Univerzita Hradec Králové

Filozofická fakulta

Katedra Filozofie

Searle a funkcionalismus

Diplomová práce

Autor: Zuzana Sixtová

Studijní program: N6101

Studijní obor: Filozofie

Forma studia: Prezenční

Vedoucí práce: prof. RNDr. Jaroslav Peregrin, CSc.

Hradec Králové, 2019

Univerzita Hradec Králové
Filozofická fakulta

Zadání diplomové práce

Autor:	**Zuzana Sixtová**
Studium:	F17NP0044
Studijní program:	N6101 Filozofie
Studijní obor:	Filozofie
Název diplomové práce:	**Searle a funkcionalismus**
Název diplomové práce AJ:	Searle and functionalism

Cíl, metody, literatura, předpoklady:

Searlův myšlenkový experiment s čínským pokojem je pokládán za velmi pádný argument proti funkcionalistickému chápání mysli: Searle argumentuje, že z toho, že něco funguje zcela jako myslící bytost ještě neplyne, že to skutečně myslí. Funkcionalisté mají naproti tomu, že pojem mysli se svou podstatou týká právě fungování, takže fungovat jako myslící bytost není nic jiného než myslet. Cílem práce je shrnout argumenty Searla i jeho vybraných odpůrců a kriticky zhodnotit argumenty obou stran.

Cole, David, "The Chinese Room Argument", The Stanford Encyclopedia of Philosophy (Winter 2015 Edition), Edward N. Zalta (ed.), URL = . Levin, Janet, "Functionalism", The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.), URL = . Preston, J. and M. Bishop (eds.), 2002, Views into the Chinese Room: New Essays on Searle and Artificial Intelligence, New York: Oxford University Press. Searle, J., 1980, 'Minds, Brains and Programs', Behavioral and Brain Sciences, 3: 41757 Searle, J., 1984, Minds, Brains and Science, Cambridge, MA: Harvard University Press; český překlad Mysl, mozek, věda, Mladá fronta, Praha, 1994

Garantující pracoviště:	Katedra filosofie a společenských věd, Filozofická fakulta
Vedoucí práce:	prof. RNDr. Jaroslav Peregrin, CSc.
Oponent:	Mgr. Ladislav Koreň, Ph.D.
Datum zadání závěrečné práce:	14.2.2018

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracovala (pod vedením vedoucího diplomové práce) samostatně a uvedla jsem všechny použité prameny a literaturu.

V Hradci Králové dne Sixtová Zuzana

Tímto bych chtěla poděkovat panu prof. RNDr. Jaroslavu Peregrinovi, CSc. za velmi vstřícnou a obětavou pomoc při vedení této práce a rovněž za poskytnutí velkého množství rad a odkazů k literatuře.

Anotace

SIXTOVÁ, Zuzana. *Searle a funkcionalismus*. Hradec Králové: Filosofická fakulta, Univerzita Hradec Králové, 2019, str. 68. Diplomová práce.

Diplomová práce se zaměřuje na pojetí funkcionalismu ve vztahu k filosofii mysli a zkoumá filosofickou možnost připsat myšlení i jiným než lidským entitám. Tento pohled je v textu představen na myšlenkách Alana Turinga, jenž se zabýval otázkou myslících strojů, kterou se pokouší zodpovědět pomocí imitační hry. Dále se diplomová práce zaměřuje na opačná stanoviska, která jsou zde podpořena obzvláště názory Johna Searla, který Turingovu snahu vyvrací svým protiargumentem čínského pokoje. Do práce je začleněno i pojetí mysli a těla, které je zde nahlíženo buď dualisticky, nebo monisticky. Diplomová práce následně představuje debatu, která se kolem možného funkcionalistického pojetí vytvořila.

Klíčová slova: Funkcionalismus, Alan Turing, myslící stroje, Turingův test, John Searle, argument čínského pokoje

Annotation

SIXTOVÁ, Zuzana. *Searle and Functionalism*. Hradec Králové: Philosophical Faculty, University of Hradec Králové, 2019, pp. 68. Diploma Dissertation.

This master's thesis focuses on the notion of Functionalism in relation to the philosophy of mind and explores the possibility of attributing thinking to non-human entities. This view is introduced through the thoughts of Alan Turing, who tries to answer the question of thinking machines with the help of the imitation game. The thesis also deals with some of the opposing stances, supported mainly by the views of John Searle, who tries to disprove Turing's effort by using the Chinese Room Argument. The work also includes different conceptions of the body-mind problem, both dualist, as well as monist. Lastly, the thesis introduces the debate that has formed around the possible functionalist conceptions.

Keywords: Functionalism, Alan Turing, thinking machines, Turing test, John Searle, Chinese Room Argument

Obsah

Úvod....9
1. Vztah mysli a těla....12
1.1 Searlův pohled....15
2. Funkcionalismus....20
2.1 Vývoj pojmu funkcionalismus....21
2.2 Alan Mathison Turing....25
2.2.1 Imitační hra a Turingův test....27
3. John Rogers Searle....31
3.1 Argument čínského pokoje....33
3.2 Může počítač myslet?....37
3.3 Intencionalita....38
3.3.1 Daniel Dennett....41
3.4 Reakce autorit k argumentu čínského pokoje....46
4. Kritika funkcionalismu....55
4.1 Ned Block....55
4.2 David J. Chalmers....57
4.3 Frank Jackson....60
4.4 Problém eliminativismu....62
4.4.1 Problém svobodné vůle....64
Závěr....68
Zdroje:....70

Úvod

Tato diplomová práce nesoucí název Searle a funkcionalismus je reakcí na stále se rozšiřující myšlenku inteligentních entit, které nejsou lidskými bytostmi. V souvislosti s touto myšlenkou se rozvíjí nekončící filosofická debata, zda je možné říkat, že tyto entity doopravdy myslí v tom smyslu, v jakém přisuzujeme mysl ostatním lidem. Zastánci umělých inteligencí mají pocit, že není problém jim tuto mysl přisoudit, protože stejně tak nevíme, jak myslí ostatní lidé a v jejich případě, na základě jakéhosi zdravého rozumu, neváháme a mysl jim přiznáváme. V případě lidských osob je pouze problematická podskupina nemocných nebo postižených lidí, kde si nejsme úplně jistí. Příkladem takových osob jsou například lidé v kómatu nebo děti ve velmi raných fázích jejich vývoje. Ostatním ovšem mysl bez rozmyslu přisuzujeme, protože většinou považujeme za zbytečné polemizovat nebo rozvracet tento typ tvrzení. Tento druh tvrzení nazýval G. E. Moore truismy a označuje tak tvrzení, která běžného člověka nenapadne zpochybňovat, příkladem se jedná o výroky jako: mám tělo, narodil jsem se, vedle mne existují další lidé a podobně. V tomto smyslu každý člověk vnímá ostatní jako myslící bytosti, a proto je zbytečné tento fakt zpochybňovat.[1] A to i přes to, že je extrémně těžké nějakým způsobem dokázat, že kdokoli kromě nás mysl má a dokáže na jejím základě myslet. Ovšem pokud bychom se zdráhali přisoudit mysl lidem a pokud bychom zpochybňovali i takto zjevná tvrzení, zabředli bychom do nekončícího skepticismu a jediné čeho bychom dosáhli, by byl fakt, že doopravdy nemůžeme nic považovat za skutečné a reálně poznatelné.

Vraťme se ale k otázce jiných než lidských entit, kterým jsou někteří myslitelé ochotní mysl také přisoudit. V této práci, tuto skupinu myslitelů zastupují funkcionalisté, kteří měli velký vliv na filosofii mysli a kognitivní vědy obecně. Jejich myšlenky se pro mnohé staly velmi přitažlivými, ale vzbudily i vlnu kritiky, která bude v této diplomové práci také představena. Velmi stručně řečeno se funkcionalisté domnívají, že funkčnost mysli jsme schopni pozorovat

[1] MOORE, G. E. A Defence of Common Sense. In: *Contemporary British Philosophy*.

pouze jejími vnějšími projevy, a tedy pokud se budou ne-lidské entity chovat stejně jako lidé, pak je jejich chování založeno na podobném základě a je možné je považovat za myslící. Tento obraz je možné pozorovat například u Alana Turinga, který myšlenku myslících strojů nastolil do podoby, která je aktuální i v dnešní době. A právě filosofickou možností inteligentních artefaktů se budeme v práci dále zabývat.

V případě ne-lidských entit, které označujeme za myslící, je vhodné alespoň krátce zmínit i druhy, které jsou stejně jako my biologického původu, tedy zvířata. Pro většinu filosofů je totiž jednodušší přisoudit mysl spíše zvířatům než mechanickým strojům, jež jsou sestrojeny lidmi. Obzvláště v posledních několika desetiletích se ochrana zvířat výrazně zvýšila a obecně se má za to, že přinejmenším savci jsou schopni cítit bolest a jednat tak aby byli schopni přežívat, což je do určité míry možno považovat za inteligentní chování, i když jde samozřejmě o „inteligenci," která se s tou lidskou nedá srovnávat. Například Stanislav Komárek k tomu říká, že každé zvíře je inteligentní svým vlastním způsobem, tedy pes je chytrý tak, aby mohl plnit roli psa.[2] Alespoň v tomto smyslu potom máme více druhů myslí, které se od sebe v určitých stupních vývoje liší.

Diplomová práce je rozdělena do čtyř hlavních kapitol. První kapitola nese název Vztah mysli a těla a bude zde nastíněn dvojí přístup, který se k tomuto vztahu dá uplatnit, dualismus a monismus. Dualistický přístup bude představen na práci René Descarta, který je s tímto termínem nejspíše nejvíce spojován a jehož vliv do určité míry přetrvává v myslích lidí dodnes. Následně se přesunu k monistickému přístupu, jenž je v moderní době přijímán jako přístup, který se více přibližuje realitě. Zmíním zde také přístup Johna Searla, jehož pohled na tuto problematiku bude pro zbytek práce důležité přiblížit.

Druhá kapitola, která je pojmenovaná Funkcionalismus, bude obsahovat nejprve vývoj pojmu, jehož základy je možné pozorovat už u Aristotela a jehož

[2] KOMÁREK, Stanislav. *Ochlupení bližní: zvířata v kulturních kontextech.* Str. 46.

vývoj vedl k různým větvím, které se v určitých ohledech liší. Následně přiblížím myšlenky Alana Turinga, jehož práce je pro dnešní představu o umělé inteligenci odrazovým můstkem. Proto představím i jeho myšlenkový experiment nazvaný Imitační hra, který vedl ke vzniku známého Turingova testu a stručně také uvedu představu, kterou Turing měl vzhledem k budoucím inteligentním strojům.

Ve třetí kapitole se zaměřím na myšlenky Johna Searla, který je nejznámějším autorem, jenž na Turingovu práci reagoval. Pokusím se představit jeho myšlenkový protiargument Čínského pokoje a jeho důvody, proč podle něho počítače nikdy nebudu schopni myslet tak jako lidé. Nastíním také Searlovu představu o tom, jakým způsobem lidská mysl funguje. V tomto bodě bude nejdůležitější jeho pojem intencionality, který v práci porovnám s intencionalitou Daniela Dennetta, jež pojem užíval jiným způsobem. Dále v této kapitole uvedu reakce autorit k zmíněnému argumentu Čínského pokoje.

V poslední čtvrté kapitole nazvané Kritika funkcionalismu se pokusím na základě různých myšlenkových experimentů různých autorů představit obecnou kritiku, která se k tomuto směru váže. Bude zde zmíněn Ned Block, David Chalmers a Frank Jackson. A také se pokusím zkonstruovat odpovědi funkcionalistů, které na tyto argumenty reagovali. Jedna z podkapitol bude věnována směru eliminativismus, jenž by se dal považovat za nejradikálnější obranu funkcionalismu a s tímto směrem uvedu také problematiku svobodné vůle, která bude zmíněna z pohledu Johna Searla.

1. Vztah mysli a těla

Problematika vztahu mysli a těla se dá považovat za jedno z věčných témat filosofie, které se dotýká i dnešní moderní představy o možné umělé inteligenci, jež je spjatá s tématem této diplomové práce. Všeobecně se považuje zkoumání mysli a těla za velmi obtížné, protože pojmy mysl a tělo často chápeme spíše intuitivně, a navíc se naše chápání váže na různé kontexty, v kterých se zrovna nacházíme. Při zkoumání vztahu mysli a těla se můžeme zaměřit buď na shody, nebo na odlišnosti mezi tělem a myslí. Na tuto problematiku se dá nahlížet dvěma způsoby, a to buď dualisticky, nebo monisticky. Dualismus je obecně označení pro filosofické nebo náboženské přesvědčení, že skutečnost je vysvětlitelná redukcí na dva většinou protikladné principy (zde je to duše nebo mysl – tělo).[3] Naopak monismus je označení filosofických názorů, které vysvětlují celek skutečnosti z jednoho principu a všechny další jevy považují za rozvinutí či modifikaci, tudíž se od svého původního principu nijak zásadně neliší.[4] Pomocí dualismu nebo monismu se snažíme vysvětlit právě vztah mysli a těla, který je pro filosofii mysli klíčový.

Co se týče dualistického pohledu v souvislosti s filosofií mysli je na místě zmínit René Descarta, který je jedním z nejznámějších filosofů, jež tento pohled rozpracoval, a jehož karteziánský dualismus má ve vztahu mysli a těla do dnes velký vliv. Descartes byl racionalistou, což znamená, že podle něho lidské poznání vychází především z rozumu. Postupuje metodou meditací, kdy se snaží pomocí svého rozumu dojít k nezpochybnitelným jistotám. Poté, co skoro vše zpochybní, včetně světa kolem sebe, a dokonce svého vlastního těla, dojde k první jistotě a tou je myšlení, potažmo pochybování, protože to je jediná věc, o které není možné pochybovat. Když mohu pochybovat, nemohu zde nebýt, jelikož jinak by zde nemohly být ani ony pochybnosti[5] *„…je to myšlení, jedině to se ode mne nedá odloučit. Já jsem, já existuji, to je jisté. Ale jak dlouho? Inu, pokud myslím;*

[3] Kol. Autorů. Dualismus In. *Filosofický slovník.*
[4] Kol. Autorů. Monismus In. *Filosofický slovník.*
[5] DESCARTES, René. *Meditace o první filosofii.* Praha: OIKOYMENH. Str.29-44.

neboť je snad možné, že jakmile bych naprosto přestal myslet, zmizel bych.“[6] Z tohoto hrubého náčrtu Descartovy práce je patrné, že člověka Descartes vysvětloval dualisticky. Popisoval lidskou bytost jako věc myslící (*res cogitans*) a věc rozprostraněnou (*res extensa*), přičemž tyto dvě části existují vedle sebe a navzájem na sebe působí. Tělo je pro něho jakýsi mechanismus, který funguje na základě zákonitostí. Mysl, která je potřebná k ovlivňování těla, potom dokáže tahat za určité páky ve strojovém těle a tím ho ovládat. Spojení těchto dvou substancí přitom podle něho probíhá v šišince mozkové, což je jedna z nejproblematičtějších částí Descartovy filosofie.[7]

Tento problematický dualistický pohled se nám navíc zdá velmi intuitivní, protože lidé mají přirozeně pocit, že mají jak fyzické, tak duševní vlastnosti, které se nám zdají být diametrálně odlišné. Jinak řečeno si lidé všeobecně myslí, že povaha mysli je nehmotná, kdežto svět je hmotný, což hraje do karet dualismu. Nejlépe je to nejspíše možné demonstrovat na příkladu medicíny. Lidé totiž tradičně berou duševní choroby jinak než choroby fyzického těla. Pokud si někdo zlomí nohu, pak pro něho není problém jít k lékaři, který problém vyřeší jakousi fyzickou opravou dané části těla. Ovšem pokud má někdo psychické problémy, pak se má často za to, že zde není potřeba nic fyzicky opravovat, a proto se spíše domníváme, že člověk postižený psychickou nemocí má pouze pocit, že je s ním něco v nepořádku. Dále je dualismus silný v našem přístupu ke smrti. Máme naději, že dokud existuje nehmotná duše, pak nemůžeme doopravdy zemřít, zemře pouze naše hmotné tělo, ale duše bude žít dál v jiných sférách bytí. Tento pohled nám nabízí představu, že naše závislost na tělech je buď podmíněná, nebo nutná pouze v případě tohoto života. Myšlenka nesmrtelnosti duše nebo mysli je v nás přítomna kvůli židovsko-křesťanské tradici, z které vycházel také Descertes, jenž byl samozřejmě věřícím člověkem.

Většina filosofů, kteří se touto problematikou zabývají, se kvůli nejasnosti vzájemného ovlivňování mysli a těla hlásí spíše k různým monistickým teoriím,

[6] DESCARTES, René. *Meditace o první filosofii.* Praha: OIKOYMENH. Str. 41.
[7] ROBINSON, Howard, „Dualism“, *The Stanford Encyclopedia of Philosophy*

tedy k myšlence, že mysl a tělo tvoří určitým způsobem jednotu. Ač se různí myslitelé v duchu monismu domnívají, že mysl nemůže být samostatná entita, stejně tak ji nechtějí úplně vypustit. Mysl je totiž něco, co každý z nás sám v sobě pozoruje, proto vznikly různé teorie identity, které jednoduše řečeno říkají, že mysl a činnost mozku jsou tou samou věcí.

Kupříkladu Donald Davidson vytváří teorii anomálního monismu. Je to návrh, jakým způsobem bychom mohli chápat mentální události jako součást fyzikálního světa a zároveň jako těmito zákony nevysvětlitelné. Mentální událostí myslí každou změnu, která se v našem psychickém životě odehraje a lze je popsat jako propoziční postoje. Fyzikální události jsou potom běžné události v čase a prostoru. Pro Davidsona je klíčové, že rozlišuje mezi samotnými událostmi a našimi popisy těchto událostí. Události jsou tedy mentální pouze jako mentálně popsané a podobně fyzická událost je fyzická, pouze pokud je popsaná fyzikálně. Tyto události existují nezávisle na našich popisech a stanou se mentálními nebo fyzikálními, až když jsou mentálně nebo fyzikálně popsané.[8] Davidson se tedy domnívá, že to, co se děje v našem mozku je možné popsat dvěma různými na sebe nepřevoditelnými způsoby. A to buď vědeckým způsobem jako vztahy neuronů anebo jakýmsi lidovým způsobem jako mentální dění. Vědecký popis je určený k popisu fungování mozku ve stejném smyslu, v jakém se popisují fyzikální zákony vnějšího světa, a lidový popis vznikl za účelem každodenního popisu lidského chování, které nepotřebuje přesné deterministické zákony, jako tomu je v případě vědy. Přitom podle něho jsou oba popisy určeny k úplně jiným účelům, a proto nemá smysl je nějak porovnávat nebo se je pokoušet překládat. V tomto případě se jedná o jednu substanci, o totožnou událost, která v sobě má dualismus ve smyslu různých popisů.[9]

[8] YALOWITZ, Steven, „Anomalous Monism“, *The Stanford Encyclopedia of Philosophy*
[9] PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Str. 244-245.

1.1 Searlův pohled

Pro tuto práci je nejdůležitější pohled, který ke vztahu mysli a těla zastával John Searle. Searle se také domníval, že problém mysli a těla je zapříčiněn právě dualismem, který je v nás stále zakořeněn. Vychází z toho, že lidé sami sebe vylučují z vědeckého přístupu ke světu. Máme samozřejmě vžitou představu, že jakožto lidské bytosti máme mysl, jednáme vědomě a racionálně. Jak však můžeme tuto představu skloubit s vědeckým pohledem na fyzikální svět, který se skládá výhradně z fyzikálních částic? Jak vysvětlit, že svět stvořený z fyzikálních částic obsahuje něco jako neuchopitelnou mysl nebo vědomí? Searle tuto problematiku shrnuje do otázky: *„Jaký je vztah naší mysli a ostatního univerza?*" neboli jaký je vztah mezi naší mysli a mozkem.[10]

To, že i dnes máme pocit, že je mysl něčím záhadným, má, dle Searla, co dočinění také s naším zastaralým slovníkem, který se užíval v sedmnáctém století, v době, kdy Descartes nastolil karteziánský dualismus. Dříve se ke studiu mysli vázaly kategorie, které by nás v souvislosti s jinými biologickými fenomény aplikovat nenapadlo. Tak například jistě by nikdo neviděl problém ve vztahu žaludku a trávení, určitě by se člověk, jenž podobné biologické fenomény přijímá, nepovažoval za dualistu. Přitom podobný vztah je i mezi myslí a mozkem, kde se to zdá být velmi problematické.[11]

Tedy souhrnem řečeno se již od Descarta táhne tradice pohlížet na lidskou mysl a na lidské tělo jako na dvě zcela odlišné substance. Na jedné straně jsou myšlenky, city a další mentální jevy, které považujeme za něco zcela subjektivního, vědomého a nemateriálního. Na straně druhé máme fyzikální jevy, jež naopak považujeme za fyzicky hmotné věci, které se nacházejí v časoprostoru a jsou v kauzálním spojení s okolním světem. Pokud se poté snažíme mysl nějak vysvětlit, pak v souvislosti s tímto naše snahy nutně musí jednu ze stran buď úplně popřít, nebo při nejmenším upozadit. Vzhledem k tomu, že sociální vědy nejsou schopny nabídnout takové výsledky, jako se to daří vědám přírodním, pak

[10] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. Str.13.
[11] Tamtéž, Str. 14.

přirozeně většinou dochází k popření, či podcenění mentálních jevů. Na této půdě pak vznikly směry, které explicitně nebo implicitně popírají, že existuje mysl v tom smyslu, jak ji běžně zdravým rozumem chápeme, tedy popírají, že bychom měli něco jako subjektivní a vědomé mentální stavy. Podle Searla je tento problém možné vysvětlit zodpovězením na otázku, proč tolik myslitelů dospívá k popření mentálních jevů. Dochází k tomu, že problém mysli je tak těžké vysvětlit, protože mentální fenomény mají čtyři vlastnosti, které se zdají být mimo vědecké pojetí fyzikálního světa.[12]

První vlastnost, která je od mysli neoddělitelná, je vědomí. Mysl obsahuje vědomé mentální stavy, které se zdají být neslučitelné s fyzikálními systémy. Pokud si představíme univerzum bez vědomí, zjistíme, že tento svět by neměl žádný význam. Vědomí je totiž výhradně lidskou vlastností, bez které by nebylo možné zažívat žádné aspekty lidské existence, lidé by nebyli schopni užívat jazyk, prožívat emoce atp.[13]

Další typickou vlastností mysli je intencionalita. Intencionalitou se myslí zaměřenost našich mentálních stavů na různé předměty, jež se vyskytují ve světě. Intencionalita se vztahuje také na přesvědčení, přání, naděje, obavy a další mentální stavy, které mohou být jak vědomé, tak nevědomé a které se vztahují mimo sebe sama na okolní svět, kterým se zabývají. Podobně jako vědomí se i intencionalita musí potýkat s problémem toho, jak může něco takového vůbec v lidském fyzikálním mozku existovat. Jak může mozek, který je složen z atomů odkazovat k něčemu, co je mimo něho.[14]

V souvislosti s bodem intencionality je vhodné zmínit Edmunda Husserla. Husserl je zakladatelem fenomenologie a jeho snahou bylo ukázat, že mezi zážitky našeho vědomí a fenoménem nějaké věci musí existovat zákonitá vazba, v které se musí prosadit i svébytnost věci samé.[15] Vztah mezi naším vědomím a věcmi nazýval právě intencionalitou. Tu, podobně jako Searle po něm, označuje

[12] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. str. 15.
[13] Tamtéž, str. 16.
[14] Tamtéž, str. 16.
[15] Kol. Autorů. Husserl Edmund In. *Filosofický slovník.*

za základní vlastnost vědomí být vědomím o něčem. Zde by se tedy mohlo zdát, že Searle by se podobně jako Husserl mohl zařadit pod fenomenologickou větev filosofie. Husserl však intencionalitu dále rozvíjí a za jeden z rysů intencionality považuje to, že *„Každý zážitek má svůj horizont, měnící se ve změně vědomých souvislostí tohoto zážitku a ve změně fází jeho vlastního zážitkového proudu – intencionální horizont poukazů na potencionality vědomí, jež přísluší jemu samému.“*[16] Je tedy znatelné, že Husserlovská intencionalita se od té Searlovské liší v tom, že Husserl k ní přidává ještě další věci, jako je například zmíněný horizont nebo introspekce, kdežto Searle používá metodu logické analýzy a intencionalitu drží pouze v rovině, jež je vypsaná v druhém bodu vlastností mysli a která se, jak později ukáži, řadí pod biologickou vlastnost mysli.

Třetí vlastností mysli je subjektivita mentálních stavů. Jedná se o to, že každý člověk vnímá svět ze svého vlastního subjektivního hlediska, přičemž tyto vnitřní mentální stavy nejsou přístupné ostatním lidem. Tato vlastnost mysli je opět problematická, jelikož to, co se považuje za zkoumatelné a skutečné, se zároveň považuje za objektivní.[17]

Čtvrtá vlastnost mysli je mentální kauzace. Lidé mají předpoklad, že to co se děje v naší mysli, ovlivňuje naše chování a naopak jsou lidé také ovlivňováni tím, co se děje s naším tělem. To znamená, že mentální obsahy kauzálně působí na fyzický svět a fyzický svět působí na mentální obsahy člověka. Psychika má v tomto smyslu vliv na materiální svět kolem, což je opět nejasný a problematický vztah.

Searle následně uvádí teze, pomocí nichž se snaží s těmito problémy vypořádat. První teze zní:

> *„Mentální fenomény, myslím tím veškeré mentální fenomény – ať už vědomé nebo nevědomé, optické nebo akustické, bolest, lechtání, svědění,*

[16] HUSSERL, Edmund. *Kartesiánské meditace*. Nakladatelství Svoboda, Praha, 1968. Str. 34-45.
[17] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. Str. 17-18.

myšlenky, tedy veškerý náš psychický život - jsou výsledkem procesů, jež se odehrávají v mozku.“[18]

To jednoduše řečeno znamená, že všechno, co se odehrává v naší psychice, je způsobeno procesy v mozku. K první tezi následně dodává tvrzení, že:

„Bolesti a jiné mentální fenomény jsou právě vlastnostmi mozku (a možná i zbytku centrální nervové soustavy).“[19]

Z těchto dvou tvrzení tedy vyplývá, že mozek způsobuje mysl a zároveň je mysl vlastností mozku. Na první pohled se zdá, že není možné, aby platila obě tvrzení zároveň. Je zde problematické připustit, že by vztah příčiny a následku fyzikálních fenoménů mohl zároveň připustit, aby byla jedna věc vlastností druhé, tedy, že by byla příčinou sebe sama. To, že je možné, aby obě tvrzení platila zároveň, se dá vysvětlit pomocí lepší interpretace kauzality. Příčinou mentálních fenoménů jsou procesy, jež se odehrávají v mozku pomocí nervových buněk a jejich různým uskupením, přičemž se v tomto systému, který je složen z buněk, zároveň realizují. Pro názornost této kauzality můžeme užít mikro a makro vysvětlení pro fyzikální věci ve světě. Můžeme vysvětlovat existenci předmětů a jejich vlastností pomocí mikro slovníku, tedy pomocí různých mřížek atomů a mikročástic nebo můžeme užít jinou úroveň makro slovníku a například tuhost stolu vysvětlit jako pevný předmět, který odolává tlaku, dají se na něho položit další věci, které jím neprostoupí atp. Tak se povrchové vlastnosti, které vnímáme, dají kauzálně vysvětlit chováním mikročástic v jiné rovině. Tyto částice v jiné rovině jsou pak příčinou vlastností vyššího řádu a zároveň se vlastnosti vyššího řádu realizují v systému, který je tvořen z mikročástic. Mikro a makro rozlišení je třeba aplikovat i na mozek, aby bylo možné vysvětlit mentální fenomény.[20]

Pokud připustíme, že toto vysvětlení je uspokojivé, pak nám nic nebrání v tom, abychom výše zmíněné čtyři vlastnosti mozku dokázali pochopit podobným způsobem, jakým chápeme ostatní fyzikální jevy. První problém s

[18] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. Str. Str. 18.
[19] Tamtéž, str. 19.
[20] Tamtéž, str. 20-24.

vědomím budeme schopni vyřešit, až přesně pochopíme procesy, které se v mozku dějí. Dokud nechápeme přesnou biologickou podstatu mozku, je těžké tyto procesy popsat, ovšem rozumíme obecné povaze procesů. Víme, že v mozku jsou nervové buňky, mezi kterými dochází k nějakým aktivitám a tím je možné vědomí. Problém intencionality je na tom podobně jako problém vědomí. Přesně ji pochopíme, až zjistíme, jak jsou intencionální stavy způsobovány biologickými procesy v mozku. Třetí problém o subjektivních mentálních stavech a jejich začlenění do objektivistického vysvětlení světa je možné vysvětlit následovně. Subjektivní mentální stavy se dají považovat za objektivní z toho důvodu, že subjektivita je jednoduše biologický fakt, který se ve světě nachází u všech živých tvorů. Každý z nás má stav vědomí, který není přístupný ostatním stavům vědomí, proto je existence subjektivity objektivním biologickým faktem. Mentální kauzace a to jak může ovlivňovat fyzické předměty, se dá vysvětlit tak, že pokud mám nějakou myšlenku, pak se v mém mozku doopravdy něco fyzicky děje. Mozková aktivita pomocí fyziologických procesů může tvořit pohyby těla. A jelikož jsou tyto mozkové aktivity vlastnostmi mozku, dají se popsat pomocí dvou úrovní, vyšší psychologické a nižší fyziologické. [21]

Searle problém mysli a těla vysvětluje tak, že nepovažuje mysl a tělo za dvě odlišné substance, protože mentální fenomény jsou podle něho vlastnostmi mozku. Tedy souhrnem řečeno podle něho neexistuje jiný druh věcí než věci fyzikálního charakteru. Pokud dokážeme pochopit tento fakt, pak se budeme schopni odpoutat od dualismu, a to nám umožní překonat problém mysli a těla takovým způsobem, že pochopíme, že takový problém tu ve skutečnosti vůbec není, protože budeme schopni definovat mentální stavy neboli vlastnosti systému neuronů jako fyzikální vlastnosti, které jsou na vyšší psychologické úrovni. Tím vrací mentální jevy zpět do vědeckého celku světa, čímž ukazuje, že je možné je objektivně zkoumat.

[21] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. str. 24-28.

2. Funkcionalismus

Funkcionalismus je myšlenka, která se objevuje v různých oborech, s tímto termínem pracuje například psychologie, filosofie, sociologie, ekonomie, biologie nebo architektura. Samotný výraz funkcionalismus, který se odvozuje od slova funkce, jež je převzatý z matematiky jako předem definovaný vztah mezi dvěma různými veličinami, je v ostatních oborech uplatňován spíše metaforicky. Většinou jde o intuitivní spojení funkce objektu (nebo procesu) s účelem.[22] Tento vztah lze volně chápat jako vztah příslušnosti, spjatosti nebo závislosti. Realitu uchopujeme jako funkční, přičemž realitou míníme něco, co považujeme za objektivní existenci. K specifické funkční závislosti se dá dojít dvěma způsoby. Prvním způsobem je, že vybereme jen oblasti reality, kde jsme si jistí, že zde funkční vztah existuje. Tuto jistotu můžeme mít v případě strojů stvořených lidmi, kde byl funkční vztah součástí záměru tvorby stroje, bez ohledu na reálnou míru funkčnosti, která se od původního záměru může lišit. Druhý způsob je charakteristický pro novověkou vědu a techniku. Vybereme oblasti reality, které uchopíme jako funkční tak, že zde funkčnost předpokládáme právě se záměrem uchopení reality, jež by mohlo vést k pravdivému poznání. V tomto případě dochází k něčemu, co lze nazvat zdvojením reality, kdy kromě původní reality vstupuje na scénu ještě realita uchopení (teoretická). Je zde možné najít funkční stavy tím, že je vytvoříme.[23]

Co se týče funkcionalismu v rámci filosofie mysli, jimž se budu zabývat zde, je možné ho definovat jako doktrínu, která duševní stavy považuje nikoliv za závislé na vnitřním prostoru, ale spíše se zaměřuje na jejich roli a způsob fungování ve vnějším systému, jehož jsou součástí. Jinak řečeno, to, co činí něco například myšlenkou, nezávisí na vnitřní konstituci, ale pouze na funkci, kterou zastává v kognitivním systému.[24] Často je uváděno, že je možné si mysl představit jako jakousi černou skříňku, kterou není možné otevřít a jediné, co lze zkoumat,

[22] NOSEK, Jiří. *Funkcionalismus ve vědě a filosofii: sborník příspěvků*. Str. 105
[23] Tamtéž, str. 165-167.
[24] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*

jsou její vstupy a výstupy, přičemž není nutné vědět co v této hypotetické černé skříňce doopravdy je, jediné, co potřebujeme znát, je způsob, jakým skříňka funguje. Funkcionalisté tedy tvrdí, že pojmy, jež užíváme k popisu mentálních procesů, jsou v podstatě funkcionální, jelikož je užíváme spíše s odkazem na funkce chování jejich majitele, než na způsob fungování jeho mozku.[25] Pro lepší přiblížení uvedu příklad, pokud něco označíme třeba za nápad, neříkáme tím, co nápad jako věc je, ale popisujeme tímto slovem funkci něčeho, tedy jak něco jako nápad funguje, proto je mentální sféra pouze předmětem fungování a to, jakým způsobem je zrealizovaná v mozku není podstatné.

Podle funkcionalistické teorie by tedy *„bolest byla stavem, který má tendenci být způsoben tělesným zraněním, produkuje přesvědčení, že je s tělem něco v nepořádku a touhu dostat se z tohoto stavu, vyvolává úzkost a pokud zde není silnější protichůdná touha, tak působí záchvaty nebo sténání.“*[26] Tato funkční role může být vyvolána stimulací určitých vláken nebo neuronů. Potom v tomto smyslu lidé mohou cítit bolest, když je tato část neuronů stimulována. Teorie tedy nevylučuje určitý typ dualismu, protože duševní stavy jsou schopny způsobovat stavy fyzické. I přes původní neutralitu funkcionalismu mezi materialistickým monismem a dualismem, je převážně užíván jako materialistická doktrína, protože je přirozeně pravděpodobnější, že stavy zastávající určitou funkci, schopné hrát nějakou roli budou spíše fyzikálního charakteru.[27] Tento pohled je možné vysvětlit pomocí teorie identity, která ztotožňuje mentální stavy se stavy mozku, tedy s materiální substancí, která je způsobuje. Podle teorie identity jde tedy o stejný druh věci vztažený ke stejnému objektu, mozku.[28]

2.1 Vývoj pojmu funkcionalismus

Funkcionalistické teorie se ovšem mezi sebou mohou lišit. Rozdíly mezi jednotlivými teoriemi je možné odhalit zkoumáním původu funkcionalismu a jeho

[25] PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Str. 247.
[26] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*
[27] Tamtéž
[28] PAVELA, Tomáš. Funkcionalismus ve filosofii mysli. *Inflow*

vývojem, o což se zde v krátkosti pokusím. Nejstarší pohled, který je možné považovat za předchůdce funkcionalismu, je Aristotelova teorie duše. Aristoteles považoval duši za formu přirozeného a organizovaného lidského těla, které je pro duši látkou. Duše umožňovala vyjádřit čestnost, což je plnění funkce, který ji definuje jako druh věci. Potom podobně jako forma sekery je to, co jí umožňuje sekat dřevo, je lidská duše identifikována se schopnostmi, které umožňují lidskému tělu plnit jeho definující funkci. Podle Aristotela je duše úzce spojená s tělem a zahrnuje všechny schopnosti, které jsou nutné k tomu, aby tělo mohlo žít, vnímat, rozumět, jednat a tak dále.[29] Jednoduše řečeno Aristoteles pojetím duše zamýšlí podstatu mentálních stavů vymezit skrze způsob jejich fungování, bez ohledu na jejich vnitřní konstituci.

Druhým pohledem jsou Hobbesovy úvahy o usuzování, které považoval za druhy výpočtů, které vychází z mechanistických principů srovnatelných s pravidly aritmetiky. Usuzování je podle něho v podstatě pouze počítání, tedy přičítání nebo odčítání důležitostí obecných názvů, dohodnutých na označování našich myšlenek. Mimo jednání lidských bytostí podle mechanistických principů, nastoluje Hobbes v Úvodu do Leviathana otázku, proč bychom nemohli připustit, že všechny automaty mají umělý život. Až zhruba v polovině dvacátého století se začalo spekulovat, v duchu Hobbesovské otázky, o tom, že myšlení možná není nic jiného, než výpočet řízený pravidly, který by mohla provádět stvoření různých fyzických typů. Této myšlenky se samozřejmě chytili zastánci umělé inteligence. Například Alan Turing, kterému bude věnována následující podkapitola k funkcionalismu, identifikuje myšlenky se stavy systému, což má mnoho společného se současnými funkcionalistickými teoriemi.[30]

Jako poslední zdroj moderního funkcionalismu uvedu behaviorismus, podle kterého jsou všechna naše tvrzení o duševních stavech a procesech rovnocenná prohlášení o dispozicích chování, přičemž se nejedná o prohlášení, která by byla pozorována pouze introspekcí. To znamená, že pokud někomu

[29] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*
[30] Tamtéž

přisuzujeme určité mentální stavy, pak to děláme v tom smyslu, že mu přisuzujeme dispozice k určitému chování. Tedy pokud někomu přisoudíme například víru, pak spíš než o duševní povaze této víry, mluvíme o pravděpodobnosti, se kterou očekáváme, že se daný člověk bude chovat takovým způsobem, který bychom v konkrétní situaci očekávali. Podle logického behaviorismu by díky výrokům o duševních stavech, které jsou ekvivalentní výrokům o chování, mohl existovat bezproblémový popis toho, jak by mohly být pojmy duševního stavu užity nejen na sebe, ale i pro ostatní. Ovšem proti tomuto pohledu vystupuje řada filosofů (např. Chisholm, Geach, Putnam), kteří namítají, že člověk může mít určité duševní stavy bez příslušných behaviorálních dispozic, a naopak může mít behaviorální dispozice bez duševních stavů. Tak například někdo může věřit, že bude pršet, a přesto si nevezme pláštěnku či deštník, protože mu déšť nevadí, nebo někdo může cítit bolest, aniž by to dával najevo a například herec dokáže předvést velmi přesvědčivě stav bolesti, aniž by bolest doopravdy cítil. Jinak řečeno, žádný duševní stav sám o sobě nemůže být zárukou, že způsobí nějaké zvláštní chování, pokud se také nepředpokládá, že subjekt má další různé duševní stavy.[31]

Vzhledem k těmto třem historickým kořenům rozlišujeme tři hlavní funkcionalistické větve: počítačový funkcionalismus, psychologický funkcionalismus a analytický funkcionalismus. Přičemž je třeba mít na paměti, že se vzájemně překrývají a že existují funkcionalistické teorie, které jsou někde mezi nimi. V následujících řádcích se tyto tři větve pokusím stručně představit. Psychologický funkcionalismus vychází z poznatků a metodologie kognitivních psychologických teorií. Jde o to, že chování je považováno za výsledek souboru duševních stavů a procesů, vnímaných z hlediska role, kterou hrají při tvorbě chování, které se snažíme vysvětlit. Jinak řečeno je funkční povaha mentálních stavů dána jejich vzájemnými vztahy a jejich vnitřním uspořádáním, což se dá empiricky v rámci kognitivní psychologie zkoumat. Vznikl jako reakce na behaviorismus, který odmítá, protože podle něho nelze komponenty naší mysli

[31] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*

definovat pouze na základě vnějšího chování. Souhrnem řečeno lze tento druh funkcionalismu považovat za metodologii kognitivní psychologie, která charakterizuje duševní stavy a procesy jako entity definované jejich rolí v kognitivní psychologické teorii. V tomto smyslu psychologický funkcionalismus považuje za duševní stavy a procesy pouze ty entity, které jsou postulovány vědeckým vysvětlením lidského chování. Proto pokud nějaký duševní stav, který popisuje lidová psychologie, nemá z hlediska kognitivní psychologie vysvětlení, pak se nedá považovat za reálný a tím pádem neexistuje.[32]

Analytický funkcionalismus se snaží poskytovat tematicky neutrální překlady a analýzy našich běžných duševních pojmů. To znamená, že o mysli chtějí mluvit pouze neduševním slovníkem, který nebude obsahovat psychologické pojmy. Snaží se teoretické pojmy popisovat v souvislosti s konkrétními situacemi. Analytický funkcionalismus při popisu kauzálních duševních stavů zohledňuje nejen stimul a reakci, ale také jiné duševní stavy. Souhrnem řečeno zde funkcionalismus stojí na zobecnění týkající se duševních stavů, jejich příčin a jejich společných účinků na chování, které popisujeme pomocí analýzy běžnými pojmy. Podle této teorie je každý mentální stav spojen s aktivitou v mozku, kterou lze pomocí různých přístrojů sledovat. Například bolest se dá identifikovat jako stimulace nějakých určitých nervových vláken. V tomto smyslu sice pojem bolest a nervová stimulace nemají stejný význam, ovšem označují stejný stav mozku.[33]

Počítačový funkcionalismus je možné brát jako podporu různých výpočetních teorií mysli a znovu jako způsob odmítnutí behaviorismu. Jedna z počátečních teorií je stvořena Hilary Putnamem, který teorii založil na myšlence, že mezi lidským myšlením a počítačem je určitá analogie. Podle Putnamovského počítačového funkcionalismu může být každá bytost s myslí považována za Turingův stroj, jehož operace mohou být interpretovány jako sada instrukcí, z nichž každá má určitá pravidla. Tento stroj je možné představit jako

[32] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*
[33] Tamtéž

nekonečně dlouhou pásku, která je dělená pomocí políček, jež skenuje skenovací zařízení. Každé políčko má kapacitu pro uložení symbolu z páskové abecedy, zaplněno je vždy konečně mnoho políček a zbytek obsahuje prázdné symboly. Skenovací zařízení, které se nachází vždy nad jedním políčkem, je schopno toto políčko přečíst, přepsat jiným symbolem a přesunout se o jeden doleva nebo doprava. Stroj funguje na základě vstupů a výstupů, které jsou podmíněny zadanými pravidly. Protože je možné symboly číst, zapisovat nebo se posouvat po pásce, je důležitý také vnitřní stav, ve kterém jsou operace čtení prováděny. Na základě čteného symbolu a vnitřního stavu se určují další akce a přechody do dalších stavů. Stroje tohoto druhu nám jsou schopny poskytnout jednoduchý model toho, jak mohou být vnitřní stavy, jejichž výstupy popisujeme pomocí mechanických procesů, považovány za reprezentace. Program tohoto stroje je zároveň možné spustit na všemožných typech počítačových hardwarů. Myšlenka vnitřních stavů, které se dají popsat mechanicky z hlediska vstupů, výstupů a vztahů mezi nimi se stala jednou z klíčových myšlenek, které se zachovaly v současných funkcionalistických teoriích.[34]

2.2 Alan Mathison Turing

Alan M. Turing se narodil v Londýně roku 1912 a zemřel v Cheshire roku 1954. Vystudoval matematiku a zabýval se obzvláště matematickou logikou, kryptoanalýzou a výpočetní technikou. Velkou část svého života věnoval počítačové vědě a zabýval se myšlenkou inteligentních strojů. Mezi jeho největší úspěchy lze zařadit Turingův stroj, který je popsán již výše. Dále byl za druhé světové války jedním z předních vědců, kteří pomocí šifrovacího stroje dokázali rozluštit tajné německé zprávy, čímž se zasloužili o záchranu mnoha lidských životů. Po válce se Turing proslavil vytvořením známého Turingova testu, který má sloužit jako ukazatel umělé inteligence, kterou zde srovnává s inteligencí lidskou. Jeho základním tvrzením bylo, že počítač, který je správně naprogramován, bude schopný soupeřit s lidským mozkem a tím pádem myslet.[35]

[34] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*
[35] HODGES, Andrew. „Alan Turing“, *The Stanford Encyclopedia of Philosophy*

Tato klíčová myšlenka, která je jasně funkcionalistického charakteru se dá považovat za základní kámen dnešní debaty o umělé inteligenci, která v posledních letech zažívá obrovský boom. Díky Turingovi se do filosofie vrátily klasické filosofické otázky o vztahu mysli a těla, o myšlení jako takovém, o intencionálním, typicky lidském způsobu myšlení a podobně. Zároveň nastolil otázku relativně novou a to otázku, zda můžeme přiznat strojům schopnost myslet ve stejném slova smyslu, v jakém myšlení připisujeme lidským bytostem.[36]

Počátky této myšlenky, kterou zde označuji jako relativně novou, můžeme ovšem pozorovat již u Descarta a jeho Rozpravě o metodě. V této rozpravě Descartes píše:

> *„… kdyby existovaly stroje, které by se podobaly našim tělům a napodobovaly by naše jednání, jak jen by to bylo prakticky (morálně) možné, měli bychom vždy dva velmi jisté způsoby, jak poznat, že to přes všecko vůbec nejsou praví lidé. První způsob je, že tyto stroje by nikdy nemohly užívat slov ani jiných znaků, nemohly by je skládat jako my ke sdělování svých myšlenek druhým. Mohli bychom si sice představit stroj, který je sestaven tak, že pronáší slova, a dokonce že pronáší některá slova odpovídající tělesným úkonům, jež vyvolávají určité změny v jeho orgánech…, ale stroj by nikdy nemohl sestavovat slova různě, aby odpovídala smyslu všeho, co se řekne v jeho přítomnosti, jak to mohou činit i lidé nejzaostalejší. – Druhý způsob je, že stroje by sice konaly řadu věcí stejně dobře, nebo možná i lépe než kdokoli z nás, ale nevyhnutelně by selhaly v některých jiných věcech, kdy by bylo patrno, že nejednají na základě poznání, ale jen podle uspořádání svých orgánů. Neboť zatímco rozum je universální nástroj, kterého lze užít za všech okolností, tyto orgány vyžadují pro každou činnost určité zvláštní uspořádání. Z toho plyne, že je prakticky (morálně) nemožné, aby v jednom stroji existovalo*

[36] PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie*. Str. 232.

dosti různých uspořádání, aby stroj mohl působit za všech životních situací stejně, jako jednáme my za vedení rozumu.[37]

Podle této citace je tedy patrné, že Descartes ve své rozpravě dochází k jinému závěru než Turing. Ten, jak bude později ukázáno, se domníval, že není problém používat slovo myslet v souvislosti se stroji a spíše se zabýval otázkou, kdy je možné říci, že stroje myslí. Descartes byl vůči inteligentním strojům dosti skeptický a pravděpodobně by byl zastával negativní stanovisko i vůči představě, že by nějaký stroj dokázal Turingovým testem projít.

Turingova práce v oblasti umělé inteligence byla nakonec přerušena kvůli jeho zatčení a odsouzení za homosexualitu, která byla v té době ještě považována za trestný čin. Vězení se vyhnul, protože se uvolil podstoupit estrogenovou léčbu, která na něho však měla tragický dopad a vedla k sebevraždě, kterou podle všeho provedl otráveným jablkem.[38]

2.2.1 Imitační hra a Turingův test

Imitační hra a Turingův test úzce souvisí se zmíněnou myšlenkou inteligentních strojů, která má své počátky v komputační teorii. Komputační teorie je jednoduše řečeno představa, že lidská mysl funguje stejně jako počítač. Verzi této teorie předložil Turing v jeho eseji „Computing Machinery and Intelligence" (1950). V tomto článku původní otázku, zda mohou stroje myslet, přepracovává do podoby imitační hry, která nám dovolí obejít diskuzi o podstatě myšlení a stojí pouze na vnějším pozorování. Pokud pomocí vnějšího pozorování usoudíme, že může myslet člověk, pak by pro nás dle Turinga neměl být problém stejným pozorováním přiznat myšlení i stroji, který jedná stejně.[39] Navíc Turing sám se domníval, že to nebude trvat příliš dlouho, než budeme mít digitální počítače, které by si mohly v imitační hře vést dobře.

[37] DESCARTES, René. *Rozprava o metodě: Jak vést správně rozum a hledat pravdu ve vědách str. 43*
[38] HODGES, Andrew. „Alan Turing", *The Stanford Encyclopedia of Philosophy*
[39] Tamtéž

Tato hra spočívá v tom, že si máme představit tři hráče. Hráče A (muže), hráče B (ženu) a hráče C (tazatele, u kterého na pohlaví nezáleží). Tazatel musí být v jiné místnosti než zbylí dva hráči a jeho úkolem je pomocí otázek zjistit, který z hráčů je žena a který je muž. Úkolem hráče A je snaha zmást tazatele a donutit ho k nesprávnému rozhodnutí. Úkolem hráče B je naopak snaha pomáhat tazateli dosáhnout správného rozhodnutí. Aby nedošlo k usnadnění rozhodnutí rozpoznáním hlasů, je rozhovor veden pomocí psacích strojů.[40] Nastává otázka, co by se stalo, kdyby byl hráč A nahrazen digitálním strojem. Pokud by došlo ke změně, odpovídal by tazatel stejným poměrem špatně nebo dobře jako v případě, kdy byli všichni hráči lidé? [41] Turing se domníval, že do padesáti let od jeho eseje bude možné naprogramovat počítač takovým způsobem, že tazatel nebude mít víc než sedmdesátiprocentní šanci rozpoznat člověka po pěti minutách dotazování.[42]

Turingův test tedy nabízí kritérium, podle kterého bychom mohli být schopni posoudit inteligentní chování počítačových programů. Dokonce byla na Turingově testu založena soutěž, nazvaná Loebnerova cena, která běží od roku 1990. Je možné sem přihlásit počítačové programy navržené na simulaci lidské konverzace. Vyhrává program, který nejlépe dokáže napodobit lidský způsob odpovědí po dobu pěti minut. Počítače však stále nejsou na úrovni, která by jim dovolila Turingovým testem projít.[43] To je do velké míry způsobeno tím, že porota, která posuzuje, zda počítač projde, může zabrousit do jakéhokoli tématu. Počítač je většinou úspěšný jen v určitých tématech, protože je velmi těžké vytvořit program, který by se orientoval ve všech věcech, které jsou pro běžného člověka naprosto přirozené.

Přesto zde stále visí otázka, zda počítač, který dokáže člověka účinně klamat takovým způsobem, že člověk nepozná, že mluví s počítačem, a tedy projde Turingovým testem, můžeme označit za myslící stroj. Podle funkcionalistů

[40] HODGES, Andrew. „Alan Turing“, *The Stanford Encyclopedia of Philosophy*
[41] OPPY, Graham and DOWE, David. „The Turing test“, *The Stanford Encyclopedia of Philosophy*
[42] TURING, Alan M. Computing Machinery and Intelligence. *Mind 49.*
[43] OPPY, Graham and DOWE, David. „The Turing test“, *The Stanford Encyclopedia of Philosophy*

nebo také behavioristů by samozřejmě bylo možné říci, že myslí, protože záleží pouze na funkci (nebo chování), která je pozorovatelná a která je stejná jak u člověka, tak u počítače. Podle mnohých filosofů, například Johna Searla, je však jediná věc, kterou počítač dělá to, že myšlení „předstírá." Počítač pouze budí zdání, že myslí, přičemž doopravdy pouze provádí naprogramované mechanické úkony, které mají s myšlením v lidském slova smyslu pramálo společného.

Turing počítal s tím, že jeho tvrzení o potencionálních inteligentních strojích vyvolá v mnoha lidech nevoli či nedůvěru, a proto ve svém článku rovnou předjímá námitky, které by mohly být proti jeho teorii vzneseny. Následně k nim přidává i protiargumenty pomocí kterých se snaží tyto námitky vyvrátit. Dohromady jich uvádí devět, a to námitku teologickou, která vychází z víry v Boha. Námitku hlavy v písku, která je založena na strachu z myslících strojů. Matematickou námitku, vycházející z poznatků v matematice. Dále námitku z vědomí, která stojí na neschopnosti stroje pociťovat emoce a uvědomění si sebe sama. Námitku z různých neschopností, kde je uveden seznam věcí, které stroj nebude nikdy schopen dělat. Námitku lady Lovelaceové, související s neschopností být iniciativní a tvořit nové věci. Námitku ze spojitosti nervové soustavy, kde jde o rozdíl mezi lidskou nervovou soustavou a počítačovým obvodem. Námitku z neformálnosti chování, která se zabývá myšlenkou, že zatímco stroje jsou určovány a omezovány svým programem, lidé mají možnost rozhodování se i v případech, pro které neexistují pravidla. A nakonec námitku z mimosmyslového vnímání, která se opírá o schopnost extrasenzorického vnímání u lidí.[44]

Nejzajímavější se mi zdá námitka z neformálnosti chování a to proto, že tato námitka souvisí se svobodnou vůlí, o jejíž existenci se stále vedou živé diskuze. Existuje mnoho pohledů na problematiku svobodné vůle, jež se povětšinou odvíjí od toho, do jaké míry svobodnou vůli připustíme. Tvrdí zastánci determinismu zastávají názor, že ani lidé svobodnou vůli nemají, a proto je pro ně nemyslitelné svobodné jednání připsat dokonce strojům. Pokud by byl svět

[44] TURING, Alan M. Computing Machinery and Intelligence. *Mind 49*.

opravdu deterministický, pak existují pravidla o tom, co budeme dělat jak pro lidi, tak pro stroje a nezáleží na tom, zda jsou někde sepsaná či ne. Na druhou stranu pokud svět není deterministický a svobodná vůle existuje, pak neexistují žádná nutná pravidla pro lidi, ale stejně tak by nemusela existovat ani pro stroje, protože lidé i stroje mohou podléhat nedeterministickým procesům při tvorbě chování. Není důvod si myslet, že by nebylo možné počítače naprogramovat tak, aby se chovali nedeterministicky. Tento způsob naprogramování by mohl být vystaven na učících se strojích, které Turing navrhuje na podporu svých argumentů.[45] O problematice svobodné vůle se zmíním ještě ke konci diplomové práce.

Turing podotýká, že některé jeho argumenty se mohou zdát slabé a dodává, že problematické námitky se do budoucna vyřeší lepším programováním počítačů a zvýšením úložné kapacity. Tento proces je ale podle Turinga příliš zdlouhavý. Proto navrhuje, abychom se místo snahy o napodobení dospělé lidské mysli, která v sobě nese počáteční stav při narození, vzdělání, kterým prošla, a další zkušenosti, které nejsou způsobené klasickým vzděláním, snažili spíše o simulaci dětských myslí. Takto naprogramovaný stroj by následně prošel vzděláním, podobně jako lidské děti a tím bychom mohli získat učící se umělý mozek. Podle Turinga by mělo být snazší dětský mozek naprogramovat, protože je to menší mechanismus, který se teprve zaplní. Šlo by tedy o dětský program se vzdělávacím procesem.[46]

Je nepochybné, že Turingův filosofický příspěvek byl obrovský a reagovalo na něj velké množství myslitelů. Co se týče kritiky, tak se nejvíce poukazuje na to, že ač počítač projde Turingovým testem, stále to neznamená, že doopravdy myslí. Podle mnohých nestačí jen porovnávat výsledek činnosti stoje a člověka, ale záleží také na způsobu, jakým se k tomuto výsledku došlo. Nehledě na to, že i kdybychom stroji uznali nějakou míru inteligence na základě Turingova testu, pravděpodobně by to neznamenalo, že dokáže obstát v dalších běžných situacích, kde je inteligence nezbytná a které člověk bez problémů zvládá.

[45] OPPY, Graham and DOWE, David. „The Turing test“
[46] Tamtéž

3. John Rogers Searle

John Searle je významným americkým profesorem filosofie, který se narodil roku 1932. Zabývá se převážně filosofií jazyka a filosofií mysli. Vzhledem k povaze diplomové práce se budeme zabývat právě jeho příspěvky k filozofii mysli, kde se zaměřoval na problematiku vědomí, intencionalitu a také umělou inteligenci. Klíčový bude jeho argument čínského pokoje, který formuloval roku 1980, a který měl potenciální možnost umělou inteligence úplně vyvrátit.

Searlovo pojetí mysli a těla, které je popsáno v první kapitole této práce je výchozím bodem k tomu, abychom si vytvořili představu o vztahu mezi mentálními a neurofyziologickými procesy v tom smyslu, jak to myslel Searle. Tedy souhrnem jsou mentální procesy výsledkem chování mozku a současně se v mozku také realizují. Tento postoj je protinázorem k zmíněnému funkcionalismu, který mezi fungováním mozku a počítačem vidí jasnou analogii, tvrdí, že funkce je zde v podstatě stejná, a tudíž je možné sestrojit počítač, který bude schopen myšlení. To v podstatě znamená, že mysl má v mozku stejnou funkci jako má program v hardwaru.[47] V souvislosti s touto myšlenkou Searle uvádí pojmy silná a slabá umělá inteligence. Slabou umělou inteligencí myslí počítače, které nám jsou schopny pomáhat ve studiu mysli tím, že jsou velmi užitečnými nástroji. Kupříkladu nám pomohou tak, že na nich můžeme simulovat duševní schopnosti nebo nám dovolí přesněji testovat vytvořené hypotézy. Ale neumožňují nám tvrdit, že počítače skutečně rozumí a jsou inteligentní. Na druhou stranu podle silné umělé inteligence nejsou počítače pouhými nástroji, sestrojenými jen k užitku člověka, spíše se jedná o vhodně naprogramovaný stroj, jenž je schopen chápat a má i jiné kognitivní schopnosti. Proti slabé umělé inteligenci nemá Searle v podstatě žádné námitky, proto se soustředí hlavně na

[47] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. Str. 29

problematiku silné umělé inteligence, která je představována například Turingem a dále v práci je popisován tento druh inteligentních počítačů.[48]

Problematika, se kterou se snaží Searle v souvislosti se silnou umělou inteligencí vypořádat, lze shrnout následovně. Program zastánců UI často předkládá analogii mezi fungováním mozku a počítače. Mozek je v tomto smyslu spíše jedním z typů hardwarového počítače, který v sobě má program, jenž je zodpovědný za lidskou mysl a inteligenci. Z čehož je možné vyvodit, že mysl není zcela závislá na svém biologickém charakteru. V souvislosti s Turingovým strojem by pak jakýkoliv systém stvořený z jakéhokoliv materiálu, který má správný program s vhodnými vstupy a výstupy, bylo možné považovat za myslící, protože mít myšlenky, znamená mít správný program. Zastánci tohoto pohledu se potom domnívají, že je pouze otázkou času, než se nám podaří vytvořit program shodný s lidskou myslí.[49]

Searle si nemyslí, že silná umělá inteligence je možná a jeho kritika je vedena bez ohledu na stupeň vývoje, ve kterém se digitální počítače nacházejí, protože nevěří ani v možný potenciál budoucích inteligentních počítačů. Kritiku vede na samotnou definici digitálních počítačů, protože je zde podstatné to, že lze jeho operační kroky specifikovat pouze pomocí abstraktních symbolů, tedy naprosto formálně. Programy digitálních počítačů je možné naprogramovat výhradně prostřednictvím syntaktických struktur. Symboly na syntaktické úrovni ovšem postrádají jakékoliv významy nebo sémantické obsahy, a tudíž se nemohou k ničemu vztahovat. To znamená, že i když bude počítač schopen v Turingovu testu užívat slova, tak stejně nebude vědět, co daná slova znamenají, což je ale v rozporu s myšlenkou funkcionalismu, že lidskou mysl a počítačový program je možné považovat za souměřitelné. Protože mít mysl znamená něco více než mít pouhé syntaktické procesy. Procesy naší mysli v sobě mají navíc také obsahy, tedy jsou vždy o něčem a tím pádem musejí mít významy. Jednoduše řečeno má mysl kromě syntaxe také sémantiku, z čehož vyplývá, že žádný počítač nebude

[48] SEARLE, J., 1980, 'Minds, Brains and Programs', *Behavioral and Brain Sciences*,
[49] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. Str. 29- 30.

nikdy schopen myslet, protože jeho programy jsou založené pouze na syntaktickém charakteru.[50]

Proti závěru, že pouze lidská mysl může mít sémantiku, je možné namítat, že i počítač je schopen určitého stupně sémantiky, protože pokud je schopen v testu smysluplně odpovídat, i když pouze na velmi omezený počet témat, pak musí chápat minimálně význam spojek. Tedy kupříkladu musí vědět, že pokud se v souvětí objeví spojka a, pak bude toto souvětí pravdivé pouze v případě, že jsou pravdivá obě tvrzení, která se zde vyskytují. A rozumět této spojce znamená znát její význam. Searle by na to nejspíše odpověděl, že znát významy spojek nestačí k tomu, abychom počítači přiznali sémantiku ve stejném smyslu, v jakém ji mají lidé, protože v jeho případě je to velmi omezená vlastnost, kterou není schopen uplatnit na nic jiného. Jak později ukáži, Searle považuje sémantiku za schopnost, kterou máme díky intencionalitě, která může vzniknout pouze z biologické hmoty.

Výše uvedená tvrzení Searle podporuje myšlenkovým experimentem, který je považován za jeden z nejznámějších argumentů moderní filosofie mysli.

3.1 Argument čínského pokoje

Argument čínského pokoje, publikovaný roku 1980, je postaven jako protipříklad k Turingovu testu, který je podle něho nedostatečný a má jednoduše řečeno dokazovat, že správné naprogramování digitálního počítače může způsobit zdání inteligence a porozumění, ale ve skutečnosti skutečné pochopení neposkytuje.

Myšlenkový experiment je založen na tom, že si máme představit skupinu programátorů, kteří napsali počítačový program simulující porozumění čínskému jazyku. Počítač dostává otázky v čínštině, které následně porovnává se svou pamětí a díky tomu je schopen vytvořit vhodnou čínskou odpověď. Tyto odpovědi jsou přitom stejně dobré jako odpovědi kteréhokoliv rodilého mluvčího čínštiny. Poté si položíme otázku, zda počítač na základě předchozího popisu rozumí

[50] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. str. 30-33.

čínsky stejně jako člověk, který čínsky doopravdy mluví. Následně si místo počítače představme člověka, který vůbec nerozumí čínštině a který je zavřený v místnosti, kde má k dispozici koše, v nichž jsou naskládány čínské znaky a manuál napsaný v jeho rodném jazyce, pomocí kterého je schopen se znaky manipulovat. Díky manuálu bude člověk v místnosti schopen manipulace se znaky na čistě formální úrovni, tedy pouze syntakticky bez schopnosti sémantického pochopení slov. Pravidla pak budou vypadat tak, že člověk vezme znak AB z koše číslo jedna a přiloží ho ke znaku XY z koše číslo dva. Pomocí těchto pravidel bude možné na otázky, které jsou předkládány lidmi venku odpovídat stejně dobře, jako by odpovídal rodilý mluvčí. Tak je pomocí pouhé manipulace se symboly člověk schopen mluvit čínsky, aniž by rozuměl jedinému slovu, a aniž by měl možnost se tímto způsobem čínsky naučit. Výsledkem potom je, že při provozování tohoto formálního programu se z pohledu vnějších pozorovatelů člověk chová tak, jako by čínskému jazyku doopravdy rozuměl, ale jak bylo již řečeno, čínsky vůbec nerozumí. A jestliže čínsky nerozumí člověk, pak nemůže být řeč o tom, že by čínsky rozuměl digitální počítač, protože dělá v podstatě to samé, tedy manipuluje znaky na základě formálních pravidel a nemá nic navíc, co by mu schopnost porozumění dodalo. Z argumentu vyplývá fakt, že k porozumění jazyku nestačí schopnost manipulace s formálními symboly, ale je třeba navíc významu, tedy sémantického obsahu, čehož počítač není schopen.[51]

Proti argumentu bylo vzneseno několik námitek, které Searle vyvrací jako neadekvátní. Kritici například poukazují na to, že existuje více stupňů a úrovní porozumění, tedy že porozumění není pouze dvoumístný predikát. Podle Searla jsou však jasné případy, kdy porozumění platí a kdy neplatí, a to je vše, co je pro platnost argumentu třeba. Přestože někdy říkáme věci typu, počítač ví, jak počítat, říkáme to spíše kvůli tomu, že do artefaktů promítáme vlastní úmyslnost. Je tomu tak proto, že nástroje používáme k dosažení vlastních cílů a je tedy pro nás přirozené, že na ně metaforicky přenášíme atributy úmyslnosti. Jenže smysl,

[51] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. Str. 33-34.

v jakém tyto vlastnosti počítačům připisujeme, je jiný, než jaký připisujeme lidem.[52]

Další myslitelé se domnívají, že jeho myšlenkový experiment bude vyvrácen, pokud se uzná, že čínskému jazyku rozumí celý systém jako celek. V tomto smyslu je člověk jakýmsi centrálním procesorem, který sice čínsky nerozumí, ale dohromady s koši, manuálem a seznamem programů a vším co je k dispozici, čínsky rozumí. Searle námitku odmítá s tím, že ani systém jako celek není schopen posunout se od syntaxe k sémantice. Další námitka je založena na představě, že pokud program čínského jazyka vložíme do robota, který bude schopen chodit po světě a postupně se učit významům, pak není důvod mu porozumění upírat. I na tento argument Searle odpovídá rozdílem mezi syntaxí a sémantikou, protože robot bude stejně mít místo mozku počítač, který tuto propast prostě nedokáže překročit. Tím, že mu přidáme další schopnosti navíc, mu nepomůžeme v ničem, co se týče obecného pochopení a inteligence. Pokud má robot pouhý formální počítačový program, pak mu ani kauzální interakce se světem neumožní získat významy.[53]

Následuje námitka, která se zaměřuje na simulaci skutečných nervových synapsí, které jsou v mozku rodilého mluvčího z Číny. Pokud budeme schopni nasimulovat počítač stejně jako mozek, pak bude také stejným způsobem zpracovávat informace a bude čínštině rozumět. V reakci na to Searle navrhuje variantu, kdy má člověk v pokoji soubor ventilů a vodovodních trubek, které jsou uspořádány stejně jako neurony v mozku rodilého mluvčího a program potom určuje, které ventily se budou otevírat v reakci na vstupy. V tomto případě Searle poukazuje na kauzální působení systému a tvrdí, že v tom nebude žádný rozdíl a čínštině člověk stejně nebude rozumět, protože simulace mozku není skutečný mozek.[54]

[52] SEARLE, J., 1980, 'Minds, Brains and Programs', *Behavioral and Brain Sciences,*
[53] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. str. 35-36.
[54] COLE, David, "The Chinese Room Argument", *The Stanford Encyclopedia of Philosophy*

Další námitka kombinuje ostatní námitky a jde v ní o to, že si máme představit robota s počítačem místo mozku, přičemž jeho „mozek“ bude naprogramovaný stejnými synapsemi, které mají v mozku lidé. Chování tohoto robota by mělo být od chování člověka nerozeznatelné a toto všechno dohromady by mělo fungovat jako jednotný systém, tedy by nešlo o pouhý počítač se vstupy a výstupy. V tomto případě se zdá rozumné připsat robotovi inteligentní chování. Searle reaguje, že ve skutečnosti jsou tyto prvky kombinace předchozích námitek irelevantní, protože to myšlence silné umělé inteligence nijak nepomůže. Je tomu tak proto, že robot stále nebude mít nic jiného než formální programy a nebude tedy ničím jiným než mechanickou figurínou. Jako protiargument uvádějí kritici nemožnost zjištění, zda i ostatní lidé mají mentální vlastnosti, protože tyto vlastnosti vidí pouze jednotlivci sami v sobě. A pokud nám nevadí přisuzovat mysl lidem pouze na základě jejich chování, neměl by to být problém ani v případě počítačů. V tomto případě je však protiargument špatně položený, protože v této diskuzi nejde o to, jak můžeme vědět, že jiní lidé mají mentální vlastnosti, ale jde o to, co lidem připisujeme, když jim ony vlastnosti přiznáváme. A to, co jim přiznáváme, nejsou pouze výpočetní procesy a výstupy, protože toho je schopen i stroj. Jak Searle dodává, tak i v kognitivních vědách lidé předpokládají realitu a znalost mentálních vlastností stejně, jako ve fyzikálních vědách musíme předpokládat skutečnost a poznání fyzických objektů.[55]

Největším nedostatkem výše zmíněných námitek je to, že opomíjejí, co je pro mozkové operace opravdu klíčové. Jejich pozornost se stále upíná k formálním synapsím mozku, které se snaží přenést na uměle vytvořené počítače, ale opomíjejí skutečné vlastnosti těchto synapsí, které nám jsou schopny naši mysl nějakým doposud neznámým způsobem zprostředkovat. Protože podle Searla a nejen podle něho je nám inteligentní chování dáno chápáním obsahů, nikoliv formy.

[55] SEARLE, J., 1980, ‘Minds, Brains and Programs’, *Behavioral and Brain Sciences,*

3.2 Může počítač myslet?

Než je možné na otázku o inteligentních strojích odpovědět, je nutné uvést, v jakém smyslu ji Searle pokládá. Protože v určitém smyslu je možné říci, že i lidé jsou stroje, a to ve významu, že stroj je fyzikální systém, který je schopen provádět určité operace. V tomto případě je potom možné říci, že stroje myslí. Proto bychom měli otázku specifikovat na stroje vytvořené člověkem, jenže i zde zůstává prostor pro inteligentní stroje, protože bychom mohli vytvořit stroj, který je na molekulu stejný jako člověk, a tedy by nejspíše mohl myslet. Proto otázku formuluje do podoby, zda může myslet digitální počítač. Ovšem i v tomto případě je důležité zůstat opatrný, protože jako digitální počítač se dá popsat v podstatě všechno, co realizuje nějaký počítačový program. Tím pádem je dle funkcionalistické tradice i lidský mozek v tomto smyslu digitální počítač, protože je schopen provádět určité množství programů. Proto je nutné, aby všichni měli na paměti, že Searle mluví o digitálním počítači, který pomocí správného naprogramování se vstupy a výstupy bude dostatečný k tomu, abychom o něm mohli říci, že skutečně myslí. Na což odpovídá jednoznačným ne, protože syntax nikdy nebude stačit na sémantiku a jediné čeho je počítač schopen, je napodobování myšlení, stejně jako napodoboval porozumění člověk v čínském pokoji.[56]

Ze všech Searlových myšlenek je možné pomocí čisté logiky dojít k závěru, který nám jasně ukazuje, že podle něho počítače myslet nemohou. Searle píše:

> *„Ať už bychom stvořili jakýkoli artefakt mající mentální stavy ekvivalentní mentálním stavům, které má člověk, nemohli bychom si vystačit s počítačovým programem. Spíše by tento artefakt musil mít síly ekvivalentní kauzálním silám lidského mozku.„*[57]

[56] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. Str. 37-38.
[57] Tamtéž. 43.

Z tohoto závěru vyplývá fakt, že Searle zdůrazňuje mentální stavy, které jsou přísně biologického charakteru a obsahují takové věci, jako je vědomí, intencionalita, subjektivita a mentální kauzace, které jsou z jeho pohledu popsány v první kapitole. Proto zde zůstává možnost pro inteligentní stroje, které by ovšem musely být stvořeny nikoliv z libovolného materiálu, ale pouze, stejně jako člověk, z materiálu biologického. Potom je otázka, zda by se toto stvoření mělo nazývat strojem v tom smyslu, v jakém běžně stroj myslíme. A pokud bychom ho nenazvali strojem, ale kupříkladu klonem, pak je zde téma myslících strojů zbytečné.

Searle sám považuje pouhou myšlenku na myslící počítače za absurdní, a proto se ptá, jak je vůbec možné, že podobné úvahy někdo může brát vážně. Vše podle něho vzniklo z toho, že zastánci umělé inteligence neberou mysl jako část biologického světa, ale snaží se ji vnímat stejně formálně, jako je tomu u počítačového programu, což je vzhledem k jejich odmítnutí dualismu zvláštní paradox.[58]

3.3 Intencionalita

Searle pro ujasnění svého konceptu filosofie mysli uvádí, jakým způsobem si vysvětluje lidské jednání, které tak odlišuje od „jednání“ počítače. Důležité je, že v lidském jednání nejsou zahrnuty pouze fyzické pohyby, ale má v sobě také způsob popisu, který při interpretaci lidského jednání běžně užíváme, proto lidé vědí, co dělají a nepotřebují k tomu fyzické pozorování. Tyto principy, pomocí kterých lidé jednání vysvětlují, musejí být potom součástí tohoto jednání.[59]

Klíčovým pojmem pro vysvětlení lidského chování je pojem intencionalita. Pokud má mentální stav intencionalitu, pak to stručně řečeno znamená, že je o něčem. Přitom nejde jen o intenci jako o záměr, tak jak se běžně užívá, protože záměr je v Searlově smyslu stejně jako například přesvědčení nebo naděje jen součástí intencionality jako celku. Pro intencionální stav je důležitý

[58] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. str. 40.
[59] Tamtéž, str. 64.

obsah, který dává podmínku toho, že je stav vždy o něčem a také psychický typ, protože ten samý obsah se může objevovat v různých typech. Pro ilustraci různých psychických typů uvádí příklad, například člověk může chtít vyjít z pokoje, věřit, že vyjde z pokoje a zamýšlet vyjít z pokoje. Pokaždé jde o stejný obsah, ale ten je v jiných psychických typech, v přání, přesvědčení nebo záměru. Obsah a typ mentálních stavů nám slouží k propojení mysli a světa, to znamená, že s mentálními vztahy dokážeme různými způsoby reprezentovat svět, tak jak je, jak ho chceme mít a podobně. Přesvědčení, která máme, jsou potom pravdivá, pokud odpovídají stavu světa, to znamená, že intencionální stavy mají podmínky splnění a každý stav sám určuje, za jakých podmínek je pravdivý. Mentální stavy mají i propojení příčin a následků, protože pokud si něco přejeme a následně to provedeme, pak je příčina reprezentací stavu věcí, které způsobuje. Tyto vnitřní vztahy nazývá Searle intencionální kauzalitou.[60]

Intencionalita má tedy tři základní vlastnosti, jejím základem je určitý obsah v určitém psychickém typu. Intencionální stavy stanovují podmínky svého splnění, tedy jejich úspěšnost záleží na tom, jestli svět odpovídá obsahu daného intencionálního stavu. A intencionální stavy jsou příčinou určitých následků tak, že se pomocí kauzace tvoří stav věcí, které intencionální stavy reprezentují. Pomocí tohoto základu, který byl předložen, je vytvořen základ pro vysvětlení lidského jednání. To Searle popisuje v několika principech. První princip říká, že jednání je součtem mentálního a fyzického komponentu, to znamená, že pokud něco provádíme fyzicky, tak to doprovází i určitá mentální činnost, která se však může objevovat i samostatně, pokud například jen ležíme a přemýšlíme. Druhý princip navazuje s tím, že mentálním komponentem je jistý záměr, tedy má intencionalitu a je o něčem. Mentální komponenta v případě úspěchu, způsobuje intencionální kauzalitou fyzický komponent. Komponenty jsou běžně vedle sebe a jsou na sobě také závislé. Třetí princip mluví přímo o intencionální kauzalitě, která je podstatná jak pro strukturu, tak pro vysvětlení jednání. Pozoruhodné na intencionální kauzalitě je, že mentální stavy iniciují události, jejichž výsledek je

[60] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. str. 64-65.

stavem věcí reprezentovaný mentálním stavem, jenž tuto událost vyvolal. Čtvrtý princip rozlišuje mezi jednáním, které je dopředu promyšlené, to znamená, že uvažujeme o tom, čeho chceme dosáhnout a jak toho chceme dosáhnout a jednáním, které je spontánní, a tedy bez předchozího rozmyslu. Proto rozlišuje primární záměry, které máme před samotným jednáním a záměry v jednání, které máme při průběhu realizace samotného jednání. Tyto dva způsoby rozlišuje kvůli tomu, že teorie jednání předpokládají, že jakékoli lidské činnosti jsou až posledním článkem, kterému předchází řada praktického usuzování, což v mnoha případech není pravda, protože často jednáme, aniž bychom o tom předtím dlouze přemýšleli. Záměr máme i ve spontánním jednání, ale nebyl před jednáním promýšlen, spíše jde o záměr samotného jednání. Pátý princip s tím souvisí a říká, že tvorba primárních záměrů je ve většině případů výsledkem praktického usuzování, při kterém musíme volit mezi přáními, které jsou v konfliktu. Přání jsou pro většinu lidí motivující silou, a proto jsou v jednání důležitým prvkem. Dodává, že praktické usuzování, které hledá nejlepší rozhodnutí s ohledem na naše rozdílná a konfliktní přání, je jiné než usuzování teoretické, při kterém jde v první řadě o pravdu. [61]

Tyto principy ukazují, jakým způsobem Searle vidí obraz intencionální kauzality. Jde o jakousi mentální energii, která iniciuje naše jednání. Pro lepší obraznost uvádí ještě tři principy, které lépe vysvětlují záměrnost jednání. Princip šest říká, že:

> *„Vysvětlení jednání musí mít stejný obsah, jaký měl jednající na mysli při jednání nebo při rozmýšlení, z něhož vzešel záměr provést jednání. Je-li vysvětlení skutečným vysvětlením, pak obsah, který skrze intencionální kauzalitu vyvolal jisté chování, musí být identický s obsahem vysvětlení tohoto chování.“*[62]

Díky tomu je možné odlišit jednání od dalších přirozených událostí našeho světa a stejně tak se liší i jejich vysvětlení. Pokud vysvětlujeme lidské chování,

[61] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994, str. 67-72.
[62] Tamtéž. Str. 72.

tak jak příčina, tak vysvětlení mají jeden obsah a vysvětlení je opravdu vysvětlením pouze za předpokladu, že má ten samý obsah jako příčina. Vzhledem k tomu, že záměry se nám do hlavy nedostávají samostatně, vysvětluje v sedmém principu sítě intencionálních stavů, ve kterých funguje každý intencionální stav jen jako část sítě jiných intencionálních stavů. To znamená, že podmínky splnění se určují ve vztahu k velkému množství jiných intencionálních stavů. Na základě toho dochází k tomu, že aktivita naší mysli není složena jen z mentálních stavů, ale skládá se také z toho, jak s nimi operujeme, tedy ze schopností, dovedností, návyků, postojů ke světu a podobně. I když máme nějaký záměr, například jet autem, pak to lze uskutečnit pouze za předpokladu, že to umíme, tedy máme nějakou dovednost, což intencionální stav není. V tomto případě se liší vědění jak a vědění že. Celé to shrnuje v posledním principu, který je o tom, že soubor těchto typů věcí, jako jsou dovednosti, schopnosti a podobně, nazývá Searle pozadím intencionality, protože na jejich základě intencionální stavy fungují, a přitom to nejsou mentální stavy.[63]

3.3.1 Daniel Dennett

V souvislosti se Searlovou intencionalitou je vhodné zmínit i Daniela Dennetta, který s tímto pojmem také pracuje. A ač se oba autoři dají zařadit pod větev monistického materialismu, je jejich přístup v mnoha ohledech rozdílný. Searle vnímá vědomí jako východisko k vysvětlení intencionality, to znamená, že jeho způsob interpretace vede od první osoby ke třetí osobě, od vědomí k intencionalitě. Dennett oproti tomu postupuje opačným směrem, z pohledu třetí osoby, tedy od intencionality k vědomí. Dalším rozdílem je, že Dennett zaujímá intencionalitu spíše ve funkcionalistickém postoji, to znamená, že obsahy přisuzuje na základě jejich funkčních rolí.

Dennett považuje intencionální postoj za náš způsob interpretace chování nějaké entity, kdy tuto entitu vnímáme, jako by byla racionální. Tedy jako by své chování podřizovala úvahám o svých přáních a přesvědčeních. Tento postoj běžně

[63] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994. str. 72-74.

zaujímáme k ostatním lidem a někdy také z antropomorfického hlediska k entitám, jejichž chování si nejsme schopni vysvětlit jiným způsobem. Vedle intencionálního postoje uvádí ještě fyzikální a projektový postoj, které užíváme k vysvětlování toho, co se v přírodě kolem nás děje. [64]

Fyzikální postoj běžně používáme ve fyzikálních vědách. Díky němu jsme schopni předvídat na základě fyzikálních zákonů a fyzikálních stavů daných systémů, k nimž se postoj vztahuje. V podstatě pro každou fyzickou věc, která se chová podle určitých fyzikálních zákonů, je možné užít předpověď a vysvětlení v podobě fyzikálního postoje. Ovšem co se týče lidských artefaktů, tak si mnohdy nemůžeme být jejich fyzikálními zákony a vlastnostmi jistí a v tu chvíli přichází na řadu projektový postoj. Ten je postaven na předvídání způsobu, jakým byl tento artefakt vyprojektován. Jinak řečeno, neznáme a ani nepotřebujeme znát fyzikální vlastnosti předmětu, stačí nám vědět, k čemu byl vyprojektován a jakým způsobem můžeme ovládat jeho funkce. Samozřejmě musíme brát v potaz, že předmět může být vyprojektován špatně, a tudíž svou funkci, kterou předvídáme, plnit nebude. V tomto ohledu je projektový postoj o něco riskantnější než postoj fyzikální, ale přesto ho denně užíváme, protože se v drtivé většině případů dobře vyplácí. Kromě artefaktů vyprojektovaných lidmi, se z projektového postoje dají předvídat i živé artefakty stvořené Matkou přírodou. Příkladem je sázení rostlin, kdy nás nezajímá, jak je dáno to, že rostou. Jediné na čem záleží, je naše zkušenost s tím, že pokud dáme semena do země, tak za několik měsíců vyrostou.[65]

Co se týče riskantnosti tak právě intencionální postoj je považován za nejvíce nejistý. Lze ho popsat jako podmnožinu projektového postoje zaměřenou na konatele. Z intencionálního postoje se dá vysvětlit jak chování ostatních osob, tak chování složitých artefaktů. Příkladem uvádí šachový počítač, kterému připisujeme racionální chování, založené na znalosti pravidel šachu a touze vyhrát. Počítačové chování je potom pro nás jednodušší vysvětlovat právě takto,

[64] DENNETT, C. Daniel. *Druhy myslí: k pochopení vědomí*. Str. 33-34.
[65] Tamtéž, Str. 34-36.

než s použitím předchozích dvou postojů, protože dokážeme předvídat tahy, které se z jeho pozice zdají být nejracionálnější, pokud chce vyhrát. Klíčové potom je, že nezáleží na tom, zda počítač skutečně „chce“ vyhrát, stačí nám, že pomocí intencionálního postoje dokážeme jeho kroky úspěšně předvídat. Již zmíněný antropomorfismus je v případě intencionálního postoje patrný v tom, že se snažíme toto předvídání jakoby vidět z vlastního pohledu, tedy jak bychom se zachovali my v pozici daného organismus nebo artefaktu, což je užitečné i pro další vědecké bádání.[66]

Pomocí intencionálního postoje potom připisujeme mysl i jiným než lidským entitám, protože jak píše Dennett *„…aby se mohly předvídat její činy, musíme předpokládat, že je konatelem chytrým, protože hloupý konatel by mohl udělat jakoukoli hloupost. Tento smělý krok, spočívající v našem předpokladě, že konatel bude jednat chytře…, je tím, co nám umožňuje předpovědi činit.“*[67] Z toho je již patrný funkcionalistický přístup, protože Dennett přisuzuje mysl všemu, co je nějakým způsobem fungující. Přičemž funkcí je myšlena schopnost zpracování informací, čehož je schopen i počítač. Z jeho pohledu jde samozřejmě o jiný druh mysli, ovšem stále je to mysl. Podle Dennetta tedy můžeme připustit, že počítač myslí v tom smyslu, že k němu zaujmeme intencionální postoj, a tedy budeme předpokládat, že je racionálním konatelem stejně jako člověk.

Dennett poukazuje na to, že Searle intencionalitu používá ne příliš šťastným způsobem. Existují totiž dvě varianty intencionality, vlastní a odvozená. Vlastní intencionalita je zjednodušeně řečeno o tom, že naše mentální obsahy jsou o něčem a odvozená se od této vlastní odvozuje v tom smyslu, že si „bere“ naše myšlenky.[68] To si lze představit tak, že pokud myslím na něco konkrétního, pak se jedná o intencionalitu vlastní, pokud však vezmeme v úvahu myšlenku, která je někde napsaná (v nějakém reprezentačním artefaktu, například v knize), pak jde již o intencionalitu odvozenou z původních myšlenek tvůrce. To je přístup, který

[66] DENNETT, C. Daniel. *Druhy myslí: k pochopení vědomí*. Str. 36-39.
[67] Tamtéž, str. 40.
[68] Tamtéž, str. 54.

Searle zastává. Jenže problém, který v tom Dennett vidí je ten, že pokud artefakty získávají intencionalitu díky aktivitě jejich tvůrce, pak tento tvůrce má úplně stejnou intencionalitu v mysli, z které artefakt stvořil. K vysvětlení používá evoluční vývoj, ve kterém je i mozek artefaktem stvořeným Matkou přírodou. Lidská intencionalita je potom stejně odvozená jako intencionalita vnějšího artefaktu.[69]

V rámci evoluce pomocí přirozeného výběru se tak u lidských bytostí postupně vyvinula mysl a s ní také intencionalita. Nejprve tu byly velmi jednoduché organismy, které soupeřily o zdroje spolu s jinými organismy. Jejich snaha přežít nebyla nijak plánovaná, spíše jen tak bloudily, dokud narazily na zdroj potravy, tedy o mysli zde zatím nemůže být řeč. Z těchto původních primitivních organismů zcela závislých na vnějším prostředí přežily jen ty, které se byly schopny rozmnožit. Organismy pak postupně získávaly různá vylepšení, protože se dokázaly flexibilněji přizpůsobovat situacím v jejich přirozeném prostředí. Adaptace na prostředí probíhala metodou pokus omyl, dokud ti lépe vybavení nepřišli na nejlepší možný způsob prospívání. Tvorové, kteří se z těchto lépe vybavených vyvinuli, pak již byli o něco chytřejší a špatné modely chování hned vynechávali. Z modelů, které nebyly katastrofální, pak dokázali před-vybírat pomocí vnitřního prostředí, kde za pomocí nasbíraných informací o světě zjišťovali, která taktika bude nejvýhodnější v daném prostředí. V tomto smyslu se již dá mluvit o inteligenci, protože si dokáží dopředu promyslet, která taktika povede k nejlepšímu výsledku, například podle okolností rozhodnou, zda je lepší před nepřítelem utéct nebo na něho zaútočit. Do této skupiny patří savci, ptáci, ryby nebo plazi.[70]

Poslední druh tvorů, které v rámci evoluce Dennett zmiňuje a kam zařazuje také lidské bytosti, jsou potom tvorové, jež jsou schopni užívat různé nástroje. Mezi tyto nástroje patří i mysl, která nám umožnila užívat jazyka, což zase přispívá k velkému zkvalitnění vnitřního prostředí a k možnosti učit se od

[69] DENNETT, C. Daniel. *Druhy myslí: k pochopení vědomí*. Str. 56.
[70] Tamtéž, str. 82-90.

ostatních. Schopnost užívání nástrojů tak musela být doprovázena velkým nárůstem inteligence.[71]

Z výše zmíněného tedy můžeme vyvodit, že vlastní intencionalita v Dennettově přístupu vůbec nemusí existovat. Intencionalita totiž vznikla v rámci evoluce jako adaptace organismů v jejich boji o přežití. Člověk, který je pak pomyslným vrcholem této evoluce, je taktéž přírodou vyprojektovaný tvor, tudíž se i jeho intencionalita odvozuje z přírody. Pomocí intencionálního postoje je schopný předvídat chování jak ostatních lidí, tak také artefaktů, jež sám vyprojektoval. A pokud počítač plní funkce, které považujeme za inteligentní, pak není problém mu určitý druh mysli přiznat.

Podle tohoto evolučního postoje je také možné vyvrátit Searlovu tezi: Syntax nestačí na sémantiku, protože pokud je člověk pouze konečným článkem v dlouhé evoluci, pak je jeho mysl pouze výsledkem procesu, který na počátku fungoval jako velmi jednoduchý organismus, jehož „mysl“ byla pouze formální soubor těch nejjednodušších pravidel. Sémantika se tak stejně jako mysl a intencionalita postupně vyvinula v průběhu evoluce.[72]

To ve mně vyvolává otázku, zda by nebylo možné vytvořit také umělou evoluci počítačů nebo spíše robotů, protože pro nás je přirozenější přiznat myšlení něčemu, co vypadá více jako člověk. Představme si tedy robota, který bude, stejně jako v námitce proti argumentu čínského pokoje, schopen chodit mezi lidmi, hovořit s nimi a postupně získávat informace o tom, co určitá slova znamenají v interakci s předměty, se kterými se v přírodě setká. Bude tedy schopen například říci „toto je pes,“ pokud jeho vizuální sensory zachytí psa. V tomto bodě ještě spolu se Searlem můžeme tvrdit, že robot má pouze syntax a pravé významy slov mu unikají, tudíž muže říci „toto je pes,“ ale význam slova pes ve skutečnosti nechápe. Pak ale robotovi naprogramujeme schopnost vytvořit dalšího robota, kterého bude sám „učit“ stejným způsobem, jako se učí malé děti. Tento nový

[71] DENNETT, C. Daniel. *Druhy myslí: k pochopení vědomí*. Str. 95-97.
[72] COLE, David, "The Chinese Room Argument", *The Stanford Encyclopedia of Philosophy*

robot vytvoří dalšího a tak stále dokola. Není v tomto případě myslitelné, že by se objevila nová vlastnost, která by robotům dovolila pochopit významy, a tedy získat mysl, která bude souměřitelná s tou lidskou? Protože pokud se v přirozené evoluci mohla objevit mysl jako výhodná adaptace, pak by se i zde mohlo objevit něco navíc. Searle by s tímto samozřejmě nesouhlasil, protože takový robot není biologického charakteru, jenž je nutný k tomu, aby se mysl mohla objevit.[73]

Souhrnem řečeno jsou oba pohledy v určitých částech velmi silné. Dennettovo evoluční vysvětlení vzniku mysli a sémantiky Searlovu koncepci překonává, protože Searle sice vysvětluje mentální stavy naturalisticky, ovšem o jejich vzniku nic neříká. Dennett tím navíc přesvědčivě překonává Searlovu tezi o syntaxi, která na sémantiku nestačí. Na druhou stranu Dennett až příliš ochotně přiznává mysl komplikovanějším artefaktům, jejichž způsobu fungování nerozumíme. To by mohlo vést k tomu, že mysl přisoudíme stroji pouze proto, že se nám tak může jevit, aniž by nějakou mysl doopravdy měl, na což nás upozorňuje Searle jeho myšlenkovým experimentem. Proto si myslím, že Dennett sice dokázal některé Searlovy teze vyvrátit, ovšem nedokázal se vypořádat s argumentem čínského pokoje.

3.4 Reakce autorit k argumentu čínského pokoje

Přesto, že je argument čínského pokoje Johna Searla jedním z největších filosofických úspěchů, který je stále těžké vyvrátit a jenž je mnohými považován za filosoficky nesmrtelný, snaží se ho mnozí myslitelé zpochybnit anebo přinejmenším upravit tak, aby bylo možné mysl umělým strojům v nějakém smyslu přisoudit.

Jedním z těch, kdo na Searlovu koncepci argumentu čínského pokoje reagoval, je Terry Winograd, který se pouští do kritiky samotného základu otázky myslících počítačů, která je podle něho bezvýznamná, protože je zbavená vhodného kontextu. Winograd souhlasí s tím, že umělá inteligence není schopná

[73] Nevím, jestli podobnou otázku již někdo nezformuloval, ovšem zde je to popsáno pouze z mého pohledu.

pojmout lidské kognitivní schopnosti a že počítače nejsou schopny těmto schopnostem porozumět a oplývat jimi. Nesouhlasí ovšem s argumenty, které Searle pokládá jako důvod, proč tomu tak není. Otázky, které si Searle pokládá, podle něho totiž nejsou koherentní, a tudíž na ně nelze objektivně odpovědět. Je totiž chybné předpokládat, že na otázku, zda počítač rozumí jazyku, můžeme odpovědět kladně nebo záporně. Chyba přitom není v logice argumentu, ale je mnohem elementárnější, protože je v samotném základu našeho chápání pravdivosti a nepravdivosti. Searle se drží dvoumístného predikátu porozumění, tedy že v objektivním smyslu dokážeme poznat případy, kdy X rozumí nebo nerozumí Y. Podle Winograda ale potřebujeme porozumět porozumění jiným způsobem, protože náš jazyk není připraven na nové otázky, se kterými jsme doposud neměli zkušenost. Musíme s ohledem na okolnosti a na naše zkušenosti zvažovat pro a proti. Potom v určitých případech je možné připustit, že něco je v určitém ohledu pravdivé a v jiném zase nepravdivé. Máme sice vrozenou intuici, že mechanické věci nejsou schopny rozumět, kdežto lidé oplývající inteligencí rozumět mohou. Na druhou stranu je pro nás počítač stále ještě nejednoznačnou entitou, pro kterou nemáme vymezené přesné termíny. Navíc termín porozumění užíváme v běžném jazyce v mnoha různých významech a perspektivách.[74]

Proto Winograd tvrdí, že *„…každý význam je inherentně kontextuální a otevřený, ale v mnoha každodenních případech je kontext tak pevně sdílen a je považován mezi účastníky za samozřejmost, že existuje iluze objektivity. Jazyk je vždy neúplný, částečný a nejednoznačný odraz zkušeností. Proto odpověď na otázku "Může počítač rozumět?" není jednoznačné „ano“ nebo „ne“, ale závisí na pozadí předpokladů a interpretací…“*[75]

Na otázku, kdy je vhodné říci, že někdo nebo něco rozumí, odpovídá Winograd, že je to spíše věc sociální. Jak píše, tak mnohdy se přistihneme, že zvířatům nebo mechanickým věcem přisuzujeme mentální stavy, což je pro nás efektivní způsob, jak popsat jejich „chování“, když nerozumíme procesům, které

[74] PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence.* Str. 80-82.
[75] Tamtéž, str. 83.

se v nich odehrávají. V tomto případě je nám jasné, že se nejedná o skutečné porozumění počítače, ale spíše o metaforu. Termín porozumění tedy v těchto případech aplikujeme relativním způsobem vzhledem k pozadí a kontextu.[76] K podobnému závěru ostatně došel také sám Searle, který ale v souvislosti s tím upozorňuje na to, že si máme dát pozor na doslovné chápání našich metafor.

Dalšími mysliteli, kteří reagují na Searla jsou Herbert A. Simon a Stuart A. Eisenstadt. Jejich myšlenka je taková, že počítač může být naprogramovaný takovým způsobem, že rozumí přirozenému jazyku. A to v tom smyslu, že dokáže jeden jazyk přeložit do jiného, dokáže odpovídat na otázky, jež mu byly v přirozeném jazyce předloženy a dělat úkony, které by v případě, že by je vykonával člověk, byly přesvědčivě považovány za důkaz, že člověk tomuto přirozenému jazyku rozumí. Svou tezi tito dva myslitelé nazývají empirickou silnou umělou inteligencí (Empirical Strong AI; SAI-E), kterou předkládají jako alternativu k Searlově logické silné umělé inteligenci (Logical Strong AI; SAI-L).[77]

Pro další práci, je nejprve důležité nastínit definici termínu porozumění, tak jak ho užívají oni, protože termín je klíčový pro rozhodování, zda je hypotéza pravdivá či nepravdivá. Z jejich pohledu je porozumění definováno jako porozumění v obvyklém smyslu slova, které používáme, když chceme popsat chování člověka a stejná kritéria budou vztahovat i pro přítomnost porozumění v chování počítače. Jejich testování přítomnosti porozumění je proto stejné jak pro lidi, tak pro počítače. To jestli člověk, potažmo počítač, něčemu rozumí v tomto smyslu, je podle nich empirická otázka, kterou potom není možné logickým způsobem vyvrátit. Proto je argument čínského pokoje ve své podstatě jakýsi trik, který nás může zavést do zmateného vnímání porozumění.[78]

Chtějí získat empirické důkazy, že počítače jsou schopny v určitých situacích porozumění a to přinejmenším stejného porozumění jakého jsou schopni

[76] PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence.*, str. 84-85.
[77] Tamtéž, str. 95.
[78] Tamtéž, str 95-96.

lidé. Proto nejprve hledají důkazy o porozumění u lidských bytostí. Vzhledem k současné úrovni technologie, v hodnocení lidského porozumění, jsme omezeni pouze na pozorování vnějšího lidského chování v určitých situacích. Proto používají externí testy také k tomu, aby zjistili, jestli i počítače mají porozumění. Následně používají i interní testy, které jsou dostupné pro počítače, ale nejsou dostupné pro lidi a tím určí povahu duševních stavů, které toto porozumění umožňují. Takto mají v plánu získat empirickou podporu pro teorii toho, jak jsou duševní stavy člověka v určitých situacích podobné úrovni informačního zpracování, aniž by bylo třeba úrovně neuronů, které jsou doposud těžko pozorovatelné. Příklad takového testu je požádat osobu, aby uvedla jméno nějakého konkrétního objektu, procesu nebo vztahu mezi objekty, přičemž jsou tyto věci dobře viditelné. Pokud tázaná osoba dokáže odpovědět správné jméno, dojdeme k závěru, že osoba je v určitém duševním stavu porozumění tohoto jména, tedy chápe, jaký objekt, proces nebo vztah označuje jménem. Přičemž míra jasnosti porozumění může být dále testována.[79]

Tyto empirické testy o porozumění jazyku, ač mohou být omylné a nepřesné, lze aplikovat jak na lidské bytosti, tak na počítače. Ovšem jak bylo již řečeno, tak v případě počítačů se můžeme podívat i dovnitř, kde můžeme pozorovat, co se kde a jak při aplikování testů na přítomnost porozumění děje. To znamená, že můžeme zkoumat části jejich programů a pamětí, při vykonávání testů. Navíc v případech, kdy programátor neposkytuje testy, jež definují porozumění, můžeme objevit proces učení, který je počítačem používán k získání těchto testů. Tento aspekt učení je klíčový, protože jinak by se dalo namítat, že počítač nemůže mít porozumění, pokud je spojení otázek a odpovědí přímé, bez toho, aby bylo zprostředkované zájmy. A právě toto zprostředkování probíhá během procesu učení.[80] Simon a Eisenstand uvádějí pro demonstraci tři počítačové programy, na nichž můžeme pozorovat „*(1) zda počítač používá*

[79] PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence.*, str. 97-98.
[80] Tamtéž, str. 99.

porozumění k plnění úkolu a (2) zda získá jeho porozumění programováním nebo učením.“[81]

Prvním programem je EPAM a byl původně sestrojen k simulaci lidského chování v experimentech verbálního učení. V průběhu času mu byly přidávány další věci jako například paměťové schopnosti. Tento program se dokáže učit rozlišovat mezi podněty, kategorizovat podněty a spojovat páry podnětů. Druhý program je ZBIE, který je schopen reprezentace jednoduchých scén, které pomocí jednoduchých vět dokáže popsat. Program je schopen naučit se, jak spojovat objekty, procesy a vztahy se slovy a větami v přirozeném jazyce. To znamená, že pokud se ukáže nová scéna, je schopen ji popsat větou přirozeného jazyka. Poslední program CaMeRa dokáže tvořit něco jako obraz mysli v diagramu, který je prezentován jako mapa (podobně jako obraz na sítnici). Obraz přetvoří například do linií a pak pomocí průsečíků spojuje slova s obrazem, čímž ho popíše. Všechny tři programy přitom lze podrobit jak vnitřnímu, tak vnějšímu empirickému testování, a proto je Simon a Eisenstand považují za podporu pro svou tezi SAI-E,[82] protože na základě této empirické evidence ukazují, že počítače jsou schopni porozumění a učení lingvistických struktur detekovaných pomocí sensorů.[83]

Úplně jiným způsobem reagují na Searlovu práci Selmer Bringsjord a Ron Noel, kteří se naopak snaží myšlenku čínského pokoje podpořit. To činí tím, že vyvrací moderní přesvědčení zastánců UI, že námitka robota je přeci jenom platná, což je podle nich dáno stále se zlepšujícími roboty, kteří jsou stále chytřejší.[84]

Jedním z těch kdo navrhl modernizovanou námitku robota je Stevan Harnad, který uznává, že Searle dokázal vyvrátit Turingův test tím, že ukázal, že stroj může projít testem bezmyšlenkovitým manipulováním se symboly bez

[81] PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence.*, str. str. 100.
[82] Tamtéž, str. 100-101.
[83] Tamtéž, str. 106.
[84] Tamtéž, str. 144-145.

nutnosti jakéhokoli porozumění těmto symbolům. Proto navrhuje nový argument, nazvaný Total Turing Test (TTT), který je podle něho lepší, protože vyžaduje, aby robot dokázal odpovědět na všechny naše vstupy. Tudíž mu k úspěšnému absolvování testu nebude stačit schopnost pouhých textových odpovědí. Aby toho byl robot schopný, musíme mu navíc přidat senzomotorické schopnosti. To by mělo zajistit, že robot získá nejen nový způsob, jak odpovídat na otázky, například týkající se pohybových aktivit, ale také těchto nových aktivit bude sám schopen.[85]

Harnad navrhuje, abychom se vrátili do čínské místnosti a tentokrát použili měřítko TTT. Místo abychom se ptali, zda kandidát (v tomto případě sensomotoriky schopný robot), procházející testem skutečně chápe čínštinu, nebo ji pouze formálně popisuje, a tudíž pouze budí zdání porozumění, zeptáme se, zda kandidát skutečně vidí například sochu, nebo ji pouze formálně popisuje, tedy jako kdyby ji viděl. Harnad se domnívá, že ať bude odpověď jakákoliv, tak to povede k popření argumentu čínského pokoje. Tento argument totiž odpovídá pravidlu odvození, které se nazývá konstruktivní dilema. Podle tohoto pravidla, pokud je tvrzení p pravda (robot se opravdu dívá na sochu) pak r (argument je vyvrácen) a pokud je tvrzení q pravda (robot pouze popíše výstup optického snímače) pak r. Obě možnosti tak vedou k popření argumentu. Bringsjord a Noel však argumentují, že jediné co se změnilo, je to, že porozumění nahradíme viděním, to ale neznamená, že robot ve skutečnosti vidí stejně jako by danou věc viděl člověk. Můžeme si „vidění“ robota představit následujícím způsobem. Představme si, že před nás někdo položí úplně neznámý předmět, u kterého nemáme tušení, k čemu slouží nebo co má představovat. Předmět sice vidíme, ale nemáme žádné porozumění, které by se k němu vázalo. Tímto způsobem vidí robot, a proto můžeme Harnadův argument označit za neplatný.[86]

[85] PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence.,* str. str. 151.
[86] Tamtéž, str. 151-153.

Pro lepší vysvětlení argumentu čínského pokoje uvádějí Bringsjord a Noel zombie zvířata, která jsou založena na myšlence zombie argumentu,[87] který velmi stručně předpokládá, že stvoření zombie jsou logicky možná ovšem fyzicky nemožná. V případě zombie zvířat předvádějí, že jsou možná i fyzicky. Mělo by se jednat o velmi jednoduché zvíře na úrovni mnohobuněčného organismu, které když zaznamená světlo, pohání ho to vpřed. Podle úrovně intensity světla se pohybuje buď rychle, nebo pomalu. Tedy v případě sestrojení takového jednoduchého robota je chování takové, že čím více je senzorem detekován zdroj světla, tím rychleji běží motor. Následně představují dvě složitější biologické bytosti. První z nich se snaží světlu vyhnout a druhá se chová agresivně, jako by světlo chtěla napadnout. Postupem času bude možné vytvořit zombie zvířata, která budou reagovat na více podnětů najednou a budou tudíž stále složitější.[88]

Předpokládejme, že sestrojíme dostatečně sofistikovaného robota, který projde Harnadovým TTT, ale který je založený na stejném principu jako zombie zvířata a jediné co má navíc, je složitý program, který vykonává rozhodnutí o toku informací do a z mozku robota. Následně si představme člověka, kterého budeme postupně předělávat, aby fungoval stejně jako takový robot. Nejprve ho zdrogujeme, aby neměl žádný hmat a nemohl chodit. V případě robota existuje nějaký realizátor, který je zodpovědný za jeho schopnost chůze pomocí signálů, a proto člověku operací aktivujeme nohy pro chůzi na základě podobných signálů. Potom bude schopen chodit, aniž by měl nad chůzí kontrolu. To samé můžeme udělat i s ostatními realizátory, takže například způsob, jakým robot pohybuje rukou, naoperujeme i člověku. Vizualizaci u člověka vytvoříme tak, že ho oslepíme, ale buňky vnímající světlo budou dále fungovat podobně jako snímače a budou přenášet stimuly do mentálních obrazů. Takže se bude schopen hýbat na základě vizualizace mentálních obrazů. Takový „člověk“ by každého přesvědčil o tom, že je schopen věci vidět, chápat a podobně, aniž by ve skutečnosti byl

[87] Argument Zombie bude konkrétněji v práci uveden v poslední kapitole jako příklad obecné kritiky funkcionalismu.
[88] PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence.,* str., str. 157-162.

schopen vnímání. Bringsjord a Noel považují tento myšlenkový experiment za konečné vyvrácení námitky robota.[89]

Stevan Harnad později znovu reagoval a označil Searla za původce nedorozumění, které se k jeho argumentu čínského pokoje váže, protože podle něho neuvádí věci přesvědčivým způsobem. Kupříkladu Searlova definice silné umělé inteligence, ke které se vyhrazuje, je problémová z toho důvodu, že takhle ji zastánci UI nepřijímají. Snaží se tedy definici přeformulovat tak, aby lépe odpovídala principům komputační teorie mysli: [90]

Searlovy tři principy silné UI jsou:
První princip: Mysl je počítačovým programem.
Druhý princip: Mozek je irelevantní.
Třetí princip: Turingův test je rozhodující.
Harnadova upravená verze:
První princip: Duševní stavy jsou pouhým uskutečněním správných počítačových programů.
Druhý princip: Výpočetní stavy jsou na uskutečnění nezávislé, neboli jinak řečeno, software je nezávislý na hardwaru.
Třetí princip: Neexistuje silnější empirický test na přítomnost duševních stavů než Turingův test, a proto je rozhodujícím testem pro komputační teorii duševních stavů.[91]

Z prvních dvou principů pak podle Harnada vyplývá, že počítačový program musí být fyzicky uskutečněn jako dynamický systém, aby se mohl stát výpočetním stavem, přičemž fyzické detaily uskutečnění jsou pro tento výpočetní stav, který uskutečňuje, irelevantní. Realizace a nezávislost jsou totiž pro komputační teorii důležité, protože podle této teorie jsou duševní stavy výpočetními stavy, které jsou sice fyzicky realizovány, ale mentalita se nehledá v samotné věci (hardwaru), nýbrž v počítačovém programu (softwaru). Tudíž se

[89] PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence.* Str. 162-164.
[90] Tamtéž, str. 297.
[91] Tamtéž, str. 297-299.

vyhýbáme struktuře a zůstává nám pouze funkce. Harnad podotýká, že tento důležitý princip Searle nijak explicitně nevysvětluje, a proto jeho původní znění vede spíše ke zmatení. A nově formulovaný třetí princip nám nyní dovoluje interpretaci, že i když počítač Turingovým testem projde, tak to nutně neznamená, že mysl mít musí, a naopak pokud neprojde, tak to nemusí znamenat, že mysl nemá. Z třetího principu vyplývá pouze fakt, že doposud nemáme žádný lepší empiricky založený test.[92]

Harnad si myslí, že takto předělaná definice silné UI má větší šanci na vytvoření počítače, který by mohl být považován za chápající stroj. Protože duševní stavy stroje bude možné uskutečnit nezávisle na realizaci výpočetních stavů. Harnad přesto uznává, že existuje mnoho způsobů, jak tuto premisu odmítnou, a přijetí některého z nich by znamenalo potvrzení Searlových závěrů. Searle by samozřejmě nepřipustil myšlenku, že komputační teorie je pravdivá, protože pokud člověk zavřený v čínském pokoji není schopen porozumět znakům, se kterými operuje, pak toho nebude schopen ani počítač. A ani veškeré ad hoc námitky, které byly postupně vymýšleny, nejsou schopny tento argument věrohodně vyvrátit.[93]

Harnadova práce nám každopádně ukazuje zajímavou věc a to fakt, že argument čínského pokoje nemusí být dokonalým nezvratitelným důkazem, že jiné entity nemohou mít mysl, protože jak již bylo v práci několikrát řečeno, nikdy nemůžeme mít jistotu o schopnosti myslet ani v případě jiných osob, ale je jedinou věrohodnou predikcí, která je založená na tom, co víme.[94]

[92] PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence.* Str. 297-299.
[93] Tamtéž, str. 300-301.
[94] Tamtéž, str. 301-303.

4. Kritika funkcionalismu

Kritika funkcionalismu se většinou zakládá na tom, že mentální stavy nejsou pouhou funkcí, ale mají také nějakou podobu, něčím jsou. Jedním z problémů, se kterým se funkcionalismus musí vypořádat, je problém kválií. Tato kategorie duševních stavů je i pro zastánce funkcionalismu těžká na odmítnutí. Veškeré funkcionalistické teorie se snaží charakterizovat mentální stavy v souvislých kauzálních podmínkách, taková charakterizace však nedokáže zachytit kvalitativní charakter zkušenostních stavů, nedokáže říci, jak základní vlastnosti jsou, jak se nám kvalitativně jeví.[95]

S problémem kválií souvisí myšlenkový experiment, který nám dokazuje, že kvalitativní vlastnosti nelze považovat za funkce. Jedná se o experiment s obrácenými kválii. Máme si představit člověka, který má obrácené vnímání červené a zelené barvy. To znamená, že zelené předměty u něho vyvolávají vjemy, které jiní lidé mají při pohledu na červené předměty. Člověk, který takovým způsobem vnímá, se pak v rámci jazyka naučí slovo zelená užívat na zelené předměty a slovo červená na červené předměty. Z čehož vyplývá, že jeho zelený vjem má úplně stejnou funkci jako u ostatních lidí červený vjem, přičemž se ale tyto jevy očividně kvalitativně liší.[96]

4.1 Ned Block

S tímto myšlenkovým experimentem souvisí námitka nepřítomnosti, která tvrdí, že mohou existovat tvorové, kteří jsou funkčně ekvivalentní lidem, ale jejich mentální stavy nemají žádný kvalitativní charakter. Tato námitka je rozpracována Nedem Blockem, který uvádí myšlenkový experiment s čínským národem, potažmo s homunkulárním mozkem[97]. Blockova kritika funkcionalismu je reakcí na Putnamovu teorii počítačového funkcionalismu, který se domníval, že každá bytost s myslí může být považována za Turingův stroj. Myšlenkový experiment

[95] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*
[96] PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie.* Str. 247.
[97] Tento experiment je považován za předchůdce argumentu Čínského pokoje, jenž byl představen Johnem Searlem a je popsán v předchozí kapitole.

spočívá v tom, že si máme představit člověka, jehož zevnějšek je podobný jako u ostatních lidí, ale tento člověk se liší tím, že v hlavě nemá mozek. Místo mozku má v hlavě skupinu malinkých lidí, které nazývá homunkuly. Každý z těchto malých lidí představuje jedno políčko strojové pásky, která člověka popisuje. Vstupy, které dostává za smyslových orgánů, jsou napojeny na světýlka umístěná v lebce a pohyb je umožněn řízením pák. Každý z těchto človíčků má na starost vždy jednu operaci, takže když tělo dostane nějaký konkrétní smyslový vjem, v hlavě se rozsvítí konkrétní světlo, na které reaguje konkrétní človíček, jenž současný stav těla přepíše zatáhnutím za konkrétní páku. Takto jednoduché úkoly nevyžadují u homunkulů velkou inteligenci a zároveň zde není nutné, aby navzájem věděli o činnostech, které provádějí ostatní. Každý pouze dělá jeden konkrétní úkol, což stačí k tomu, aby to zvenčí vypadalo jako chování běžného člověka. Z čehož vyplývá, že chování homunkulárního člověka je funkčně stejné jako chování běžných lidí.[98]

Block dále navrhuje upravenou verzi experimentu. Protože je v mozku přibližně miliarda neuronů, používá ve své úvaze čínský národ, který se stal funkcionalistickým a rozhodl se dokázat, že lidská mysl je možná i mimo mozek. Každý Číňan je vybaven vysílačkou, díky které je ve spojení s ostatními lidmi a také s umělým tělem, stejně jako v předchozím experimentu. Jednoduše řečeno jsou homunkulové nahrazeni obyvateli Číny, kteří zastávají stejnou činnost. I zde je možné říci, že tento systém je nám funkčně ekvivalentní, i když jen po určitou dobu. Takový systém by sice byl mnohem více v nevýhodě, co se týče různých vnějších vlivů, jako je například povodeň, zemětřesení nebo třeba ztráta signálu a spojení, ale náš mozek může také podléhat různým poškozením, které by vedly ke špatnému vnímání vstupů, což by zapříčinilo celkově špatné fungování systému. Blockovi je v souvislosti s experimentem vytýkáno, že jeho čínský systém by pracoval příliš pomalu. Na to Block odpovídá, že časová škála není tolik důležitá, protože je lehké si představit naše duševní operace zpomaleně, a přesto bychom neřekli, že to již nejsou duševní operace. Stejně tak jsou myslitelná

[98] BLOCK, Ned. *Troubles with Functionalism*. str. 278

stvoření, která by byla na první pohled nehybná a nějaké reakce by byly vidět pouze skrze různé časosběrné fotografie nebo zrychlený videozáznam. Časové měřítko Block označuje za „hrubě behavioristické“ a námitce se vyhýbá tak, že si ponechává experiment alespoň v metafyzické rovině. Metafyzická možnost je to, co jeho argumentu stačí k prokázání protipříkladu k funkcionalismu. Protože u obou verzí myšlenkového experimentu zůstává pochybnost, zda mají nějaké mentální stavy a hlavně pak, zda mají kvália.[99]

Block se pro názornost odvolává na článek „Jaké to je být netopýrem?“ od Thomase Nagela, který zde udává, že ač popíšeme fungování mozku a nervové soustavy u netopýra, stejně nebudeme nikdy vědět, jak netopýr doopravdy myslí, tedy jaká je jeho mysl zevnitř.[100] Počítačový funkcionalismus totiž podléhá potížím s duševními stavy ve smyslu, pokud se stav M rovná stavu P, pak má někdo stav M pouze, když prožívá stav P. Mentální stavy tuto kauzalitu, ale nesplňují, protože, jak zdůrazňuje Block a také Fodor, lidé se velmi často nacházejí ve více než pouze jednom psychologickém stavu. Například můžeme si něco přát a zároveň něčemu věřit, čehož Turingův stroj schopen není, protože se může nacházet vždy pouze v jednom stavu. Funkcionalita tedy nemá způsob na uspokojivé zacházení s obsahem mezi duševními stavy. Protože podle počítačového funkcionalismu má být kvalitativní stav totožný se strojovým stavem, jenže pokud zde chybí tento kvalitativní stav pak je jeho existence pochybná a tím pádem je zde pochybnost i o strojovém stavu a o pravdivosti funkcionalismu.[101] Experiment s homunkulární hlavou je tak v první řadě protiargumentem k počítačovému funkcionalismu.

4.2 David J. Chalmers

Dalším kritikem funkcionalismu je David Chalmers, který ke své kritice užívá mimo jiné argumentu zombie, jenž byl popisován více autory. Zombie jsou obecně v rámci filosofie mysli vymyšlená stvoření, která mají pomoci vysvětlit

[99] BLOCK, Ned. *Troubles with Functionalism*., str. 279-281.
[100] PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie. Str. 248*
[101] BLOCK, Ned. *Troubles with Functionalism*., str. 281-285.

problémy týkající se vědomí a jeho vztahu k fyzikálnímu světu. Argument staví na tom, že zombie jsou od těl lidí nerozeznatelní, co se týče jejich funkčního uspořádání a fyzické struktury na molekulární úrovni, ale nemají na rozdíl od člověka žádné vědomé zkušenosti, nemají kvalitativní vědomí. Počátky myšlenky zombie je možné nalézt již u Descarta, který považoval zvířata za nemyslící automaty fungující na principu fyzikálních mechanismů.[102] Člověk byl něčím více právě díky mentálním schopnostem, které mu dodává nesmrtelná duše. Myslící zombie nebo jiný stroj, jež vypadá a chová se a jako člověk byl pro něho nemyslitelný, protože jak bylo zmíněno již v kapitole věnující se funkcionalismu, by se od člověka dal lehce rozeznat pomocí neschopnosti souvisle a srozumitelně užívat jazyk a celkově by selhával v oblastech, kterých je schopen člověk právě na základě myšlení.[103] Argument zombie byl ovšem postaven tak, aby byl možný nebo myslitelný. Právě myslitelnost podle Chalmerse znamená metafyzickou možnost, z toho je patrné, že svůj argument staví na intuici. A pokud jsou zombie metafyzicky možní, pak to znamená, že funkcionalismus je nesprávný.[104] Chalmers podobně jako Block zdůrazňuje problematiku nefyzikálních kválií, která jsou osobám dána jako vlastnosti mozku.

Proti tomuto argumentu lze namítnout verifikovatelnost vět, navrhovanou především pozitivisty. Podle pravidla verifikace je věta smysluplná pouze v případě, že lze její pravdivostní hodnotu ověřit. Pokud věta ověřit nelze, pak se v tomto smyslu považuje za bezvýznamnou, tedy žádné tvrzení založené na metafyzice nemůže být pravdivé. Ovšem zůstává zde ten problém, že naše schopnost přemýšlet a mluvit o různých zkušenostech je sama o sobě ve své podstatě neověřitelná, a proto je tato námitka slabá.[105]

Na myšlenkový experiment reagoval Daniel Dennett, který myšlenku zombie považuje za absurdní. Ve svém článku The Unimagined Preposterousness

[102] KIRK, Robert. „Zombie“, *The Stanford Encyclopedia of Philosophy*
[103] DESCARTES, René. *Rozprava o metodě: Jak vést správně rozum a hledat pravdu ve vědách str. 43*
[104] KIRK, Robert. „Zombie“, *The Stanford Encyclopedia of Philosophy*
[105] Tamtéž

of Zombies, popisuje tento argument dokonce jako nejzábavnější chybu v současném myšlení. Podle něho každý, kdo připouští myslitelnost zombie, podceňuje představivost a končí s představováním něčeho, co je proti jejich vlastní definici. Navíc dodává, že spoléhání se na intuici se považuje za kontroverzní. Stejně tak samotné kvalitativní stavy jsou podle Dennetta zmatené a neodpovídají skutečnému vnitřnímu fungování našeho kognitivního systému. To je dáno především tím, že si nemyslí, že by ve vědomí existovalo něco nad rámec funkčnosti a fyzičnosti. Na podporu své kritiky představuje stvoření „zimbo." Tato stvoření jsou schopná reflektovat své vlastní činnosti, včetně svých vlastních vnitřních stavů, které jsou ovšem nevědomé (podle argumentu zombie). Zimbo je také zombie, akorát je složitější v tom, že má schopnost sebe-reprezentace, je tedy nevědomou bytostí se schopností vyšších řádů myšlenek. Dokáže například věřit, že je vědomý, i když ve skutečnosti vědomý není. Zastánci, kteří věří, že argument zombie je koherentní, by měli také akceptovat možnost zimbo, tedy musí připustit pomyslné kategorie různých zombie, což by ukázalo buď rozporuplnost jejich konceptu, nebo by museli přiznat, že jejich koncept vědomí není udržitelný.[106]

Chalmers na Dennettovu kritiku dále reagoval tím, že rozhodujícím faktorem je fakt, že jsme s našimi zkušenostmi obeznámení, to znamená, že na ně můžeme odkazovat. Vzhledem k tomu, že zombie tyto zkušenosti nemají, jsou jejich kvazi – fenomenální úsudky neoprávněné. Chalmers se domnívá, že i kdyby na naše soudy kvália vliv neměla, stačí jejich pouhá přítomnost ve fyzickém kontextu, aby zajistila, že naše myšlenky jsou o těchto kváliích.[107] Což ovšem podle mého názoru dostatečně nevysvětluje hypotetickou možnost Dennettových stvoření zimbo, protože pokud si zimbo myslí, že je vědomé, pak stejně tak lidé mohou mít pouhou iluzi těchto kválií.

Joseph Levine k tomu poznamenal, že i kdyby pravděpodobnost zombie nevedla k prokázání falešnosti funkcionalismu, minimálně otevřela vysvětlující

[106] DENNETT, C. Daniel. The Unimagined Preposterousness of Zombies.
[107] KIRK, Robert. „Zombie", *The Stanford Encyclopedia of Philosophy*

mezeru, která ukazuje, že kvalitativní charakter zkušenosti nelze odvodit z jakéhokoli fyzického nebo funkčního popisu. Funkcionalisté na to namítají, že naše neschopnost vidět způsob uzavření mezery může spíše jednoduše odrážet hranice naší současné schopnosti teoretizovat a i když se nám zombie mohou zdát myslitelné, stejně je tak nakonec nebudeme brát kvůli nárůstu empirických poznatků.[108]

Problematika teorie zombie je poměrně rozšířená a názory na ni se liší, podnítila plodnou práci na poli vztahů mezi představivostí, myslitelností a možností a pro teorie zabývající se vědomím nadále představuje velkou výzvu. Pokud je teorie koherentní, pak nemůže být o myslících strojích řeč, protože pouze lidský mozek má vlastnosti jako je vnímání kválií a podobně.

4.3 Frank Jackson

Jako poslední kritický argument proti funkcionalismu uvedu argument znalostí, který je také jedním z nejvíce diskutovaných argumentů, které proti funkcionalismu a fyzikalismu existují. Cílem argumentu je opět dokázat, že vědomá zkušenost zahrnuje i nefyzikální vlastností. Hlavní myšlenkou potom je, že i když máme úplné fyzické znalosti o jiné bytosti, která se považuje za vědomou, stále to neznamená, že máme také dostatečné znalosti o tom, jaké to je mít stejné zkušenosti jako ona bytost. Pro ilustraci znovu uvedu T. Nagela a jeho netopýry, kde Nagel tvrdí, že určitá fakta lze zachytit pouze ze subjektivní perspektivy. Protože i když víme z objektivní perspektivy všechno o tom, co to znamená lovit pomocí sonaru, stále bychom nemohli vědět, jaké to je pomocí tohoto sonaru vnímat. Tedy úplné poznání fyzických znalostí není dostatečné pro poznání fenomenálních stavů.[109]

Tuto myšlenku použil jako první Frank Jackson, který ji podporuje příkladem neurofyziložky Marie. Marie je výbornou vědkyní, která svou

[108] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*
[109] NIDA-RÜMELIN, Martine. „Qualia: The Knowledge Argument“, *The Stanford Encyclopedia of Philosophy*

vědeckou práci tvoří v černobílé místnosti pomocí černobílé televize. Specializuje se na neurofyziologii zraku a postupně získává veškeré informace o tom, co se děje, když někdo vidí červené rajče a použije výraz červená. To znamená, že je schopna pozorovat, jaké vlnové délky z rajčete stimulují sítnici a jaký to má dopad na centrální nervový systém, který způsobuje vokální akordy a vylučování vzduchu z plic, což vede k pronesení věty „rajče je červené". Jenže pak je Marie z černobílé místnosti propuštěna a náhle se o světě naučí něco nového, zjistí, jaké to je vidět červenou barvu. Z toho nutně vyplývá, že její předchozí informace nebyly úplné, ale přitom měla všechny informace fyzického i funkčního charakteru. Tedy máme víc než jen fyzické informace a funkcionalismus, stejně jako fyzikalismus, nemůže být pravdivý.[110]

Někteří kritici argumentu reagovali tím, že Marie bude schopna poznat barvy, pokud je poprvé spatří, pomocí svých úplných fyzických znalostí o barevném vidění. Jenže pokud se poprvé setká s konkrétní barvou, tak získává úplně nový fenomenální koncept dané barvy a vytvoří si přesvědčení, že to je to, co vidí ostatní. Tedy její předchozí znalosti nemohly být úplné. Proto kritici jako obranu proti těmto argumentům navrhli hypotézu schopností (Ability Hypothesis), která říká, že vědět, jak co je, třeba jako vidět červenou barvu, je pouze praktickou znalostí. Čili je to spíše svazek schopností, jakési „know-how", než znalost tvrzení nebo faktů. Například Lewis tvrdí, že vědět, jak zkušenost je, je vlastnost schopností pamatovat si, rozpoznávat, představovat si věci a podobně. Není to vědění že, ale pouze vědění jak.[111]

Na to se argumentovalo tím, že schopnost představovat si, že máme zkušenosti určitého druhu, není dostatečné proto, abychom poznali, jaké to je určitou zkušenost doopravdy mít. Například Conee a Alter se domnívají, že imaginativní schopnosti nejsou dostatečné. Argumentují v tom smyslu, že si máme představit osobu, která se doposud nesetkala s třešňově červenou barvou.

[110] NIDA-RÜMELIN, Martine. „Qualia: The Knowledge Argument", *The Stanford Encyclopedia of Philosophy*
[111] Tamtéž

Této osobě povíme, že je to barva někde mezi vínově červenou a ohnivě červenou, přičemž vzhledem k těmto informacím má možnost si představit, jak bude barva vypadat, ale dokud barvu skutečně neuvidí a nezíská ji zkušeností, nebude vědět, jaké to je ji vidět.[112]

Alternativním a rozšířenějším pohledem mezi funkcionalisty, jak se s argumentem znalostí vypořádat je to, že poznaní toho, co jak je (jako vidět červenou barvu), jsou ve skutečnosti osvojené výrokové poznatky, jež jsou nám dány zkušeností a jsou vyjadřovány pojmy v první osobě. To podle nich není pro funkcionalismus problém, protože tyto pojmy v první osobě nemusí představovat jako způsoby prezentace, jakkoliv neredukovatelné kvalitativní vlastnosti.[113] Obrana by vypadala nějak takto: kvália jsou fyzikální vlastnosti zkušeností (za předpokladu, že uznáme, že zkušenosti jsou fyzické procesy), tedy Marie může vědět ještě před propuštěním z černobílé místnosti všechno o kváliích a o tom, že dané věci kvália mají, i když s nimi ještě není seznámena. Po propuštění se s kválii seznámí, ale nezískává novou věc z propozice znalostí tím, že se s nimi seznámí. Tento názor se opět setkává s podobnými problémy jako předchozí hypotéza, protože seznámení s kválii je nezbytné pro to, abychom mohli vědět, co to je mít s nimi zkušenost. Jenže skutečnosti, které tyto nové poznatky dělají pravdivými, jsou fyzikálního charakteru a Marie je věděla ještě před propuštěním. Filosofové se většinou shodují v tom, že je těžké popřít, že Marie má po propuštění nové poznatky, které předtím mít nemohla, jenže tento pocit může být pouze iluzorní.[114]

4.4 Problém eliminativismu

Jako poslední a z mnoha pohledů nejradikálnější obrana funkcionalismu, který se převážně chápe jako materialistická doktrína, existuje námitka eliminativismu, kterou představuje například Sellars, Quine, Rorty nebo Dennett.

[112] NIDA-RÜMELIN, Martine. „Qualia: The Knowledge Argument“, *The Stanford Encyclopedia of Philosophy*
[113] LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy*
[114] NIDA-RÜMELIN, Martine. „Qualia: The Knowledge Argument“, *The Stanford Encyclopedia of Philosophy*

Eliminativismus je tvrzení, které říká, že naše běžné chápání mysli je špatné a že některé nebo dokonce všechny duševní stavy, ve skutečnosti vůbec neexistují. Za klasický argument podporující eliminativismus se považuje Sellarsova teze, že používáme teoretický rámec pro vysvětlování a předvídání inteligentního chování. Tato pozice považuje teorii, kterou používáme za lidovou psychologii, u které se předpokládá, že je tvořena ze zobecnění toho, co každodenně popisujeme psychologickými pojmy, jako je například víra. Lidová psychologie nám ale podle nich neřekne nic o skutečné podstatě mozku a její principy popisují kognitivní procesy nesprávným způsobem. Z tohoto postoje je patrné, že lidová psychologie nepopisuje nic, co je skutečné. Podle nich totiž neexistuje nic, co by se nedělo v mozku a obyčejné duševní stavy nelze redukovat na neurologické procesy. Duševní stavy se dle tohoto pohledu nedají redukovat ne proto, že by nebyly fyzické, ale proto, že duševní stavy, které jsou popsány psychologií v běžném slova smyslu, ve skutečnosti neexistují. Eliminativisté prohlašují, že různé pojmy lidové psychologie se nakonec ukáží jako prázdné a neodpovídající tomu, co skutečně existuje. Proto mohou být naše myšlenky považovány za funkčně definované teoretické pojmy a funkcionalismus tím pádem lze považovat za pravdivý. Podle eliminativismu se potom dá jednoduše popřít, že existují takové věci, jako jsou například kvália a tvrdí, že přesvědčení o tom, že existují, je pouhou iluzí nebo zmateností. Například Dennett mluví o tom, že naše kvalitativní stavy jsou subjektivní, a tedy soukromé, a že se jejich vnímání může mezi jednotlivci v určité míře lišit. Tímto způsobem naznačuje, že naše kvalitativní koncepce jsou natolik zmatené, že nemohou odpovídat našemu skutečnému vnitřnímu fungování kognitivního systému.[115]

Proti tomuto pohledu bylo namítáno, že eliminativismus je v jistém slova smyslu sebezničující, protože jakákoliv teorie vyžaduje existenci víry, víry v to, že je pravdivá. Víra je však podle eliminativismu považována za psychologický pojem, jež ve skutečnosti fyzikálně neexistuje a je spíše prázdným pojmem lidové psychologie. Bez ohledu na tuto vnitřní rozporuplnost se dá také namítat, že

[115] RAMSEY, Wiliam. „Eliminative Materialism“ “, *The Stanford Encyclopedia of Philosophy*

v rámci lidové psychologie používáme výroky, které se vztahují k mnoha věcem, které mají velký význam. Úspěch, který lidová psychologie má, naznačuje, že nabízí přesnější popis duševních procesů, než jaký jí eliminativisté přisuzují. Pokud pomocí zdravého rozumu předvídáme chování lidí, pak máme poměrně vysoký stupeň úspěchu, což nám poskytuje argument pro závěr, že lidová psychologie poskytuje lepší vysvětlení než eliminativismus. Odpověď eliminativismus na tuto námitku je, že i když se nám teorie zdá úspěšná, nemusí být v posledku pravdivá, protože se obecně stává, že určitým populárním teoriím přisuzujeme větší váhu, než si ve skutečnosti zaslouží a ignorujeme anomálie, které pak považujeme za bezvýznamné. Podle nich jsme slepí k nedostatkům, které lidová psychologie představuje.[116] Problémem potom je, že pokud uznáme eliminativismus, pak zde nebude žádné místo například pro morálku nebo svobodnou vůli, protože by zde neexistoval žádný rozměr lidské psychologie, který by odpovídal našemu pojetí morálky a svobody.

4.4.1 Problém svobodné vůle

S problémem svobodné vůle se setkáváme v průběhu celé diplomové práce, a proto se pokusím tuto problematiku shrnout ze Searlova pohledu. Jak říká Searle, tak představa lidské svobody je zásadní pro naše pojetí sebe sama, a přesto není stále prokázána. Přeci jen se nám zdá přirozené, že je příroda složená z částic a vzájemných vztahů, pomocí nichž je možné všechno vysvětlit. Z tohoto pohledu zde prostor pro svobodnou vůli není. Co se týče lidské svobody, existovaly v minulosti mnohé pokusy se s ní vypořádat. Proti existenci svobodné vůle stojí determinismus, který je například zajímavým způsobem představován Laplacem, který navazoval na Newtonovu fyziku. Popisuje hypotetickou osobu, většinou označovanou jako Laplaceův démon, která by v určitém okamžiku znala polohu veškerých existujících částic a zároveň také všechny jejich pohybové zákony. Tato osoba by mohla předvídat budoucnost a zároveň také znát minulost celého univerza, tedy svobodná vůle by zde své místo neměla. Ve prospěch svobody vůle hovoří lidská zkušenost, že naše rozhodnutí a myšlenky mají vliv na naše chování.

[116] RAMSEY, Wiliam. „Eliminative Materialism“ “, *The Stanford Encyclopedia of Philosophy*

To znamená, že něco z nějakých důvodů děláme, ale víme, že jsme se mohli rozhodnout i jinak. Lidské chování nelze předvídat, protože se lidé mohou sami rozhodovat, co udělají a co ne. Vyplývá z toho fakt, že lidé sice na jedné straně uznávají vědecké argumenty, které nás nabádají věřit, že ve světě svobodná vůle není, ale na druhou stranu z vlastní zkušenosti věříme, že svobodná vůle existuje.[117]

Proto existuje směr nazývající se kompatibilismus, podle kterého je determinismus se svobodnou vůlí slučitelný. Svět je pokládán za deterministický, ale lidská rozhodnutí jsou založená na svobodné vůli, čímž se ovšem nepopírá jejich determinovanost, protože i lidské chování podléhá fyzikálním silám, které působí na vše, tedy i na lidské tělo. Spíše se tím myslí, že k lidskému jednání nedochází pod nátlakem a můžeme se tedy svobodně rozhodovat. Searle ovšem kompatibilismus jako vysvětlení svobodné vůle nepovažuje za dostatečné, protože pravou podstatu problému ve skutečnosti neřeší.[118] Podle Searla totiž *„…nejde o to, zda vedle vnějších fyzikálních příčin a vnitřních nutkání jsou při způsobování našeho chování ve hře ještě jiné vnitřní 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.“*[119] V tomto smyslu není ani v kompatibilismu pro svobodnou vůli v běžném slova smyslu dostatek prostoru.

Dále vedle kompatibilismu uvádí také mentální determinismus. Svobodné jednání, které běžně známe z našich zkušeností, totiž také nemusí být ve skutečnosti svobodné, protože není jisté, zda se na své zkušenosti můžeme úplně spoléhat. Jako příklad uvádí hypnotické stavy, kdy má zhypnotizovaný člověk pocit, jako by jednal svobodně, ačkoli jeho pozorovatelé vědí, že ve skutečnosti jedná tak, jak mu bylo řečeno. Jak ovšem dodává, tak je velmi nepravděpodobné, že by veškeré lidské jednání probíhalo pod něčím jako hypnózou a musíme naše vysvětlení svobodné vůle vztahovat k tomu chování, které je pro lidi běžné a nesmíme se soustředit na takovéto výjimečné případy chování, jako je chování

[117] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994, str. 93-95.
[118] Tamtéž, str. 95-96.
[119] Tamtéž, str. 96.

pod hypnózou. Na druhou stanu máme někdy psychické nutkání nějakým způsobem jednat a mentální determinismus zastává názor, že právě tyto psychické příčiny určují naše chování podobným způsobem jako bychom pod hypnózou byli, a proto je naše chování nutné psychicky. Na to ovšem namítá, že lidé sice jednají na základě intencionálních stavů, jako jsou přesvědčení, obavy nebo přání a v tomto smyslu mentální stavy fungují kauzálně, ale za žádných okolností se nedá říci, že by byl vztah příčiny a následku přísně determinován, protože v úplně stejném mentálním stavu bychom stejně dobře mohli udělat něco jiného. Psychické nutkání nebo hypnóza jsou spíše patologickými stavy, které lze od běžného chování odlišit, a proto máme i z hlediska psychologie pro lidskou svobodu ve světě místo.[120]

Pro svobodnou vůli je zásadní její vázanost na vědomí, to znamená, že ji připisujeme pouze stvořením, která mají vědomí.[121] Jak dodává, tak racionální rozhodování, které se obvykle v souvislosti s argumentem svobodné vůle udává, není ten nejzásadnější proces. Podle Searla je pouze jedním z typů zkušeností, které nás v naší svobodné vůli utvrzují. To, co považuje za nejdůležitější druh zkušeností, je vázanost svobody na záměrné intencionální jednání, jejichž podmínky splnění spočívají v pohybech těla, které jsou způsobeny bezprostředně záměrem v jednání. Protože při běžném chování se každá věc, kterou se rozhodneme udělat, pojí s přesvědčením, že za těch samých podmínek mohu udělat také jinou věc. V tomto smyslu je možné vysvětlit i chování pod hypnózou, protože zhypnotizovaný člověk v určitém smyslu neví, co dělá, protože jeho záměr v jednání je nevědomý, tedy si záměr svého jednání jednoduše neuvědomuje. Intencionální jednání je tak pro Searla podstatou lidské svobody.[122]

Jednání ze svobodné vůle se podle Searla nedokážeme vzdát tak jednoduše jako některých teorií, které se nám na základě našeho vnímání sice zdají pravdivé,

[120] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994., str. 96-98.
[121] Proto pokud sestrojíme robota, kde vědomí nepředpokládáme, pak mu ani nepřipíšeme svobodnou vůli a to ani za předpokladu, že jeho chování nebudeme schopni předvídat. Svobodu v tom smyslu v jakém ji přisuzujeme lidem, bychom připsali pouze robotovi, u kterého bychom byli přesvědčeni, že má stejné vědomí jako lidé. Tamtéž, str. 98.
[122] Tamtéž, str. 98-104.

ale nahrazují je teorie, které toto naivní zdání přesvědčivě vyvrací. Je tomu tak proto, že přesvědčení svobody je zabudováno do každého našeho vědomého intencionálního jednání, tedy není pouze věcí uvažování, ale je součástí našeho jednání.[123]

Jak později sám Searle přiznává, tak ani tento pohled na svobodnou vůli, která je zásadní pro intencionální jednání, samotný problém svobodné vůle neřeší. Protože ve světě, kde se každý jev zdá být determinovaný příčinami, které mu předcházejí, je těžké přiznat existenci čistě svobodného jednání. A jak dodává, je problém svobodné vůle specifický tím, že jeho řešení nejsme schopni přiblížit se, a přesto nejsme schopni v běžném životě existovat, aniž bychom svobodnou vůli nepředpokládali. Kdykoli se nacházíme v situaci, kdy se musíme pro něco rozhodnout, tak musíme předpokládat naši vlastní svobodu, bez které by tento úkon nebyl možný.[124]

Přesto, že Searle konečné vysvětlení svobody vůle neuvádí, je pro mne jeho pojetí zajímavé především v tom, že poukazuje na problémy kompatibilismu, který se běžně považuje za překonání problému svobodné vůle, a který ve skutečnosti tuto svobodu uznává pouze teoreticky, přičemž je jeho vysvětlení také založeno na determinismu.

[123] SEARLE, John. *Mysl, mozek, věda*, Mladá fronta, Praha, 1994, str. 104-105.
[124] SEARLE, John. *Freedom and Neurobiology. Reflection on Free Will, Language, and Political Power*. New York: Columbia University Press, 2004. Str. 10-11.

Závěr

Jedním z hlavních cílů této diplomové práce bylo představit myšlenku funkcionalismu ve vztahu k filosofii mysli, která připouští možnost připsání myšlení i jiným entitám na základě funkčnosti. Otázkou, zda i něco jiného než člověk, může mít kognitivní vlastnosti, se zabýval již Descartes. Jeho systém myšlení je založen na odlišném přístupu mezi funkcí těla a mysli, což mělo za následek dědictví dualismu, které je v určité míře vnímáno dodnes. Vzhledem k tomu, že Descartes připisoval lidskému duchu-myšlení vyšší míru důležitosti než tělu, je právě tato vyšší substance tím, co nás od ostatních entit odlišuje. Proto je pro něho připsání této vlastnosti zvířatům nebo strojům nepřípustné. Právě vztahu mezi tělem a myšlením je věnována první kapitola této práce. Pokusila jsem se zde nastínit novější pohled monismu, který je dnes obecně přijímaný. Konkrétněji jsem se věnovala pohledu Searla, který velmi zjednodušeně řečeno, považuje mysl, potažmo mentální vlastnosti za „mikrovlastnosti“ mozku, které zde vznikají a zároveň se zde také realizují.

V druhé kapitole jsem se snažila představit myšlenku funkcionalismu samotného, jenž mentální stavy posuzuje z hlediska jejich funkčnosti. V tomto smyslu je nějaký určitý mentální stav posuzován z hlediska role, kterou hraje ve vnějším systému. Je tomu tak proto, že mysl samotná je zde brána pouze jako jakási černá skříňka, kterou není možné prozkoumat a k tomu abychom mohli popsat myšlení, nám stačí znát způsob, jakým se projevuje. Pojmy, které k popisu mentálních stavu užíváme, potom mají funkční charakter, jelikož se běžně užívají spíš s odkazem na funkce chování jejich majitele než na způsob fungování jeho mozku. K otázce, zda je možné strojům připsat myšlení, se v rámci funkcionalistické tradice vyjádřil také Alan Turing, kterého v práci zmiňuji. Turing se snaží na otázku odpovědět imitační hrou, která podle něho ukáže, zda je myšlení stroje od myšlení člověka rozeznatelné. Jeho závěr zní, že pokud v testu nedokážeme rozeznat počítač od člověka, pak počítač myslí.

Tím se dostáváme ke třetí kapitole a k druhému hlavnímu cíli práce, což bylo ukázat protinázor, který myšlení strojům nepřiznává. Tento názor je zde představen na myšlenkách Johna Searla, který dokazuje, že Turing se ve svém závěru dopustil omylu. To je demonstrováno na argumentu čínského pokoje, jenž ukazuje, že Turingovým testem dokáže projít i systém, který vůbec nerozumí obsahu a o kterém je tedy těžké říci, že by doopravdy myslel. Toho je podle Searla schopen pouze tvor, jenž je vytvořen z biologického materiálu, protože mysl je možná díky biochemické struktuře mozku, který obsahuje kauzální schopnosti. Digitální počítač, který má pouze formální programy, logicky není schopen ničeho jiného než manipulace se symboly na této formální rovině, která však ke skutečnému porozumění, a tedy myšlení nestačí. Skutečné porozumění je dáno až obsahem potažmo významy, je k němu tedy potřeba také sémantika, která digitálním počítačům chybí. Systém čínského pokoje tak zvenčí sice budí zdání porozumění, tedy vypadá jako by měl intencionalitu, ovšem ve skutečnosti tuto pro myšlení podstatnou vlastnost postrádá. Intencionalita je systému spíše připsána lidmi, což ovšem podle Dennetta, jehož myšlenky jsem se také pokusila v této práci přiblížit, neznamená, že by systém určitý druh mysli neměl. V textu byli následně představeni někteří myslitelé, kteří na argument čínského pokoje reagovali v posledních letech a kteří se pokouší na argument nahlížet z různých a velice zajímavých pohledů ať už negativně nebo pozitivně.

V poslední kapitole byla popsána obecná kritika, která se k funkcionalistickému způsobu myšlení váže. Převážně je zde představena kritika, která pramení z nemožnosti vysvětlení kvalitativních stavů pomocí tohoto způsobu myšlení. Nemožnost kválií je podpořena několika nejznámějšími myšlenkovými experimenty, jež jsou zde popsány z hlediska Blocka, Chalmerse a Jacksona. Následně byl v práci představen i problém eliminativismu, který považuje naše běžné chápání mysli za špatné a podle kterého některé nebo dokonce všechny duševní stavy, které pojímá společný smysl, ve skutečnosti vůbec neexistují. Následkem přijmutí eliminativismu by potom bylo, že zde nebude žádné místo například pro morálku nebo svobodnou vůli.

Zdroje:

BLOCK, Ned. *Troubles with Functionalism*. [online]. Minnesota Studies in the Philosophy of Science, 1978. str. 261-325 [cit. 2019-02-05]. Dostupné z: http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/1978.troubles.with.f.pdf

COLE, David, "The Chinese Room Argument", *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2015/entries/chinese-room/>.

DESCARTES, René. *Rozprava o metodě: Jak vést správně rozum a hledat pravdu ve vědách*. Překlad Karel Šprunk. [1. vyd.]. Praha: OIKOYMENH, 2016. 77 s. ISBN 978-80-7298-212-7.

DESCARTES, René. *Meditace o první filosofii*. Překlad Petr Glombíček a Tomáš Marvan. [3. vyd.]. Praha: OIKOYMENH, 2015. 127 s. ISBN 978-80-7298-202-8.

DENNETT, C. Daniel. *Druhy myslí: k pochopení vědomí*. Překlad Jaroslav Peregrin. [2. vyd.]. Praha: Academia, 2004. ISBN 80-200-1177-3.

DENNETT, C. Daniel. The Unimagined Preposterousness of Zombies. *Journal of Consciousness Studies*, vol. 2, no. 4, 1995, pp. 322-326. Dostupné z: http://pp.kpnet.fi/seirioa/cdenn/unzombie.htm

HODGES, Andrew. „Alan Turing“, *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2015/entries/turing/.

HUSSERL, Edmund. *Karteziánské meditace.* Překlad M. Bayerová. [1. vyd.]. Nakladatelství Svoboda. Praha, 1968. ISBN 25-060-68 02/6.

KIRK, Robert. „Zombie“, *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2017/entries/zombies/

Kol. Autorů. *Filosofický slovník.* [2. vyd.]. Olomouc, 2002. ISBN 80-7182-064-4.

KOMÁREK, Stanislav. *Ochlupení bližní: zvířata v kulturních kontextech.* Vyd. 2. Praha: Academia, 2012, 262 s. ISBN 978-80-200-2113-7.

LEVIN, Janet, "Functionalism", *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2017/entries/functionalism/

MOORE, G. E. A Defence of Common Sense. In: *Contemporary British Philosophy*. Ed. J. H. Muirhead, 1925. Dostupné z: http://www.ditext.com/moore/common-sense.html

NIDA-RÜMELIN, Martine. „Qualia: The Knowledge Argument", *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), URL = „https://plato.stanford.edu/archives/win2017/entries/qualia-knowledge/

NOSEK, Jiří. *Funkcionalismus ve vědě a filosofii: sborník příspěvků*. Praha: Filosofia, 2001, 241 s. ISBN 80-7007-147-8.

OPPY, Graham and DOWE, David. „The Turing test", *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2015/entries/turing-test/

PAVELA, Tomáš. Funkcionalismus ve filosofii mysli. *Inflow* [online]. 2008-06-01 [cit. 2019-01-17]. Dostupné z: www.inflow.cz/funkcionalismus-ve-filozofii-mysli

PEREGRIN, Jaroslav. *Kapitoly z analytické filosofie.* Vyd. 2. Praha: Filosofia, 2014, 319 s. ISBN 978-80-7007-420-6.

PRESTON, J. and M. BISHOP (eds.), 2002, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*, New York: Oxford University Press. ISBN 978-0-19-925277-0.

RAMSEY, William, "Eliminative Materialism", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/materialism-eliminative/>.

ROBINSON, Howard, „Dualism“, *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2015/entries/dualism/

SEARLE, J., 1984, *Minds, Brains and Science*, Cambridge, MA: Harvard University Press; český překlad *Mysl, mozek, věda*, Mladá fronta, Praha, 1994

SEARLE, J., 1980, ‘Minds, Brains and Programs’, *Behavioral and Brain Sciences*, 3: 417–57

SEARLE, John. *Freedom and Neurobiology. Reflection on Free Will, Language, and Political Power*. New York: Columbia University Press, 2004. Dostupné z: http://www.summer12.isc.uqam.ca

TURING, Alan M. Computing Machinery and Intelligence. *Mind 49*. 1950. s. 433-460. Dostupné z: https://www.csee.umbc.edu/courses/471/papers/turing.pdf

YALOWITZ, Steven, „Anomalous Monism“, *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2015/entries/anomalous-monism/