intencionalitě“, v níž je Searlova filosofie interpretována jako bližší fenomenologii než analytické filosofii, a to hlavně kvůli tematizování pojmu intencionalita. Seminární práce se mimo Searlových článků odvolává na *Kapitoly z analytické filozofie* J. Peregrina.

Je třeba stručně nastínit pojem intencionality z hlediska dějin filosofie a představit tak, alespoň ve zkratce, historii pojmu, a následně se pokusit osvětlit podobnost Searlovy koncepce intencionality s fenomenologií, včetně případného vyvození důsledků.

Pojem *intencionalita* poprvé použil Jeremy Bentham v rámci rozlišení mezi záměrným (intentional) jednáním a jednáním, které záměrné není. Dalším, kdo se intencionalitou zabýval, byl Edmund Husserl, a to na základě myšlenek, které představil Franz Brentano ve svém *Psychologie vom empirischen Standpunkt* (1874). Brentano zde jednak představil definici intencionality a zároveň uvedl, že intencionalita je znakem mentálního. Intencionalita v Brentanově pojetí představuje vlastnost mnoha mentálních stavů, jako jsou například přesvědčení, soudy, tužby a percepce. Brentanova definice intencionality identifikuje zmíněný pojem s tím, co mentálnímu stavu nebo jednání umožňuje být zaměřeno na tzv. intencionalní objekt, kterému připsal vlastnost zvanou intencionalní neexistence nebo immanentní objektivita. Intencionalní neexistence přitom pravděpodobně poukazuje na fakt, že intencionalní objekty se nemusejí nacházet v prostoru a čase; dostačuje, že existují v mysli. Brentano přisuzuje intencionalitu jakožto vlastnost výlučně mentálním stavům, což přineslo později otázku, zda skutečně pouze mentální stavy mohou být intencionalní, a také snahu rozlišovat intencionalitu vlastní a intencionalitu odvozenou (*Encyclopedia of Philosophy*, Vol. 4, pp.704, 708-709).

Edmund Husserl, zakladatel fenomenologie<sup>34</sup>, navazuje na Brentanův pojem intencionality. Domnívám se však, že v případě Searla o návaznosti na Husserlovu koncepci intencionality hovořit nelze<sup>35</sup>, a to jednak proto, že metoda obou myslitelů je diametrálně odlišná<sup>36</sup> a také proto, že sám Searle odmítá, že by se Husserlem nějak inspiroval:

---

<sup>34</sup> „Phenomenology in the narrow sense begins in 1900 with the publication of the first volume of Husserl's *Logical Investigations* [...]” Smith (2003), p. 262

<sup>35</sup> Ačkoli například v *Encyclopedia of Philosophy* tvrdí pravý opak: „[...]John Searle has found the Husserlian intentionality thesis useful for his own work [...]” Borchert (2006), Vol.7, p. 299

<sup>36</sup> „[...] we can say that the method of Husserlian phenomenology is to describe the noema by giving its essential structure. The method of logical analysis is to state conditions – truth conditions,

„[...] přečetl jsem první díl *Logických zkoumání* a upřímně řečeno byl jsem velmi zklamaný. Nabyl jsem dojmu, že spis Frege v ničem nepřekonává a že není příliš dobře napsáný, naopak byl spíše nejasný a zmatený. Tudíž jsem zanechal snahy dozvědět se něco o intencionalitě od předchozích myslitelů a pustil se do vlastní práce. [...] od Husserla jsem se nanaučil nic, doslova vůbec nic [...]“<sup>37</sup>

Searle fenomenologii explicitně uznává, pouze pokud bychom ji definovali tak, že se zabývá zkoumáním vědomí:

„Hned na začátku bych chtěl zdůraznit, že *pokud bude fenomenologie definována jako zkoumání struktury vědomí, pak proti ní vcelku nic nenamítám.*“<sup>38</sup>

Searle se také konkrétně vymezuje proti Husserlově metodě transcendentální redukce a zdůrazňuje, že jeho metodou je logická analýza vycházející z formulování podmínek splnění určitého intencionálního stavu; intencionální stavy samotné zároveň představují podmínky svého splnění. Naproti tomu o Husserlovi říká, že se naopak snaží dobrat výsledku zkoumáním struktury svého vědomí. Searle odmítá introspekci a nahrazuje ji naturalistickým pojetím intencionality, což znamená, že intencionalitu chápe jako biologickou vlastnost, stejně jako zažívání nebo fotosyntézu. Searle odmítá introspekci s tím, že jeho cílem je vymanit intencionalitu z prostředí tradice kontinentální filosofie a otevřít ji pro zkoumání analytické filosofie (Searle 2008, p. 114).

Na rozdílnost Husserlovy metody a jeho neochotu zabývat se podmínkami splnění upozorňuje také Barry Smith:

---

performance conditions, conditions of constitution, etc.“ Searle (2008), p.115; „[...]Searle's philosophy of mind differs on most central points from that of Husserl. Smith (2003), p. 261

<sup>37</sup> Srov. „[...]I read the *First Logical Investigation*, and, frankly, I was very disappointed. It seemed to me that it was in no way an advance on Frege and was, in fact, rather badly written, unclear, and confused. So I abandoned the effort to try to learn something about intentionality from previous writers and just went to work on my own. [...]I learned nothing from Husserl, literally, nothing [...]“ Searle (2008), pp. 111-112

<sup>38</sup> Srov. „I want to emphasize at the start that *if phenomenology is defined as the examination of the structure of consciousness, I have no objections whatever to phenomenology.*“ Searle (2008), p. 108



„Husserl zaujímal k analýze – ve významu udávání definic, nutných a dostatečných podmínek analyzovaného – často skeptický postoj.“<sup>39</sup>

Fenomenologická metoda introspekce nebo transcendentální redukce není podle Searla schopna postihnout vždy totéž co jeho logická analýza. Searle totiž tvrdí, že:

„[...] některé z velmi podstatných logických vlastností intencionality se nacházejí mimo možnosti fenomenologie, protože nemají žádnou okamžitou fenomenologickou realitu.“<sup>40</sup>

Zásadní rozdíl mezi Searlovou koncepcí intencionálního obsahu a Husserlovým noematem spočívá podle Searla v tom, že noema může obsahovat pouze věci, které jsou z fenomenologického hlediska skutečné. Pokud něco není fenomenologicky reálné, pak to není reálné vůbec, a naopak, pokud je něco fenomenologicky reálné, pak je to reálně dostatečně. Searle v této souvislosti hovoří o tzv. fenomenologické iluzi<sup>41</sup>. Jednou z vlastností, které není fenomenologie podle Searla schopna postihnout je například kauzální sebereferece vnímání a jednání, která vzejde až z logické analýzy a nikoli z fenomenologické analýzy, jelikož nepředstavuje fenomenologicky skutečnou vlastnost v tom smyslu, že není vždy přítomna ve vědomí, když je prožívána nějaká zkušenost (Searle 2008, p. 120).

Pro Searla je fenomenologie přijatelná pouze jako odrazový bod analýzy intencionality, ale v tomto bodě také přijatelnost fenomenologie končí. Searle fenomenologii nezavrhuje úplně, jelikož poukazuje, že bychom vždy měli používat odpovídající metodu, ať už analytickou nebo fenomenologickou; obě metody Searle vnímá jako alternativy, a tudíž ani jednu z nich nevylučuje (Searle 2008, p. 115-116). Searle se také zamýšlí nad rolí fenomenologie v dnešní filosofii

---

<sup>39</sup> Srov. „Husserl was, however, often sceptical about analysis where this means providing definitions, the necessary and sufficient conditions of an analysandum“ Smith (2003), p. 262

<sup>40</sup> Srov. „[...]some of the most important logical features of intentionality are beyond the reach of phenomenology because they have no immediate phenomenological reality.“

<sup>41</sup>Fenomenologická iluze pramení z odmítnutí základních faktů skutečnosti jako opravdu základních. Odmítají nadřazenost těchto základních faktů, protože jinak by museli například zařadit Dasein do zbytku biologického světa a vysvětlovat jaké jsou jeho evoluční výhody a zda jej mají i jiní živočichové apod. Searle (2008), p.126; detailněji Searle (2008), pp.116 – 126.

a dochází k závěru, že fenomenologie hraje velmi důležitou roli poté, co uznáme existenci základních faktů skutečnosti a také závislost mysli na těchto základních faktech, poté začneme zkoumat běžnou zkušenost prostředky fenomenologie. Je však znovu třeba zdůraznit, že sama fenomenologie dostačuje pouze ze začátku a další zkoumání se musí zabývat analýzou logických struktur. Searle fenomenologii nezavrhuje, jeho kritika není zaměřena na fenomenologii jako celek, ale pouze na několik chybných částí (Searle 2008, p.135).

I přes viditelný částečně pozitivní přístup k fenomenologické metodologii však nelze Searla interpretovat primárně prizmatem fenomenologie, jelikož Searle se pokouší založit a provozovat naturalizovanou filosofii, kterou charakterizuje používání několika adekvátních metod a přijetí základních faktů skutečnosti.

### 3.2 Kartezianismus

Searlova filosofie mysli je vnímána prizmatem fenomenologie; kvůli kritice materialismu a odporu k redukci mentálních jevů se však nabízí také interpretace, jež hovoří o Searlově nevyhnutelném příklonu k dualismu<sup>42</sup>. Jak Searle sám uvádí<sup>43</sup> odmítnutí materialismu pro mnohé nutně znamená příklon k dualismu, protože žádná třetí volba neexistuje. V případě Searlovy filosofie mysli však o dualismu nemůže být řeč primárně proto, že Searle sám dualismus mysli a těla explicitně odmítá. Abychom se však mohli vymezit proti možným dualistickým interpretacím, bude nutno vysvětlit, v čem přesně spočívá Searlovo odmítnutí dualismu.

Není pochyb o tom, že Searle dualismus v jakékoli podobě odmítá. Searle popírá tradiční rozlišení na mentální a fyzikální a přichází z vlastním pojetím, které tradiční kategorie překonává a charakterizuje je stanovisko, že mentalismus je ve skutečnosti pouze speciálním typem fyzikalismu (Searle 1992, Searle 2002) – pro

---

<sup>42</sup> „[...] he has to either (a) give up his claim about the irreducibility of consciousness (and this would allow the analogy with liquidity and digestion to hold) or (b) accept a dualist conception of mind, where consciousness is an ontologically distinct feature of the world.” Smith (2003), p. 145

<sup>43</sup> Podle Searla hraje v otázce statutu mentálních jevů jistou roli nevyslovený předpoklad, že jedinou alternativou k vědeckým teoriím, které symbolizuje materialismus, jsou nevědecké teorie, jež reprezentuje právě kartezánský dualismus. Viz Searle (1992), pp. 3-4

Searla existuje jen jeden druh věcí a tím jsou věci fyzikální. Zde Searle naráží na problémy související se zastaralým a pro moderní teorii mysli nevhodným dualistickým slovníkem, který nás nutí si myslet, že věci jsou dvojího druhu – buď fyzikální nebo mentální – což je potřeba překonat. Odpoutání se od dualistického slovníku nám následně umožní vyřešení problému vztahu mysli a těla, respektive popření existence takového problému, jelikož jsme následně schopni definovat mentální stavy – neboli emergentní vlastnosti systému neuronů – jako určitý druh fyzikální vlastnosti, ovšem na vyšší úrovni.<sup>44</sup>

Searlova filosofie mysli tedy překonává nejen dualismus, ale i tradiční dualistický slovník, čímž se náhle ztratí propast mezi mentálním a fyzikálním a s ní i problém vymezení vztahu mezi myslí a tělem, potažmo vědomím a mozkem, který představuje ohnisko Searlovy filosofie mysli.

Searle se snaží v *Rediscovery of the Mind* (1992) zasadit mentální jevy, respektive vědomí, zpět do celku světa a zrušit tak velmi dlouhou tradici, která se objektivnímu zkoumání vědomí a mentálních stavů vyhýbala nebo je považovala za nemožné a neadekvátní. Domnívám se, že i díky tomuto ohledu nelze hovořit u Searla o již zmiňovaném příklonu k fenomenologii, a to hlavně proto, že se jeho filosofie mysli neodděluje od celku světa. Searle hovoří o tzv. základních faktech<sup>45</sup> (basic facts) neboli základní realitě (basic reality) z oblasti částicové fyziky, evoluční biologie a neurobiologie. Filosofie má dle Searla za úkol řešit otázku: Jaké povahy jsou vztahy mezi lidskou realitou a základní realitou. Otázka však v žádném případě neimplikuje existenci dvou oddělených světů a tedy druh karteziánského dualismu: „Lidský svět je součástí jednoho světa, a ne nějakého jiného.“<sup>46</sup> (Searle 2008, p. 109).

---

<sup>44</sup> Searle, John R., „Intencionalita a jej miesto v prírode“, In Gáliková, Silvia – Gál, Egon (eds.), *Antológia filozofie mysle*. Bratislava 2003, pp. 494-495.

<sup>45</sup> „We know that the basic structure of the entire universe consists in entities that we find it convenient (if not entirely accurate) to call “particles,” and these exist in fields of force and are typically organized into systems. We know furthermore that we and all living systems have evolved over a period of somewhere between three and five billion years by processes of Darwinian natural selection.[...]we have to add the neurobiological basis of all human and animal mental life. All of our consciousness, intentionality, and all the rest of our mental life are caused by neurobiological processes and realized in neurobiological systems.“ Searle (2008), p. 109

<sup>46</sup> Srov. „The human world is part of one world, not something different.“ Searle (2008), p. 109

Searle explicitně odmítá jak dualismus, tak i materialismus a jakoukoli formu redukcionismu; své vlastní stanovisko označuje za biologický naturalismus (Searle 2002, p. 47), čímž postuluje dosud neredukcionistickou alternativu k převládajícím redukcionistickým přístupům ve filosofii mysli. Směry, proti nimž se Searle vymezuje, jsou de facto všechny, které popírají existenci výše zmíněných základních faktů a nebo hovoří o dvou (dualismus), případně třech (trialismus<sup>47</sup>) světech<sup>48</sup>. Primárním důvodem pro odmítnutí těchto směrů je pro Searla snaha rozvíjet naturalistickou filosofii (Searle 2004, pp. 20-26), která mimo jiné vychází z předpokladu, že svět je jeden a že jeho součástí je také vědomí, které je možno objektivně zkoumat.

V druhé kapitole *Rediscovery of the Mind* Searle představuje přehled jednotlivých směrů současného materialismu, kde u každého ze směrů uvádí jednak protiargumenty ze strany zdravého rozumu (common sense) a také námitky technického ražení. Dovolím si zde zmíněné směry v rámci přehlednosti alespoň vyjmenovat: zmiňuje zde logický behaviorismus, typovou teorii identity (type identity theory), partikulární teorii identity (token identity theory), funkcionalismus černé skříňky (black box functionalism), koncepci silné umělé inteligence neboli funkcionalismus Turingova stroje, eliminativní materialismus a naturalizaci intencionality (Searle 1992, p. 54).

Searle se však nezabývá pouze jejich vyvrácením, ale pokouší se také vystopovat jejich původ, který nachází v sedmnáctém století v již zmíněném dualismu. Searle staví kritiku současných materialistických směrů na jejich historickém původci, který v těchto novodobých směrech podle Searla přežívá především v používání tradičního dualistického slovníku. Současný materialismus dle Searla převzal původní debatu tak, jak ji formuloval René Descartes – tedy hlavně adekvátnost rozlišování na materiální – imateriální, tělo a mysl, fyzikální a mentální – což znamená, že pokud dnes přijmeme existenci vědomí, přijímáme s ním i jistou formu dualismu (Searle 1992, p. 54).

---

<sup>47</sup> Termín „trialismus“ (trialism) přebírá Searle od Johna Ecclese. Viz Searle (2004), pp. 22-23

<sup>48</sup> Searle ve stručnosti nastiňuje také svůj pohled na problematiku univerzálií, která podle něj nevyžadují postulování třetího světa, jelikož když používáme univerzálie, pouze se tím vyjadřujeme o jediném existujícím světě jiným způsobem. Searle (2004), p. 25

Descartes prostřednictvím *res extensa* a *res cogitans* rozdělil skutečnost na dvě části, z nichž první byla doménou vědy, zatímco druhá doménou církve a náboženství. Descartovo tehdy užitečné rozdělení, které Searle interpretuje jako snahu vyřešit konflikt mezi vědou a náboženstvím, však vedlo k vyloučení vnitřních kvalitativních subjektivních zkušeností, jež utvářejí vědomí, protože věda se může zabývat pouze objektivně pozorovatelnými jevy. Rozdělení skutečnosti bylo užitečné v sedmnáctém století, dnes však přetrvávající zbytky karteziánské tradice brání vědomí stát se jedním z adekvátních předmětů zkoumání přírodních věd (Searle 2002, p. 48).

Spolu se Searlem si můžeme dnes položit otázku, proč nejsou ostatní filosofové schopni opustit tři sta let starý dualistický slovník a přijmout tak bez obav z dualismu existenci vědomí. Searle se ptá, proč se dnešní materialisté bojí vědomí a následně si odpovídá, že možný důvod spočívá v subjektivní kvalitě vědomí, která je pro materialisty přímo děsivá, jelikož její přijetí by pravděpodobně znamenalo konec jejich pojetí světa, a to proto, že jejich pojetí reality, které je založeno na popření existence subjektivity, je podle jejich představ jediné možné (Searle 1992, p. 55).

Odmítavému postoji analytické filosofie k problematice vědomí Searle věnoval v *Consciousness and Language* (2002) kapitolu „Analytic Philosophy and Mental Phenomena“ (Searle 2002, pp. 203-226), v níž identifikuje hlavní důvody, proč je pro analytickou filosofii problém přijmout mentální jevy; jsou to verifikacionismus, kognitivní věda a strach z karteziánství (Searle 2002, pp. 222-224). Searle na závěr zmíněné kapitoly odmítá nutnost postulovat pro interpretaci vědomí jakoukoli teorii – ať už monismus, dualismus nebo fyzikalismus či epifenomenalismus – což zdůvodňuje poukazem na fakt, že pro vysvětlení trávení není také žádná podobná teorie zapotřebí, a tudíž ani pro vědomí, které je stejně jako trávení biologickým jevem, nejsou nutné (Searle 2002, p. 225).

Ačkoli Searle materialismus masivně kritizuje, považuje jej nicméně za teorii, která v rámci současného světonázoru zaujímá v podstatě monopolní postavení; zatímco dualismus je obecně odmítán, protože se na rozdíl od materialismu neslučuje se současnou představou přírodního světa. S výhradním

postavením, které se materialismu dnes dostává, Searle samozřejmě nesouhlasí a vyslovuje opět otázku, proč právě materialismus by měl mít jako jediný nárok na formulování obecně přijímaného světónázoru (Searle 1992, p. 3).

Přijetí materialismu a s ním i dualistického slovníku má podle Searla jasný negativní dopad nejen na zkoumání vědomí, čímž se detailně zabývá v rámci „Six Unlikely Theories of Mind“ (Searle 1992, pp. 5-10), kde představuje celkem šest přístupů, které se profilují v rámci eliminativního materialismu a funkcionalismu<sup>49</sup>. Jednotlivé přístupy charakterizují následující tvrzení: mentální stavy jako takové neexistují; lidová psychologie je úplně mylná; na tzv. mentálních stavech není nic mentálního; se správným programem a s příslušnými vstupy a výstupy je možné, aby měl počítač myšlenky, pocity, atd.; mentalistický slovník („představa“, „přesvědčení“,...) bychom neměli chápat v tom smyslu, že skutečně zastupuje vlastní mentální jevy, ale pouze jako určitý způsob vyjádření; vědomí, v tom smyslu, v jakém o něm běžně uvažujeme, možná ani neexistuje.

---

<sup>49</sup> Funkcionalismus představuje dominantní formu současného materialismu; na jedné straně formuluje teorii mysli zastávající stanovisko, že zásadní vlastnost mysli představuje její kauzalita, a na druhé straně se zabývá teoretickou identifikací mysli s centrální nervovou soustavou. Srov. Borchert (2006), Vol. 6, p. 14. Pravděpodobně právě kvůli převaze funkcionalismu v rámci materialistických teorií současnosti se Searle ve svých pracech soustřeďuje na jeho detailní kritiku.

## 4. Svoboda vůle a intencionalita

Searlova koncepce intencionality, kterou jsme detailněji představili v druhé kapitole práce, je kromě kritiky programu silné umělé inteligence a snahy o zasazení vědomí zpět do celku přírody a zkoumání přírodních věd také zásadní pro odpověď na otázku: „Proč se nechceme vzdát víry ve svobodnou vůli?“

Problémem svobodné vůle se Searle detailně zabývá v šesté kapitole knihy *Mysl, mozek a věda* (1994) a ve *Freedom and Neurobiology* (2004).

V první části této kapitoly se budeme zabývat Searlovou kritikou determinismu a kompatibilismu a představíme jeho analýzu vysvětlení jednání; v druhé části se soustředíme na Searlův pokus převést problematiku svobodné vůle z roviny psychologické do roviny neurobiologické – přístup, kterým se Searle úspěšně vypořádal již s problémem vysvětlení vztahu *mysl – tělo* a v důsledku i *vědomí – mozek*.

### 4.1 Determinismus, kompatibilismus a kauzální vysvětlení jednání

Podle Searla je představa svobodné vůle, respektive představa lidí jako svobodně jednajících bytostí, zásadní pro naše globální pojetí sebe sama. Zároveň však zastáváme názor, že se všechno skládá z částic a jejich vzájemných vztahů – jimiž také všechno vysvětlujeme – což svobodě vůli příliš prostoru neponechává (Searle 1994, p. 93).

Standardní řešení problému svobodné vůle představuje podle Searla tzv. kompatibilismus, který v podstatě slučuje představu svobodné vůle s determinismem. Kompatibilismus nepopírá, že je vše ve světě determinováno, na druhou stranu ale hovoří o svobodném lidském jednání – což znamená, že k jednání nejsme nuceni nátlakem (Searle 1994, p. 95). Kompatibilismus ovšem podle Searlových měřítek jako vysvětlení svobodné vůle nedostačuje z toho hlediska, že pravou podstatu problému ve skutečnosti neřeší, protože:

„[...] nejde o to, zda vedle vnějších fyzikálních příčin a vnitřních nutkání jsou při způsobování našeho chování ve hře ještě jiné psychické důvody. Spíše jde o to, zda příčiny našeho chování, ať už jakékoliv, budou s to *determinovat* chování tak, že se děje, co se dít *musí*.“

Kompatibilismus se ve skutečnosti s běžným pojmem svobody vůle neslučuje (Searle 1994, p. 96), protože determinismus spočívá na tvrzení, že každé jednání předchází dostatečně kauzálně silné podmínky, jimiž je jednání determinováno; zatímco teorie svobodné vůle zastává stanovisko, že určitá jednání dostatečně kauzálně silnými podmínkami předcházena nejsou – na základě zmíněných definic je vyvrácena teze kompatibilismu, jelikož svobodná vůle představuje popření determinismu (Searle 2004, p. 47).

Searle argumentuje ve prospěch svobodné vůle primárně na základě každodenní zkušenosti, ačkoli je otázkou, do jaké míry se můžeme na naši zkušenost spolehnout. Protiargumentem k Searlově každodenní zkušenosti svobodné vůle jsou příklady hypnózy, kdy se zhypnotizovanému subjektu zdá, že jedná svobodně, ačkoli pozorovatelům je jasné, že se ve skutečnosti nalézá v posthypnotickém stavu a svobodně tedy nejedná (Searle 1994, p. 97). Zdánlivý protiargument je Searlem vysvětlován jako argument ve prospěch svobodné vůle poukazem na empirickou nepravděpodobnost, že by veškeré lidské chování probíhalo jako v hypnóze (Searle 1994, pp. 97-98). Tímto stanoviskem chce Searle pravděpodobně upozornit na to, že pokud skutečně chceme vysvětlit (ne)existenci svobodné vůle, je třeba se soustředit na chování, tak jak běžně probíhá, a nikoli na výjimečné případy typu hypnotického chování.

Vedle kompatibilismu Searle vylučuje také mentální determinismus zastávající názor, že veškeré chování je psychicky nutné – opět na základě běžné evidence. Searle nepopírá, že jednání probíhá na základě intencionálních stavů, které fungují kauzálně, upozorňuje ale, že u nich nelze hovořit o přísné determinaci vztahu příčiny a následku:

„Pokud jde o psychické příčiny, mohli bychom jednat také jinak. Případy hypnózy a psychicky nutkavého chování jsou naproti tomu obvykle patologické a lze je snadno odlišit od normálního svobodného jednání.“



Z hlediska psychologie máme tedy dostatek prostoru pro lidskou svobodu.“  
(Searle 1994, p. 98)

Na koncepci lidské svobody je zásadní její vázanost na vědomí, což znamená, že svobodu jsme ochotni připsat pouze stvořením s vědomím. V souvislosti s argumenty pro svobodu vůle se obvykle zmiňuje proces racionálního rozhodování, tento proces však podle Searla není tím nejzásadnějším, je pouze jedním typem zkušenosti, která nás v předpokladu svobodné vůle utvrzuje. Nejdůležitějším druhem zkušenosti v otázce svobodné vůle je pro Searla zkušenost vázanosti svobody na intencionální jednání, přičemž podmínky pro splnění intencionality „spočívají ve výskytu jistých pohybů těla způsobených bezprostředně záměrem v jednání.“ Ani příklad posthypnotického stavu se nevyklučuje se zmíněnou koncepcí, jelikož člověk, který se v posthypnotickém stavu nachází, v podstatě neví, co dělá, protože skutečný záměr svého jednání si neuvědomuje – je nevědomý (Searle 1994, pp.102-104). V Searlově koncepci intencionálního jednání představuje zkušenost svobody zásadní součást jednání, respektive jeho podstatu.

Jednání ze svobodné vůle podle Searla nepředstavuje tvrzení, kterého bychom se mohli vzdát tak snadno jako například heliocentrické teorie sluneční soustavy. Na rozdíl od kopernikánského pojetí sluneční soustavy není pocit svobody pouze rysem uvažování – je částí běžného každodenního jednání (Searle 1994, pp. 104-105).

Ani Searlův pohled na svobodnou vůli a vysvětlení její zásadní role v intencionálním jednání však neřeší samotný problém svobodné vůle – jak může ve světě, kde se každý jev zdá být determinovaný příčinou mu předcházející, existovat čistě svobodné jednání. Jak Searle přiznává, problém svobodné vůle je jeden z filosofických problémů, jehož řešení se ani nepřibližujeme (Searle 2004, pp. 10-11). Potíž se svobodou vůle spočívá v předpokladu, že přírodní jevy je nutné vysvětlovat deterministicky. V případě problému vztahu mysli a těla, respektive vědomí a mozku se Searle jednak zbavil tradičního dualistického slovníku, který vylučoval chápání mentálního jako zvláštního druhu fyzikálního a zbylé otázky – jak přesně mozek způsobuje vědomé stavy; které neuronové

procesy způsobují vědomou zkušenost a další – přesunul do oblasti neurobiologie. Je otázkou, zda i v případě svobodné vůle bude problém možno převést taktéž na neurobiologickou úroveň (Searle 2004, pp. 38, 40-41) a přiblížit se tak alespoň částečně jeho vyřešení.

V *Mysl, mozek a věda* (1994) Searle představil důvody, proč nejsme ochotni opustit předpoklad svobodné vůle v našem rozhodování, založené na nutnosti svobody pro intencionální chování; ve *Freedom and Neurobiology* (2004) na zmíněné důvody navazuje a hovoří o tom, že příčiny, respektive důvody, předcházející jednání nepředstavují kauzálně dostačující podmínky pro jednání. Searle vyvozuje, že mezi příčinami jednotlivých stádií při rozhodování nebo jednání existují jisté mezery nebo prodlevy (gap<sup>50</sup>):

- 1) jednak mezi důvody pro jisté rozhodnutí a vlastním rozhodnutím,
- 2) mezi rozhodnutím a počátkem jednání
- 3) (když se například snažím naučit německy) také mezi počátkem jednání a jeho pokračováním až k ukončení.

Je však třeba upozornit, že mezery se nacházejí mezi jednotlivými vědomými stavy a nikoli mezi vědomými stavy a tělesnými pohyby nebo mezi stimulem a vědomým stavem. Vědomá zkušenost podobných prodlev je podle Searla to, co nás utvrzuje v předpokladu svobodné vůle; úskalí ale spočívá ve faktu, že podobně jako jsme utrzováni ve víře ve svobodnou vůli, stejně tak se nechceme vzdát ani determinismu. Podle Searla by bez vědomé zkušenosti těchto prodlev neexistoval ani problém svobody vůle – i tento problém tedy v Searlově koncepci úzce souvisí s vědomím. Je však třeba znovu si klást otázku, zda je naše zkušenost svobodné vůle spolehlivá a zda také – mimo zkušenost samu – odpovídá i něčemu ve skutečnosti (Searle 2004, p. 42-45).

Pro potřeby bližší analýzy problému svobodné vůle Searle předpokládá, že nás zkušenost svobodné vůle neklame a zároveň, že je teze mentálního determinismu chybná. Searle navíc postuluje hypotézu, že problém svobodné vůle

---

<sup>50</sup> Searle definuje *gap* následujícím způsobem: „[...] the conscious experience of the distinctive features of free, voluntary, rational actions [...]“ Searle (2004), p. 45

nespočívá v psychologické rovině, ale v rovině neurobiologické (Searle 2004, pp. 46-47).

Analýzu problému svobodné vůle Searle začíná exkurzem do oblasti vysvětlování lidského jednání a přijetím tří spolu souvisejících předpokladů:

- 1) Vysvětlení zakládající se na důvodech jednání obvykle nezmiňují kauzálně dostačující podmínky.
- 2) Takové vysvětlení jednání představuje adekvátní vysvětlení jednání.
- 3) Adekvátní kauzální vysvětlení zmiňuje podmínky, které jsou vzhledem ke kontextu kauzálně dostačující.

Vysvětlování jednání za pomoci běžně užívané kauzality není podle Searla legitimní, jelikož logickou strukturu vysvětlení jednání nelze formulovat tak, že A způsobilo B, ale že nějaký subjekt S učinil nějaké jednání J na základě nějakého důvodu D – logická forma jednání se liší důrazem na podání důvodu (nikoliv kauzálně dostačujících podmínek) jednání, z čehož Searle ve *Freedom and Neurobiology* (2004, p. 55) vyvozuje, že vysvětlení jednání na základě podání důvodů jednání nepředstavuje běžné kauzální vysvětlení – jež za přijetí třetího Searlova předpokladu kauzálně dostačující podmínky uvádět musí. Do sledu událostí je tedy dosazen neredukovatelný racionální činitel, se kterým Searle také počítá při vyjasnění druhého předpokladu – vysvětlení jednání na základě důvodů tohoto jednání je adekvátní, protože osoba, která se jednání dopustila ví (v ideálním případě), že se její jednání na zmíněných důvodech opravdu zakládalo (Searle 2004, pp. 53-54). Adekvátnost třetího postulátu je tedy zaručena existencí racionálního činitele, který se na základě jistých důvodů daného jednání dopouští (Searle 2004, p. 57).

Vědomá zkušenost prodlevy, na níž se zakládá celý problém svobodné vůle, je na základě Searlovy analýzy reflektována i v logické struktuře vysvětlení jednání.

## 4.2 Neurobiologická rovina problému svobodné vůle

Stejně jako v případě problému vědomí se i zde Searlovo vysvětlení daného problému odvíjí od předpokladu, že pokud není svobodná vůle pouhou iluzí a naopak představuje vlastnost světa, pak musí být určitým způsobem skutečná, respektive předpokládá existenci určité vlastnosti mozku, která svobodnou vůli uskutečňuje (Searle 2004, p. 58).

Jak jsem již uvedla v první části kapitoly, Searle spojuje problém svobodné vůle s vědomou zkušeností prodlevy (gap). Jelikož se Searle snaží převést zmíněný problém do oblasti neurobiologie, musí se ptát, jak je tato prodleva reflektována na neurobiologické úrovni (Searle 2004, p. 59).

Searle pracuje se dvěma hypotézami – první předpokládá, že stav mozku je kauzálně dostačující pro jednání, druhá naopak předpokládá, že stav mozku kauzálně dostačující není. První hypotéza představuje neurobiologický determinismus v kombinaci s psychologickým libertarianismem a pohlíží na svobodnou vůli jako na něco iluzorního – máme sice pocit svobodné vůle při jednání, ale na neurobiologické úrovni žádná svobodná vůle neexistuje. Hypotéza č. 1<sup>51</sup> je problematická, jelikož popírá, že by zkušenost svobodné vůle hrála kauzální nebo explanační roli v chování. Searle naopak tvrdí, že chování se zcela odvíjí od neurobiologie podmiňující svalové kontrakce (Searle 2004, pp. 61-62). Druhá hypotéza na druhou stranu implikuje otázku, jak by musel mozek fungovat, aby byl v souladu s tím, že každá vlastnost vědomí je determinována stavem, v němž se nacházejí mikročástice a zároveň, že vědomí působí kauzálně tak, že *determinuje* následující stav systému – a to prostřednictvím procesů, které *nejsou determinovány*, ale jsou produktem svobodného rozhodování nějakého racionálního já, jež má pro své jednání jisté důvody. Hypotéza č. 2 je značně komplikovaná, ale podle Searla paradoxně přesně charakterizuje situaci, v níž se lidské bytosti nacházejí:

---

<sup>51</sup> Searle tuto hypotézu později přirovnává k tradičnímu programu umělé inteligence: „We simply put in computer programs that will give the robot an algorithmic solution to the problems posed by the input stimuli and the states of the system.“ Searle (2004), p. 66

„Jsme roboti s vědomím, jejichž vědomé stavy jsou vázány na procesy odehrávající se v neuronech, zároveň však také někdy jednáme na základě nedeterminujících vědomých procesů (odtud neuronové procesy), což se odvíjí od toho, že naše racionální já činí rozhodnutí na základě určitých důvodů.“<sup>52</sup>

Definitivní odpověď na otázku po (ne)existenci svobodné vůle Searle neformuluje. Jediný závěr, který Searle vyvozuje ze zkoumání daného problému je alternativa dvou hypotéz, z nichž první je méně komplikovaná a lépe slučitelná s všeobecnou představou, kterou máme o biologii, na druhou stranu však vede k nepřijatelným závěrům; druhá hypotéza je oproti té první značně komplikovanější a sama předpokládá vyřešení několika značných problémů (Searle 2004, pp. 77-78). I přes Searlovu snahu tedy zůstává problém (ne)existence svobodné vůle bez konkrétního návrhu na řešení. Nicméně i přesto, že Searle nepodal jasnou odpověď na otázku po svobodné vůli, považuji jeho analýzu za přínosnou z toho hlediska, že upozorňuje, že kompatibilismus ve skutečnosti nepředstavuje teorii, která by svobodu vůle skutečně vysvětlovala. Za nejzásadnější bod celé analýzy daného problému však považuji objevení tzv. prodlevy (gap), která je nedílnou součástí každodenního rozhodování a zároveň v podstatě hlavním důvodem existence otázky po svobodě vůle.

---

<sup>52</sup> Srov. „We are conscious robots whose states of consciousness are fixed by neuronal processes, and at the same time we sometimes proceed by nondeterministic conscious processes (hence neuronal processes) that are matters of our rational selves making decisions on reasons.“ Searle (2004) pp.70-71

## Závěr

Cílem diplomové práce bylo představit vybraná témata filosofie mysli Johna R. Searla a zároveň se pokusit odpovědět na otázku, zda je Searlův filosofický projekt udržitelný, případně, do jaké míry je inovativní a přínosný.

Searlova kritika programu silné umělé inteligence, jež spočívá v kritice Turingova testu neboli imitační hry na základě metafory a argumentu čínského pokoje a v kritice kognitivismu. Jak ukázal rozbor článku „Minds, Brains, and Programs“, ve kterém se *čínský pokoj* poprvé objevuje, Searle kritizuje současný program silné umělé inteligence spočívající na tvrzení, že systém složený z počítače a příslušného programu má porozumění a kognitivní stavy. Odmítnutí teze silné umělé inteligence Searle ilustruje poukazem na čistě formální syntaktickou charakteristiku programu, která nezahrnuje žádné obsahy (sémantiku), jimiž se rozdílí od formálně definovaných programů disponují naše mentální procesy. Nejzásadnější je ovšem Searlovo rozlišení na mentální a nementální systémy, jehož přijetí značí konec výrokům o *přesvědčení* termostatů a podobných přehmatů ze strany zastánců programu silné umělé inteligence. Přijetí zmíněného rozlišení představuje pro program silné umělé inteligence podle Searla nutný krok, pokud se chce program i nadále věnovat zkoumání mysli.

Dalším zásadním bodem Searlovy reakce na odpovědi zastánců silné umělé inteligence na metaforu *čínského pokoje* je přísné rozlišování simulace a duplikace, a to v tom smyslu, že pouhá počítačová simulace mozku není schopna vyprodukovat vědomí, jelikož se jedná právě o pouhou simulaci, která sama o sobě nezahrnuje duplikaci kauzálních účinků, která by byla pro produkci vědomí a mentálních stavů nutná. Simulace mentálních stavů nepočítá s intencionalitou, která je pro skutečné mentální stavy v Searlově filosofii mysli zásadní. Případné redefinování programu silné umělé inteligence již neleží v centru Searlovy kritiky jednoduše proto, že by se již nejednalo o původní program, který považuje myšlení za počítačové operace s formálně definovanými jednotkami – což Searle striktně odmítá na základě předpokladu, že formální vlastnosti nejsou pro intencionalitu konstitutivní. Zásadní problém představuje omezenost simulace,

kteřá spočívá v tom, že nikdy sama o sobě nemůže stvořit duplikát. Searlova kritika je i přes četné námitky aplikovatelná bez ohledu na technologický pokrok, jelikož je zaměřena proti samotné koncepci komputace, již představil Church a Turing. Nuly a jedničky nejsou ze své definice – jakožto čistě abstraktní syntaktické entity – součástí kauzálního mechanismu založeného na přesunu energie, jako je tomu v případě neuronových spojů.

Cílem Searlovy kritiky projektu silné umělé inteligence není popření tvrzení, že stroje mohou myslet nebo že lze vytvořit umělý mozek. Chyba programu silné umělé inteligence podle Searla spočívá v definici komputace, která je pak dále chybně používána k vysvětlování myšlení. Searle nevyklučuje existenci strojů schopných myšlení, ovšem digitální počítače těmito stroji nejsou, jelikož jsou schopny vědomí a myšlení pouze simulovat, a nikoli duplikovat. Searle se tedy nevymezuje proti simulaci myšlení za pomoci digitálního počítače, ale proti zaměňování simulace za duplikaci.

Stejně jako Searle odmítl pouze silnou verzi programu umělé inteligence a nikoli slabou verzi stejného programu, odmítá rovněž silnou verzi Turingova testu – který je chybně používán pro určení přítomnosti kognitivních stavů ve stroji, a nikoli pouze jako potvrzení či vyvrácení úspěšné emulace lidského chování. Úspěšná emulace lidského chování není pro Searla dostatečným důvodem, abychom systému připsali mentální stavy. Ačkoli Searle odmítá behaviorismus a explicitně nenabízí žádnou konkrétní alternativu k Turingově testu, zdůrazňuje, že pro rozhodnutí o připsání mentálních stavů bude třeba ne jeden test, ale několik testů, přičemž klíč k řešení představuje systém sám.

Searle nepřišel s definitivním řešením problému zjištění přítomnosti mentálních stavů u jiných osob a zvířat, ale prostřednictvím analýzy problému upozornil, že mentální stavy nelze přisuzovat *pouze* na základě chování, ale také s přihlédnutím k fyziologické podobnosti a podobnosti kauzálních mechanismů, které zprostředkovávají vztah mezi stimulem a odpovědí na stimul. Úspěšnost v Turingově testu není dostačující pro připsání mentálních stavů proto, že je nutné zabývat se i skutečností vnitřních mentálních stavů, a nikoli pouze vnějším chováním. Searle se v této souvislosti odvolává na požadavek, aby naše tvrzení

odpovídala skutečností ve světě, což považuji za zcela adekvátní. Navíc kategorie úspěšnosti a neúspěšnosti v důsledku nic neznamenají, jelikož nejsou přírodě vlastní – existují pouze v závislosti na existenci pozorovatele. Na základě aplikování zmíněných kategorií a nezájmem o skutečnost mentálních stavů můžeme vyvozovat, že Searle pravděpodobně odmítá považovat koncepci silné umělé inteligence za součást přírodních věd a jejich zkoumání.

Prostřednictvím kritiky silné umělé inteligence Searle dospěl nejprve k závěru, že syntax není vlastní sémantice (z toho důvodu čistě formálně či syntakticky definované operace programu nemohou dostačovat pro porozumění) a později zproblematizoval i samotnou syntaktickou charakteristiku, která představuje observačně-závislou vlastnost, respektive interpretaci, a tedy nikoliv vlastnost fyzice vlastní. V souladu se syntakticky založenou definicí komputace, nelze na základě fyzikálních vlastností považovat nic za digitální počítač, a tudíž ani mysl, čímž je v důsledku vyvrácen program kognitivismu, jež si kladl za cíl vysvětlovat myšlení pomocí analogie mozek – počítač, respektive vysvětlování fyzikálního systému nalezením formálního vzorce, který systém (mozek) sdílí se svou počítačovou simulací. Hlavní nedostatek kognitivismu spočívá v neschopnosti zajistit kauzální vysvětlení jevu, který počítač simuluje, jinými slovy řečeno – počítačové vysvětlení a popisy zakládající se na zpracování informací nemají žádnou kauzální realitu. Searle dává přednost vysvětlení naturalistickému, tedy takovému vysvětlení jevu, které je založeno na intencionalitě. Komputaci a zpracování informací lze považovat pouze za *popisy*, a nikoli za kauzální vysvětlení.

Searlova filosofie myslí však neznamená pouze kritiku silné umělé inteligence, ale také vlastní projekt znovuzrození mysli, který je spojen s problematikou vědomí, respektive s jeho ontologickým statusem a definicí. Druhá kapitola ukázala, že díky Searlově rozlišení na epistemickou a ontologickou rovinu protikladné dvojice subjektivní – objektivní je možné epistemicky objektivně zkoumat ontologicky subjektivní jev, jakým je vědomí, protože ontologická subjektivita pouze poukazuje na fakt, že mentální stavy, a tedy de facto vědomí, mohou existovat pouze tak, že jsou prožívány určitou osobou – jejich existence je vázána na existenci nějakého já; tím však není vyloučena možnost objektivního



zkoumání. Searlův projekt znovuobjevení mysli a vědomí pro objektivní zkoumání přírodních věd je založen na nové definici vědomí, která charakterizuje vědomí jako vlastnost mozku vyšší úrovně; bylo by však chybou domnívat se, že Searle přisuzuje vědomí epifenomenální charakter – vědomí představuje podle Searla kauzálně emergentní vlastnost chování neuronů. Vědomí je v rámci Searlova biologického naturalismu chápáno jako další z mnoha biologických jevů, který by neměl stát mimo oblast zkoumání přírodních věd – pokud chceme zkoumat skutečnost opravdu objektivně. Zásadním předpokladem Searlova naturalistického pohledu na vědomí je jeho vázanost na mozek v tom smyslu, že vědomí je ve skutečnosti vlastností mozku, a to na vyšší úrovni. Naturalistický pohled na vědomí Searlovi v důsledku umožňuje skloubit vědomí s tzv. základními fakty skutečnosti. Ani nemožnost pozorování vědomí pro Searla neznamena nemožnost daný jev vědecky zkoumat. Searlově definici vědomí nelze namítat, že by byla nedostatečně pevná a adekvátní, jelikož pro počáteční fázi zkoumání vědomí, kdy je nutné se primárně vyhnout zaměňování vědomí za jeho zvláštní formy, je Searlova naturalistická definice vědomí zcela dostačující. Searle v rámci svého pojetí vědomí nepodává jeho definitivní definici ani detailní vysvětlení. Úloha Searlova pojetí je spíše anticipační a přípravná – zaměřená na identifikaci a jasné vymezení problému vědomí; odmítnutí dualismu, který spočívá v neschopnosti považovat vědomí za součást fyzikálního světa; redukcionismu, který od vědomí odděluje jeho podstatu spočívající v subjektivitě; behaviorismu – na základě principu o vzájemné nezávislosti chování a vědomí – a v neposlední řadě rovněž vyvrácení evoluční neúčinnosti vědomí. Vlastní průběh zkoumání vědomí Searle pouze rozvrhuje do tří stádií, na jehož konci bychom měli dojít k formulaci teorie, která by formalizovala kauzální vztah mezi neurobiologickými koreláty vědomí a vědomím.

Mezi vlastnosti vědomí patří v Searlově pojetí také intencionalita, která je součástí kritiky programu silné umělé inteligence, a to v souvislosti s jedním z intencionálních jevů – porozuměním. Na základě rozlišení intencionality vlastní a odvozené na jedné straně a intencionality metaforické na straně druhé se Searle vymezuje proti přepisování stejné intencionality, jakou mají mentální stavy, přístrojům typu termostat a podobně. V rámci svého postoje, který Searle označil

za naivní mentalismus, rovněž odmítá pojetí nevědomých mentálních jevů jako dispozic k chování a interpretuje je jako dispozice k vědomým myšlenkám, přičemž všechny nevědomé stavy musejí být principiálně přístupné vědomí, protože existenci hluboce nevědomých mentálních stavů Searle vylučuje.

Díky projektu znovuoobjevení mysli a vědomí pro zkoumání přírodních věd se Searlovi podařilo proniknout až k samotnému jádru problému vědomí, které spočívalo v řešení vztahu mysli a těla, potažmo vědomí a mozku. Searle zmíněný problém vyřešil velmi elegantně, jelikož jej přenesl z roviny metafyzické a filosofické do roviny neurobiologické, kde vědomí představuje vlastnost mozku způsobovanou neurobiologickými procesy.

Pro Searlův filosofický projekt je intencionalita nepostradatelná, tento fakt nás však ještě neopravňuje k tomu, abychom jeho filosofii mysli mohli interpretovat fenomenologicky – jednak protože Searle sám explicitně odmítá, že by se Husserlovou filosofií jakkoli inspiroval a také proto, že Searle odmítá introspektivní metodu zkoumání vědomí a dává přednost logické analýze založené na formulování podmínek splnění intencionálního jevu. Stejně jako vědomí, je v Searlově filosofii mysli i intencionalita biologická vlastnost. Searle fenomenologii úplně neodmítá, ovšem její možnosti v otázce zkoumání vědomí považuje za omezené. Searlův pohled na fenomenologii můžeme interpretovat jako kritický a pragmatický – kriticky Searle pojednává o tzv. fenomenologické iluzi a pragmaticky pohlíží na možnost částečného využití fenomenologické metody.

Stejně jako nelze pohlížet na Searlovu filosofii mysli prizmatem fenomenologie, není adekvátní ani interpretace dualistická, jelikož právě dualismus představuje teorii, proti níž se Searle zasazuje snad nejvíce. Searle se odvrací od tradičního dualistického slovníku a popírá tradiční protikladnou dvojici fyzikální – mentální a s ní i zdánlivě nevyřešitelný problém vztahu mysli a těla, potažmo vědomí a mozku – jež je ústředním tématem Searlovy filosofie mysli. Jedinou adekvátní interpretaci představuje biologický naturalismus a naturalistická filosofie, která se svět nesnaží rozdělovat, nýbrž jej chápe jako jeden celek, jehož součástí je i vědomí. Searle v otázce vědomí neuznává nutnost formulovat

monistickou, dualistickou nebo jakoukoli jinou teorii, jelikož ani pro jiné biologické jevy, jako je např. trávení a fotosyntéza, nejsou žádné teorie zapotřebí.

Kromě tradičního problému vztahu mysli a těla, který Searle vyřešil popřením a odmítnutím tradičního dualistického slovníku, se Searle snaží adresovat ještě jeden tradiční problém – ne(existenci) svobodné vůle. Problém svobodné vůle, stejně jako problém vztahu mysli a těla, se Searle pokusil přesunout z roviny filosofické do roviny neurobiologické. Na rozdíl od problému vztahu mysli a těla však v případě svobodné vůle Searle nepodává návrh na jeho definitivní řešení, nýbrž podává spíše důkladnou analýzu daného problému.

Filosofie mysli Johna R. Searla představuje odvážný filosofický projekt, jehož primárním cílem není podat *definitivní* řešení vybraných filosofických problémů, ale odhalit jejich pravou podstatu. Searlovo filosofické bádání neusiluje o nalezení definitivních odpovědí na palčivé filosofické otázky, ale naopak o kladení správných otázek, které buď přispějí k vyřešení daného problému, nebo alespoň naznačí, kterou cestou je třeba se při dalším zkoumání ubírat.

## Resumé

Filosofie Johna R. Searla se zabývá filosofickými problémy ve třech hlavních oblastech – mysl, jazyk a společnost. Práce se soustřeďuje na nejjzákladnější ze jmenovaných oblastí, kterou představuje filosofie mysli. Pozornost je věnována především Searlově reakci na silnou umělou inteligenci, projektu znovuoobjevení mysli a intencionality, vyvrácení tradičního karteziánského dualismu a problému svobodné vůle.

Searlovu filosofii mysli charakterizuje aplikace analytické metody, která mu umožňuje zjistit, zdali daný filosofický problém není postaven na chybném předpokladu nebo na systematickém používání vágních pojmů. Prostřednictvím analytické metody je Searle schopen dobrat se až samotné podstaty problému. V případě vztahu mysli a těla dovedla Searla analýza až k popření existence problému, což znamenalo možnost zasadit vědomí i intencionalitu zpět do objektivního zkoumání přírodních věd, jelikož vědomí představuje biologickou vlastnost. Searlova filosofie představuje naturalistické pojetí, které v případě problému mysl – tělo, vědomí i svobodné vůle vede až k přesunu problému z oblasti filosofie do oblasti přírodních věd – respektive neurobiologie.

Searle se v rámci svého filosofického projektu nezabývá pouze analýzami filosofických problémů, ale rovněž i vyvrácením teorií, které jsou z jeho pohledu mylné. Mezi tyto teorie patří: silná umělá inteligence, kognitivismus a behaviorismus. K odmítnutí teze silné umělé inteligence se Searle dopracuje poukazem na čistě formální, syntaktickou charakteristiku počítačového programu, který na rozdíl od mentálních procesů nepracuje na bázi sémantiky. Zásadní je rovněž Searlovo rozlišení na mentální – nementální systémy a rozlišení simulace – duplikace, které je klíčové pro odmítnutí představy mozku jako digitálního počítače a tím pádem i celého programu kognitivismu.

Filosofický projekt Johna R. Searla je zcela jistě odvážný a užitečný, a to ačkoli jeho primárním cílem není vždy podat definitivní odpověď na závažné filosofické otázky, ale spíše analýza těchto otázek, která přispěje k lepšímu

uchopení daného filosofického problému, jež v důsledku může vést až k jeho vyřešení.

## Summary

The philosophy of John R. Searle deals with philosophical problems in three areas – mind, language, and society. The thesis is focused on the most basic area represented by the philosophy of mind, particular attention is being paid to Searle's reaction to the strong artificial intelligence, his project of rediscovering the mind and intentionality, the refutation of traditional Cartesian dualism, and the free will problem.

Searle's philosophy of mind is characteristic by its usage of the analytical method, which enables him to find out whether the philosophical problem in question does not stand on a false assumption or a systematical application of ambiguous terms. With the aid of analytical method, Searle is often able to discover the core of the problem, which in case of mind-body problem lead to the refutation of the problem. The refutation of the mind-body problem enabled Searle to put consciousness and intentionality back in the objective research of natural sciences because consciousness represents a biological property. Searle's philosophy embodies the naturalistic conception that in case of the mind-body problem, the problem of consciousness, and the free will problem – transferred the philosophical problems from the philosophical domain into a domain of the natural sciences, in fact, neurobiology.

In his philosophical project, Searle does not deal only with analyses of the philosophical problems but also with refutations of the theories that are false according to his view. These theories include: strong artificial intelligence, cognitivism, and behaviorism. The refutation of strong artificial intelligence is based on the distinction between purely formal, syntactic characteristic of a computer program that in contrast to mental processes does not include semantics. Another two important distinctions are: mental – nonmental systems, and simulation – duplication that represents the crucial point for the refutation of the idea that brain is a digital computer, which in fact refutes the whole cognitivist research program.

John R. Searle's philosophical project certainly ranks among the challenging and effective ones, even though, its main objective is not always to present definite answers to significant philosophical questions but rather to analyze the questions in the first place, which would lead to a better understanding of the problem and possibly to its resolution.

## Anotace

ZBÍRALOVÁ Tereza. Katedra filozofie, Filozofická fakulta. „John R. Searle: Reakce na silnou umělou inteligenci a jeho pojetí intencionality.“ Mgr. Filip Tvrdý. 75 stran. 20 titulů použité literatury.

**Klíčová slova:** umělá inteligence, intencionalita, vědomí, počítač, digitální počítač, Turingův stroj, imitační hra, svobodná vůle, karteziánský dualismus, materialismus, fenomenologie, kognitivismus, subjektivita, objektivita, porozumění, argument čínského pokoje.

Cílem diplomové práce je představit filosofii mysli Johna R. Searla, a to primárně prostřednictvím jeho reakce na koncepci silné umělé inteligence, která se odvíjí od jeho pojetí intencionality.

V úvodu práce je naznačen projekt, na který se John R. Searle v rámci své filosofie soustřeďuje, jakou roli v tomto projektu hraje Searlova filosofie mysli – na niž se zaměřuje diplomová práce – a jaký druh filosofické metodologie Searle ve svém zkoumání upřednostňuje.

Hlavní část práce sestává ze čtyř kapitol, z nichž každá se soustřeďuje na jedno z vybraných témat Searlovy filosofie mysli: kritiku koncepce silné umělé inteligence; problematiku vědomí a projekt znovuzrození mysli; tradiční karteziánský dualismus a problém svobodné vůle.



## Annotation

ZBÍRALOVÁ Tereza. Department of Philosophy, Philosophical Faculty. „John R. Searle: Reaction to the Strong Artificial Intelligence and His Conception of Intentionality.“ Mgr. Filip Tvrдый. 75 pages. 20 titles of used literature.

**Key words:** artificial intelligence, intentionality, consciousness, computer, digital computer, Turing machine, imitation game, free will, cartesian dualism, materialism, phenomenology, cognitivism, subjectivity, objectivity, understanding, Chinese Room Argument.

The aim of the diploma thesis is to present John R. Searle's philosophy of mind through his reaction to the conception of strong artificial intelligence, which is based on his view of intentionality.

The introduction of the thesis seeks to provide a view into the project that represents the main focus of John R. Searle's philosophy, what role does Searle's philosophy of mind – which represents the main focus of the thesis – play in his project, and what kind of philosophical methodology does he prefer in his philosophical inquiries.

The main part of the thesis consists of four chapters. Each chapter is focused on one of the chosen themes from Searle's philosophy of mind: criticism of the conception of the strong artificial intelligence; the problem of consciousness and the rediscovery of the mind; the refutation of traditional cartesian dualism and inappropriate interpretations of Searle's philosophy; the free will problem.

## Seznam citované a použité literatury

- Bennett, Maxwell, et al. *Neuroscience and Philosophy - Brain, Mind, and Language*. New York: Columbia University Press, 2007.
- Borchert, Donald M., ed. *Encyclopedia of Philosophy*. 2nd ed. Thomson Gale, 2006.
- Burge, Tyler. „Philosophy of Language and Mind: 1950-1990“. *The Philosophical Review* 101.1 (Jan., 1992), pp. 3-51.
- Horák, Otakar. „Moje presvedčenia o Searlovej intencionalite“. Seminárni práce. Univerzita Palackého, 2008/2009.
- Nagel, Thomas. "Aké je to byť netopierom?" In *Antológia filozofie mysle*, ed. Silvia Gáliková & Egon Gál. Bratislava: Kaligram, 2003, pp. 399-413.
- Peregrin, Jaroslav. *Kapitoly z analytické filozofie*. Praha: Filosofia, 2005.
- Prado, C. G. *Searle and Foucault on Truth*. New York: Cambridge University Press, 2006, pp. 30-59.
- Rust, Joshua. *John Searle and the Construction of Social Reality*. London, New York: Continuum, 2006, pp. 64-72.
- Searle, John R. *Consciousness and Language*. Cambridge University Press, 2002.
- Searle, John R. *Freedom and Neurobiology - Reflections on Free Will, Language, and Political Power*. New York: Columbia University Press, 2004.
- Searle, John R., „Intencionalita a jej miesto v prírode“. In *Antológia filozofie mysle*, ed. Silvia Gáliková & Egon Gál. Bratislava: Kaligram, 2003, pp. 491-503.
- Searle, John R. „Minds, Brains, and Programs“. *The Philosophy of Artificial Intelligence*. Ed. Margaret A. Boden. Oxford University Press, 1990, pp. 67-88.
- Searle, John R. *Minds, Brains, and Science*. Cambridge: Harvard University Press, 1984.
- Searle, John R. *Mysl, mozek a věda*. Praha: Mladá fronta, 1994.

Searle, John R. "Nevedomé a jeho vzťah k vedomiu". In *Antológia filozofie mysle*, ed. Silvia Gáliková & Egon Gál. Bratislava: Kaligram, 2003, pp. 239-259.

Searle, John R. *Philosophy in a New Century*. New York: Cambridge University Press, 2008.

Searle, John R. *Rationality in Action*. The MIT Press, 2003.

Searle, John R. *The Rediscovery of the Mind*. The MIT Press, 1992.

Smith, Barry, ed. *John Searle in Focus*. New York: Cambridge University Press, 2003.

Turing, Alan M. „Computing Machinery and Intelligence“. *Mind, New Series* 59.236 (Oct. 1950), pp. 433-460.