



národní  
úložiště  
šedé  
literatury

### **Chtěli byste být mozkiem v baňce?**

Wiedermann, Jiří  
2006

Dostupný z <http://www.nusl.cz/ntk/nusl-35307>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 05.05.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní [nusl.cz](http://nusl.cz) .



**Institute of Computer Science**  
**Academy of Sciences of the Czech Republic**

## **Chtěli byste být mozkiem v baňce?**

Jiří Wiedermann

Technical report No. 967

duben 2006



## **Chtěli byste být mozkiem v baňce?**

Jiří Wiedermann<sup>1</sup>

Technical report No. 967

duben 2006

### Abstrakt:

Moderní teorie kognitivních systémů pohlíží na tyto systémy jako na autonomní vtělené výpočetní systémy, které se situují v okolí prostřednictvím svých senzomotorických jednotek. Přesto zejména v kruzích počítačových teoretiků je opakovaně slyšet názory, že na kognici lze pořád možné pohlížet i "klasicky", jako na problém specifického zpracování dat a že tudíž vtělení není nezbytné pro zachycení podstaty kognice. Ukážeme, že takto zjednodušený pohled opomíjí podstatnou vlastnost kognitivních systémů - a sice jejich aktivní vliv na výběr či dokonce vznik vstupních dat. Bez této zpětné vazby si systém nemůže vytvořit svůj vnitřní model světa poznáný prostřednictvím svých akcí. Pro vysvětlení povahy zmíněného problému použijeme výpočetní model kognitivních systémů zavedený autorem v předchozích pracích. Tento model umožní na principiální úrovni přemýšlet o fungování algoritmických mechanismů imitace, komunikace, vzniku řeči, myšlení a vědomí a tím přispět i k jejich pochopení v živých systémech.

### Keywords:

kognitivní systémy; vtělenost; situovanost; vnitřní model světa

---

<sup>1</sup>Tato práce vznikla v rámci výzkumného záměru AV0Z10300504 s částečnou podporou grantu 1ET100300419

# Chtěli byste být mozkiem v baňce, aneb o důležitosti vtělenosti a situovanosti při rozvoji mentálních schopností kognitivního agenta<sup>1</sup>

Jiří Wiedermann

Ústav informatiky AV ČR  
Pod Vodárenskou věží 2, 182 07 Praha 8, Česká republika  
jiri.wiedermann@cs.cas.cz

## Abstrakt

Moderní teorie kognitivních systémů pohlíží na tyto systémy jako na autonomní vtělené výpočetní systémy, které se situují v okolí prostřednictvím svých senzomotorických jednotek. Přesto zejména v kruzích počítačových teoretiků je opakovaně slyšet názory, že na kognici lze pořád možné pohlížet i „klasicky“, jako na problém specifického zpracování dat a že tudíž vtělení není nezbytné pro zachycení podstaty kognice. Ukážeme, že takto zjednodušený pohled opomíjí podstatnou vlastnost kognitivních systémů – a sice jejich aktivní vliv na výběr či dokonce vznik vstupních dat. Bez této zpětné vazby si systém nemůže vytvořit svůj vnitřní model světa poznatý prostřednictvím svých akcí. Pro vysvětlení povahy zmíněného problému použijeme výpočetní model kognitivních systémů zavedený autorem v předchozích pracích. Tento model umožní na principiální úrovni přemýšlet o fungování algoritmických mechanismů imitace, komunikace, vzniku řeči, myšlení a vědomí a tím přispět i k jejich pochopení v živých systémech.

## 1 Úvod

V nedávném čísle internetového magazínu EDGE zmiňuje přední americký neurovědec V.S. Ramachandran [7] neformální průzkum, který provádí mezi svými známými. Předmětem Ramachandranova zájmu je odpověď na následující otázku. Kdybychom měli na výběr volbu mezi svou existencí v té podobě v jaké jsme, a existencí v podobě „mozku v kádince“, jakou alternativu bychom zvolili? „Mozek v kádince“ přitom znamená, že náš mozek by existoval v kádince za podmínek, kdy všechny jeho životní funkce by byly zabezpečené a dále, dle Ramachandrana, „*použitím tisícovek elektrod a vhodného způsobu elektrické stimulace ... váš mozek bude schopen myslet a cítit stejně jako v reálném životě*“. Ramachandran dochází k závěru,

že lpět na první alternativě je iracionální. Nás však v dalším nebude zajímat, která možnost je lepší. Spíše se soustředíme na otázku, jestli je skutečně možné mozek „přelstít“ způsobem, naznačeným Ramachandranem. Tato představa by totiž znamenala, že inteligenci, a v obecnosti kognitivní procesy, je možné chápat, modelovat a zkoumat v principu jako klasický problém zpracování dat: máme daný výpočetní systém (v našem případě mozek), do kterého vhodná data vstupují a z něhož po jejich zpracování nějaká data vystupují. Ramachandran ve svém pohledu na kognici jako na klasický výpočetní problém není zdaleka osamocen. Vpravdě je to pohled značné, ne-li převážné části odborníků vzdělaných ve výpočetní teorii.

Cílem článku bude ukázat, že takový „principiální“ pohled je mylný. Je přílišným zjednodušením modelované skutečnosti, které ignoruje aktivní podíl kognitivního subjektu na výběru a dokonce i na vzniku dat vstupujících do systému. Ve 2. části ukážeme, že systém, pomocí kterého kognitivní entity interagují s vnějším světem, tvoří jejich integrální součást a musí být vzat do úvahy v jakémkoliv realistickém modelu kognice. To znamená, že je nutné uvažovat vtělené kognitivní systémy. Dále ve 3. části naznačíme, že pokud má být takový systém schopen řešit složitější kognitivní úkoly, musí být systém situovaný, tj. musí být vybaven modelem světa, ve kterém se pohybuje, včetně modelu sebe sama v rámci tohoto světa. V závěrečné 4. části naznačíme algoritmickou představu, jak lze takový model postupně budovat v interakci s prostředím a jak ve výsledném systému fungují mechanismy imitace, komunikace, myšlení a dokonce se dostaneme až na práh vědomí.

Dříve, než se pustíme do příslušných úvah, zpřesníme pojem kognitivního systému do míry postačující pro naše účely.

Kognitivní procesy jsou procesy, pomocí kterých živé organismy vnímají, lokalizují, vybírají, zpracovávají, uchovávají, generují a využívají informace ve svém chování ku svému prospěchu. V tomto chápání je (lidská)

<sup>1</sup> Tato práce vznikla v rámci výzkumného záměru AV0Z10300504 s částečnou podporou grantu 1ET100300419

inteligence komplexem provázaných kognitivních procesů, jejichž projevy lidé používají ve svém životě i ve vzájemné mezilidské interakci.

Komputacionalismus věří, že podstatnou část kognitivních procesů lze modelovat pomocí výpočtů. V rámci komputacionalismu pod pojmem „(umělý či výpočetní) kognitivní systém“ budeme rozumět vtělený počítač, který realizuje kognitivní proces, pro jehož plnění byl svým konstruktérem navržen. Vtělený počítač je počítač, který má k dispozici percepční a motorické jednotky, pomocí kterých lze řešit daný kognitivní problém.

## 2 Motivační příklad

Turing si v roce 1950 jako první uvědomil, že mezi všemi artefakty jedině počítače mají potenciál „myslet“ (viz [10]), avšak nebyl schopen podat ani definici, ani „algoritmus“ myšlení. Namísto toho však navrhnul test, který nese jeho jméno a jehož cílem bylo zjistit, jestli se počítač, se kterým komunikuje pomocí vzdáleného terminálu (dnes bychom řekli, pomocí „SMSek“) chová v interview nerozlišitelně od člověka. Pokud ano, tak nezbyvá než věřit, že komunikujeme s inteligentní entitou, resp. dle naší definice, s kognitivním systémem.

Všimněme si, že v tomto scénáři do samotného počítače vstupuje digitální informace (totiž SMSky) a podobná informace (SMSky) z něj i vystupuje. Z toho se zdá být nad slunce jasnějším, že kognitivní systém, o kterém je řeč, se z hlediska zpracování informací chová jako klasický výpočetní systém: nějaké informace do něj vstupují a transformované informace z něj vystupují. Pokud počítač na naše SMSky odpovídá smysluplně, usoudíme, že na druhém konci je inteligentní entita.

Podrobně nyní předchozí inteligentní entitu dalšímu, zdánlivě jednoduššímu testu: vybavme počítač digitální kamerou, kterou může počítač nasměrovat libovolným směrem a zaostřovat na libovolný předmět, a počítačem digitálně řízeným podvozkem, který jej přesune na libovolné místo. Dostaneme tak vlastně jakési torzo robota, kterého budeme nazývat navigátorem. Kognitivním úkolem pro navigátora bude dostat se z bodu A do bodu B, který je na dohled (vyznačen např. vlajkou). Celou věc komplikují překážky, které musí navigátor při své pouti z A do B objíždět. Věřme, že takového robota lze sestavit, a soustředíme se nyní na tok dat v celém systému. Z kamery, zacílené a zaostřené na nějaký předmět v okolí, tečou digitální obrazové informace do počítače. Zde se zpracují a počítač vydá povel k zacílení kamery na jiné místo, na její zaostření, a

pokud mu v tom nebrání překážka, na přesun do nějakého jiného místa. Poté se celý cyklus opakuje. Lze se na celý kognitivní systém opět dívat jako na klasický výpočetní systém, do kterého vstupují a z kterého vystupují digitální data? Nuže: na vstupu, před kamerou a pod koly navigátora, nemáme digitální data, nýbrž „reálný svět“. Navigátor samotný neprodukuje digitální data, nýbrž pohyb, akci (buď kamery anebo podvozku). To rozhodně nevypadá jako klasické zpracování dat. Teoretik v každém z nás ale namítne, že pokud „zapomeneme“ na navigátorovo „tělo“ (kameru a podvozek) a soustředím se pouze na počítač, tak tento zřejmě jako digitální systém pracuje. Takže, v principu, pro studium výpočetních procesů v navigátorovi by snad mohlo stačit uvažovat pouze tuto digitální část (jeho „mozek“). Do počítače budou vstupovat data, která budou odpovídat tomu, co by kamera viděla, kdyby byla uvažována, a počítač bude vydávat povely nepřítomnému podvozkovi a nepřítomné kameře. Narazíme ovšem na problém přípravy dat: data nelze předem připravit, protože nevíme, jak se počítač v dané situaci, zprostředkované daty, které by v daném okamžiku dodávala kamera, zachová. Výběr dalších dat totiž záleží na rozhodnutí počítače: jak nasměruje a zaostří kameru, a kterým směrem se vydá. Pokud chceme udržet scénář zpracování digitálních dat, musíme pro jeho provoz „nasadit“ další zařízení (nazvěme ho „simulátorem reality“, anebo stručně „simulátorem“). Toto zařízení bude sledovat pokyny mozku a bude mu dodávat taková (digitální) data, jaká by dostával v případě, že by měl „tělo“. Celá situace se začíná nápadně podobat problému „mozku v kádince“, který řeší filozofové již několik staletí, počínaje snad Descartem a konče zmiňovaným Ramachandranem.

Zastavme se nyní u problému, jak by asi mohl náš simulátor reality fungovat. Těžko si lze představit jinou cestu než takovou, že simulátor by „v sobě“ obsahoval model světa, ve kterém se navigátor pohybuje. Poslušen instrukcí navigátorova „mozku“ (pro chybějící tělo) by simulátor „navigoval“ ve svém modelu a „mozku“ by vlastně prezentoval jakousi „virtuální realitu“, takovou, s jakou by navigátor interagoval prostřednictvím svých senzomotorických orgánů (kdyby je měl).

Dalším problémem je, jak asi by bylo možné vytvořit onen zmíněný model reálného světa v simulátoru. Dříve nebo později musíme připustit, že data o reálném světě je možné získat pouze z tohoto světa, a že tudíž se tato data, snímaná vhodnou kamerou, ale nejlépe touž, kterou jsme odebrali z navigátora (aby odpadly problémy s převodem dat do formátu, jaký používal navigátor), musejí vhodným způsobem v simulátoru zapamatovat, jak již bylo zmíněno. Protože předem nevíme, jaký prostor se navigátor rozhodne zkoumat, musíme v simulátoru mít

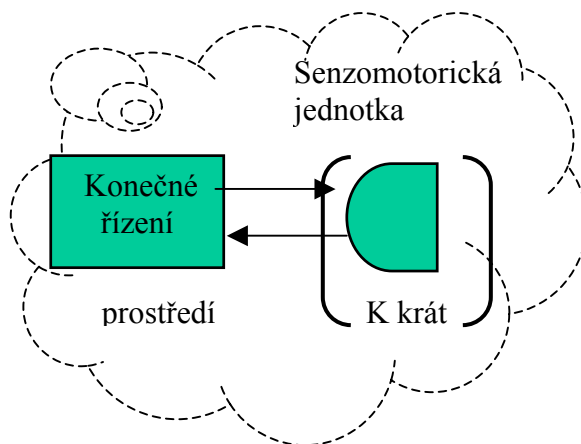
uložen „celý svět“, ve kterém se navigátor bude teď i v budoucnu pohybovat, „nafilmovaný“ ze všech možných stran v reakci na všechny v principu možné pohybové instrukce navigátora vydané kameře (kromě jiného, zaostřené i nezaostřené) a podvozku (pohyb vpřed, vzad, zastavit, mírně doleva, prudce doleva, atd.). Dovedete si to představit, nota bene pro složitější zařízení, než je náš navigátor?

Ránu z milosti zasadí posledně rozvíjené představě tento argument. Představme si, že máme v kádince mozek geniálního badatele. Tento badatel netuší, že je pouhým mozkiem v kádince, a pro svůj výzkum si staví stále důmyslnější přístroje pomocí kterých odhaluje přírodní zákony do té doby lidem neznámé. Zřejmě není v lidských silách postavit zařízení, kde by již podobné informace byly k dispozici dříve, než byl sestrojen přístroj, který je umožní získat, a dodávalo by je mozku našeho badatele. Další zajímavé simulační problémy by přinesla situace, kdyby se náš mozek v kádince zamiloval, založil rodinu a měl děti ... Mimochodem, předchozí úvahy dávají odpověď na Ramachandranovo dilema, jestli je lepší mít mozek ve vlastní těle, anebo v kádince [7]. Nikdy nevolte kádinku!

Závěr z těchto úvah je následující. Do architektury kognitivních systémů je přirozené a nutné zahrnout i jejich „tělo“, které se manifestuje prostřednictvím sensorických a motorických jednotek. Je tomu tak proto, že kognitivní systém ve vzájemné interakci a koordinaci těchto jednotek s prostředím poznává svět. Tím se stává také autonomním – nepotřebuje žádné další zařízení, které by mu poznávání umožňovalo. Tyto jednotky mu umožní volit předměty svého momentálního zájmu a tím tedy „zajímavá“ vstupní data podle svých potřeb, a generovat chování, které je na jedné straně určeno těmito daty, ale na druhé straně určuje, která data budou zkoumána v nejbližším kroku. Kognitivní systém může dokonce „sám pro sebe“ generovat svá budoucí vstupní vhodnou změnou svého okolí prostřednictvím svých motorických jednotek. Kognitivní systém tedy není pasivním konzumentem dat, které mu jsou předkládána, a generátorem dat, která jsou určena pro někoho jiného a jakmile jsou jednou vygenerována, vlastním systémem mohou klidně zůstat „nepovšimnuta“ (jak je tomu u klasických výpočetních systémů). Naopak, kognitivní systém je aktivním účastníkem nacházení svých vstupních a vyžívání svých výstupních dat – je to jeho poslání.

Vyzbrojení těmito poznatky, vraťme se k Turingovu testu: čím to je, že v jeho výpočetním modelování jsme vystačili s klasickým pohledem na výpočty? Teď již je to snad vidět: v tomto testu vystupuje kognitivní systém

jako pasivní systém. My se ptáme, on odpovídá (tak funguje interview). Předpokládejme však, že by se role zrovnoprávnily a počítač by nás mohl chtít úkolovat a vyptávat se na věci, které neví, např. „co se stane, když“; „zkuste prosím, udělat takový a takový experiment“, apod. Vbrzku bychom se ocitli v roli simulátora vnějšího světa a tím se vlastně stali i „tělem“ uvažovaného počítače. Pohled na kognici jako na klasické zpracování informace tedy nefunguje ani v případě „rovnoprávného“ Turingova testu.

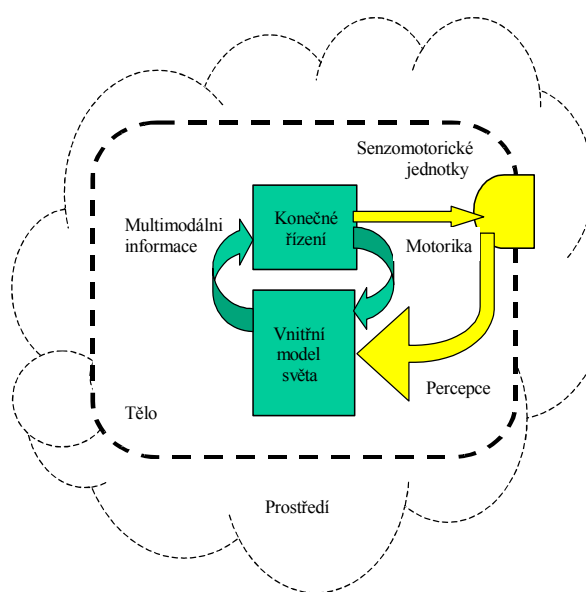


Obr. 1: Univerzální schéma vtěleného agenta

### 3 Architektura a fungování vtěleného situovaného kognitivního agenta

Na základě předchozích úvah jsme tedy dospěli k základnímu schématu kognitivního agenta: skládá se ze dvou hlavních částí – řídicí jednotky a těla, přičemž tělo je „nosičem“ sensorických a motorických jednotek agenta (Obr. 1). Toto schéma je „univerzální“ v tom smyslu, že zachycuje všechny známé živé organizmy, od těch nejjednodušších (např. bakterie) až po ty nesložitější (lidé). Náš navigátor z předchozí části měl také tuto architekturu. Zřejmě složitost řídicí jednotky roste se složitostí těla. Pokud postupujeme podél imaginární linie znázorňující růst této složitosti u živých či přepokládaných umělých organizmů, narazíme na jakýsi předěl mezi organizmy, které se zdají býti ovládané jednoduchými programky a jejichž učící schopnost je minimální, a „vyššími tvory“, které mají schopnost učit se novým kouskům, imitovat, jednoduše komunikovat, atd. Mají snad tyto „vyšší“ organizmy něco navíc, co „nižším“ chybí? Lze to zachytit již na nejnižší úrovni modelování nějakým „zjemněním“ našeho předchozího modelu?

Uvažujme tento myšlenkový experiment: vezměme si našeho navigátora a „dovybavme“ jej simulátorem (modelem) vnějšího světa, avšak nikoliv celého světa, který by navigátor eventuálně mohl navštívit, ale pouze té části světa, kterou navigátor během svého života prozkoumal. Předpokládejme dále, že v rámci svého vnitřního modelu světa bude mít navigátor k dispozici i svůj vlastní model a model sebe samého ve vnějším prozkoumaném světě. Nakonec předpokládejme, že navigátor má s každou částí modelu asociovanou (krátkou) posloupnost akcí (ve formě motorických příkazů), které lze v daném kontextu realizovat. Tuto posloupnost budeme nazývat zvyky. Výslednému zařízení říkáme navigátor+ (navigátor plus). Jeho schéma je na Obr. č. 2.



Obr. 2 Vtělený kognitivní agent s vnitřním modelem světa

Jaké výhody může mít navigátor+ v porovnání se standardním modelem? Zřejmě navigátor+ se v „poznaném“ světě lépe vyzná, protože má jeho model. Např. již na základě částečné informace o vnějším světě získané jeho senzory si navigátor+ může doplnit celek (tj. chybějící informace) podle svého vnitřního modelu, a má zde také k dispozici „návrh“ posloupnosti akcí, které lze v dané situaci realizovat (zárodek intencionality, řekli by filozofové myslí). Navigátor+ se tudíž může ve svém vnitřním světě „virtuálně pohybovat“ – může si odzkoušet výsledek nějakých svých akcí aniž by tyto akce realizoval (entitu s takovými schopnostmi nazývá Dennett „popperovská bytost“ [2]). Samozřejmě je důležité, aby navigátor+ rozlišoval mezi akcemi v rámci

vnitřního modelu, a akcemi skutečnými – totiž těmi v reálném světě. Simulovaným akcím ve vnitřním světě říkáme představivost, a schopnosti rozlišovat mezi simulovanými a pozorovanými akcemi říkáme povědomí.

Hlavní výhodou, kterou navigátor+ tímto získá, je schopnost imitovat (napodobovat) činnost jiných navigátorů. Příslušný imitační mechanismus funguje takto: navigátor+ pozoruje jiného navigátora v akci. Porovnáním se svým vnitřním modelem mechanismus zjistí, že pozorovaný objekt je stejné zařízení, jakým je navigátor+ (protože mechanismus má k dispozici model navigátora). Nyní navigátor+ pozoruje posloupnost akcí, kterou provádí pozorovaný navigátor. Pokud nadáme našeho pozorujícího navigátora schopností zapamatovat si krátké posloupnosti akcí, tak tento může právě pozorované akce zopakovat. To ovšem není nic jiného než imitace.

A dále: pokud navigátoři mají podobný repertoár akcí (čili podobné chování) — a tak tomu bude, pokud se pohybují ve stejném prostředí, tak již na základě vzájemného pozorování může jeden navigátor získat informaci o budoucích akcích (říkejme jim úmysly) druhého, protože oba dva mají podobné vnitřní modely a na základě pozorování akcí svého protějšku v nějaké situaci si dovedou doplnit (simulací ve vnitřním světě) očekávaný vývoj chování. Říkejme této schopnosti empatie. Zde začíná schopnost porozumění důsledkům akcí jiných podobných entit, a také svým vlastním akcím. Zde se formuje i „model“ sama sebe. Dále je odsud již pouze krok k primitivní komunikaci pomocí gest. Naznačením nějaké akce, charakteristickým gestem, „vyšle“ agent informaci, kterou si příjemce (pozorovatel) „doplní“ mechanismem empatie v rámci svého vnitřního modelu na celou akci. Takže pomocí jednoho gesta byla odevzdána komplexní informace, třeba příkaz „opatrně postupuj směrem, kterým ukazují“. Mimochodem, zde mohou do hry vstupovat emoce, jako součást komunikace. Na to však musí být agenti patřičně vybaveni (zvláštní mimika, změna barvy, atp.). Pokud máme agenty, kteří dovedou artikulovat, je možné gesta doplnit a postupně dokonce nahradit artikulovanými zvuky. Jsme svědky zrození řeči. Je dobré si všimnout faktu, že agenti „rozumějí“ své gestikulaci (řeči) v podstatě v termínech ukotvení příslušných gest ve stejné percepci, resp. ve zvycích a v konečném důsledku pomocí empatie. Ještě jedna důležitá poznámka: přechod od gest k artikulaci neznamená pouze to, že se gesta asociují s příslušnými zvuky, ale především to, že se asociují s pohyby mluvidel. To umožní „neslyšné mluvení“ sám k sobě a později přechod k myšlení (viz dále). Náš model tedy vysvětluje a podporuje klasickou

lingvistickou Sapir-Whorfovou hypotézu ([9], [13]), totiž domněnku, že jazyk předchází a formuje myšlení.

Od vzájemné komunikace již není daleko k myšlení. Myšlení v našem modelu není nic jiného než komunikace sám se sebou. Tím, že myslící agent komunikuje sám se sebou, spouští se mechanismus rozlišování vnějších podnětů (naslouchám, co povídám) od vnitřních (simulovaných) – povědomí je „zapnuto“. Malou modifikací (z hlediska inteligentního tvůrce) lze dosáhnout, že komunikace sama se sebou nemusí být přes vnější zpětnou vazbu (mluvím a poslouchám sám sebe) nýbrž se dá zařídit bez mluvení (např. lidé často při myšlení pohybují svými mluvidly) a dokonce i „vnitřkem“, bez pohybu mluvidel – pouze reakcí na motorické příkazy, které jsou odesílány (i) do vnitřního modelu světa. Při myšlení samozřejmě „vypneme“ percepci a realizaci motorických příkazů, takže myšlení je v našem obrázku č. 3 znázorněno jeho tmavě vyplněnou částí – cyklem mezi konečným řízením a vnitřním modelem světa. Z hlediska vnitřních mechanismů funguje agent v takovém případě podobně jako v případě, kdy dostává „reálnou“ percepční informaci a vykonává všechny motorické instrukce. V případě myšlení se děje totéž, avšak ve „virtuálním“ světě vnitřního modelu.

Nyní jsme jen krůček od vysvětlení principu vědomí v našem modelu. „Definice“ vědomí pro náš model staví na skutečnosti, že agenti již dovedou navzájem komunikovat ve vyšším jazyku, tj. nikoliv v jazyku elementárních motorických příkazů, ale v jazyku „abstraktním“, ve kterém je relativně komplexní akce (posloupnost motorických příkazů) či abstraktní pojem nahrazen slovním vyjádřením či gesty. Úroveň jazyka je tím vyšší, čím je „bohatší“ (dá se v něm komunikovat o více věcech) a čím více se oprostí od fenotypu agentů. Pokud je úroveň rozvoje takového jazyka tak vysoká, že agenti si jsou schopni referovat navzájem o svých minulých prožitcích, zážitcích, zkušenostech, popisovat své současné anebo zamýšlené činnosti, pozorované jevy, vysvětlit očekávané jevy, vykonat nějakou činnost na základě jejího slovního popisu, a jsou alespoň v principu schopni naučit se jiný vyšší jazyk či modifikovat existující jazyk, tak hovoříme, že agenti mají vědomí. Všimněme si, že takový stav nelze dosáhnout bez toho, aby agenti měli k dispozici vnitřní model světa současně se znalostmi o jeho fungování a o svém fungování v něm, a „nebyli uděláni“ tak, že se dovedou učit. Také je dobré si uvědomit, že nevyžadujeme, aby agenti byli „stejněho konstrukčního typu“ (stejný fenotyp). Jediné, co musejí mít společné, je jazyk vyšší úrovně se stejnou či podobnou sémantikou.

Z předchozího je zřejmé, že vědomí nemá povahu binární kvality – buď ho entita má, anebo nemá. Je to spíše kvalita spojitá, sahající od rudimentárních forem směrem k vyšším, které si ani nedovedeme plně představit, protože naše lidské vědomí zřejmě není jeho konečnou instancí. Je možné si např. představit vědomí vybavené mechanismem přesného vzpomínání na cokoliv, co jsme viděli, četli, slyšeli, cítili, prožili. Na druhé straně je také zřejmé, že vědomí nelze předpokládat či „zkonstruovat“ u příliš jednoduchých agentů, jejichž architektura či vtělení je příliš omezující na to, aby zvládli (zvládly?) např. imitaci.

Je zajímavé, že náš model kognitivního agenta, přes všechnu svoji jednoduchost, vlastně zesiluje Sapir-Whorfovou hypotézu: nejen že ukazuje, že jazyk je primární při rozvoji myšlení, ale dokonce naznačuje, že je primární i při rozvoji vědomí.

Teprve při velmi vysoké úrovni rozvoje abstraktního jazyka (a tím pádem i vědomí) lze uvažovat o snížení závislosti rozvoje mentálních schopností na ztělesnění a situovanosti. Tím se dostávám zpět k modelu mozku v baňce – takový mozek by snad mohl kontemplovat o matematických problémech, avšak potěšení se života by neměl ☺.

#### 4 Vnitřní model světa

Poslední problém, který musíme vyřešit, je problém vnitřního modelu. Jak takový model může vypadat, jakým mechanismem se zkonstruuje?

Při hledání odpovědi na tyto otázky je třeba vycházet z toho, že agent nemá jiné možnosti získávání dat než prostřednictvím svých senso—motorických aktivit, a pouze z těchto tzv. základních informací může vycházet při odvozování složitějších informací. Základní informace jsou tvořeny tzv. multimodálními informacemi. Je to komplex informací, které v jednom okamžiku dodávají všechny percepční (vnitřní i vnější) orgány agentova těla, společně s informacemi (ve formě motorických instrukcí), které byly vyslány všem orgánům agentova těla (tj. jak lokomočním, tak emočním i sensorickým). Není zvykem mluvit o emočních orgánech – v našem modelu jsou to mechanismy, které „generují“ emoční projevy (změnu barvy, zrychlené reakce apod.). Také se může zdát zvláštní hovořit o motorických instrukcích pro percepční oprány. Zde máme na mysli orgány, které „hýbou“ sensorickými orgány, např. okulomotorické orgány. Sloučením percepčních a motorických informací se dosahuje tzv. ukotvení percepčních (angl. grounding) informací – motorická informace je jakoby sémantikou příslušných percepčních informací, a opačně [3].



Multimodální informace jsou vybírány a zapamatovány statistickým mechanismem na základě jejich četnosti (viz např. tzv. kogitoid [11]). Podobně se pomatují často se opakující sekvence multimodálních informací. Je důležité, aby příslušná paměť měla asociativní schopnost — aby byla schopna „vybavit“ si příslušnou informaci na základě její reprezentativní části (nejčastěji motorická informace). Všimněme si, že takto získané informace slouží třem účelům. Poprvé, je zde zapamatována „syntaxe“ pozorovaného světa — jaké vjemy „hrají“ dohromady s jakou motorikou. Podruhé, jsou zde zapamatovány posloupnosti často se opakujících akcí a příslušní interní a externí vjemy (proprio- a exterocepce). Potřetí, v takto vytvořeném „modelu“ vlastními orgány poznání světa je již integrován i model agenta — jsou zde „ukotveny“ jeho vlastní vjemy jeho vlastních akcí, vnímání svého vlastního těla (pokud je patřičně vybaven příslušnými senzory), a vyskytují se zde samozřejmě vjemy odpovídající pozorování ostatních agentů. Tento model je tedy šitý na míru jak agentovi, tak i světu, je závislý na vlastnostech jeho senzorních orgánů a na jeho pohybových schopnostech a je tudíž přenosný pouze mezi věrnými funkčními kopiemi agenta (to není případ lidí).

Operace tvorby multimodálních informací z percepčních a motorických informací je základní vlastností řídicího systému agenta. Další důležitou operací nad vnitřním modelem světa je operace zjišťování podobnosti multimodální informace. Tato operace jednak zajistí podobné reakce v podobných situacích, a také přispívá k tvorbě multimodální informace odpovídající abstraktním konceptům, které subsumují konkrétní multimodální informace podobným statistickým mechanismem, který je v akci při „výběru“ často se opakujících situací. Sémantika abstraktních konceptů je ukotvena přes komunikační jazyk jak na percepční, tak i motorické informace, resp. na „shluky“ konkrétních multimodálních informací, které abstraktní pojem subsumuje (viz např. již zmiňovaný kogitoid [11]), ve kterém je abstraktní koncept chápán jako „průnik“ subsumovaných konkrétních multimodálních informací). Skladbu (syntaxi) jazyka určují zvyky.

Je třeba si uvědomit, že agent pracuje výlučně na základě informací, získaných svou vlastní aktivitou a prostřednictvím svých vlastních orgánů. Říkáme, že agent je situovaný ve svém prostředí. Stejný agent v jiném prostředí si vytvoří jiný vnitřní model světa. A naopak, agenti vybavení různými senzorními orgány ve stejném světě si také vytvoří různé vnitřní modely světa. Do třetice, agenti vybavení stejnými senzorními orgány, ale různě rozmístěnými po těle, si ve stejném světě také

vytvoří různé vnitřní modely tohoto světa. Hle, jak závisí situovanost na vtělenosti!

Zdá se, že v živých organismech na úrovni primátů a některých ptáků a snad i některých savců (psů?) mechanismem odpovídajícím tvorbě vnitřního modelu světa jsou zrcadlové neurony (objevené v 90. letech Rizzolattim et al. – viz např. [8]). V našem přístupu jsme ovšem funkci zrcadlových neuronů generalizovali na nejvyšší možnou míru (viz též [12]). Práci, zabývající se souvislostí zrcadlových neuronů a rozvojem řeči se v poslední době vyrojila celá řada, za všechny jmenujme např. [1] a [4] a další práce těchto autorů. Ramachandran ve vlivném článku [6] přirovnává objev zrcadlových neuronů k objevu struktury DNA. Pojmy a problematika vtělenosti a situovanosti jsou podrobně pojednány v učebnici [5]. Model prezentovaný v této práci je původní, ale samozřejmě vychází i z myšlenek všech výše zmíněných i dalších autorů.

## 5 Závěr

Z našich úvah vyplývá, že kognitivní systémy nelze chápat jako výpočty, nelze na ně pohlížet jako na klasické systémy zpracování informací. Je tomu tak proto, že tyto systémy se samy musejí „starat“ o to, jaká data zpracovávají a své chování musejí podřítit těmto cílům. Takový druh činnosti nelze provozovat jiným způsobem než vtělením výpočetního systému do tělesné schránky, která se stává nositelem percepčních a motorických jednotek. Akceschopnost uvažovaného kognitivního systému v daném prostředí závisí na výkonnosti, funkčnosti, množství a rozmístění těchto jednotek, na celkové tělesné architektuře a na efektivitě řízení celého systému. Akceschopnost kognitivního systému se dále zvýší, pokud je situovaný ve svém prostředí, tj. pokud je neustále informovaný o tom, co se děje, má k dispozici informace o tom, co se stalo a v jakých situacích, a má mechanismy pro odvození toho, co lze v dané situaci udělat či očekávat. To vše mu zabezpečí vnitřní model světa. Další rozvoj lze očekávat, pokud má kognitivní systém schopnost komunikovat s podobnými systémy a „vyměňovat si“ zkušenosti. V závislosti na kvalitě komunikačních schopností lze očekávat i rozvoj vědomí.

## Literatura

- [1] M. A. Arbib: The Mirror System Hypothesis: How did protolanguage evolve? In: Maggie Tallerman, editor, *Language Origins: Perspectives on Evolution*. Oxford University Press, 2005.
- [2] D. C. Dennett: *Kinds of Minds*, 1996

- [3] S. Harnad: The Symbol Grounding Problem. *Physica D* 42: 335-346, 1990.
- [4] J. R. Hurford: Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution. In: O. Kimbrough, U. Griebel, K. Plunkett (eds.): *The Evolution of Communication systems: A Comparative Approach*. The Vienna Series in Theoretical Biology, MIT Press Cambridge, MA, 2002
- [5] R. Pfeifer, C. Scheier: *Understanding Intelligence*. The MIT Press, Cambridge, Massachusetts, London, England, 1999, 697 s.
- [6] V. S Ramachandran: Mirror neurons and imitation as the driving force behind “the great leap forward” in human evolution. EDGE: The third culture, viz [http://www.edge.org/3rd\\_culture/ramachandran/ramachandran\\_p1.html](http://www.edge.org/3rd_culture/ramachandran/ramachandran_p1.html)
- [7] V. S. Ramachandran: Mirror neurons and the brain in the vat. [http://www.edge.org/3rd\\_culture/ramachandran06/ramachandran06\\_index.html](http://www.edge.org/3rd_culture/ramachandran06/ramachandran06_index.html)
- [8] G. Rizzolatti, L. Fadiga, V. Gallese, I. Fogassi: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131-141,1996
- [9] E. Sapir: 'The Status of Linguistics as a Science', 1929. In E. Sapir: *Culture, Language and Personality* (ed. D. G. Mandelbaum). Berkeley, CA: University of California Press, 1959
- [10] A. Turing: Computing machinery and intelligence. *Mind*, vol. LIX, no. 236, October 1950, pp. 433-460.
- [11] J. Wiedermann.: Towards Algorithmic Explanation of Mind Evolution and Functioning (Invited Talk). In: L. Brim, J. Gruska and J. Zlatuška (Eds.), *Mathematical Foundations of Computer Science, Proc. of the 23-rd International Symposium (MFCS'98)*, Lecture Notes in Computer Science Vol. 1450, Springer Verlag, Berlin, 1998, pp. 152—166.
- [12] J. Wiedermann: Mirror Neurons, Embodied Cognitive Agents and Imitation Learning. In: *Computing and Informatics*. Vol. 22, no. 6 (2003), p. 545-559
- [13] B. L. Whorf: 'Science and Linguistics', *Technology Review* 42(6): 229-31, 247-8, 1940. Also in B. L. Whorf: *Language, Thought and Reality* (ed. J. B. Carroll). Cambridge, MA: MIT Press, 1956