

Juraj Wiedermann

Chtěli byste být mozem v baňce?

*Pokroky matematiky, fyziky a astronomie*, Vol. 51 (2006), No. 4, 272--282

Persistent URL: <http://dml.cz/dmlcz/141327>

## Terms of use:

© Jednota českých matematiků a fyziků, 2006

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

- [9] FIALA, J.: *Jedenapůlrozměrný prostor*. *Vesmír* 84 (2005), 734–739.
- [10] HÉNON, M.: *A two dimensional mapping with a strange attractor*. *Comm. Math. Phys.* 50 (1976), 69–77.
- [11] JULIA, G.: *Mémoire sur l'itération des fonctions rationnelles*. *J. Math. Pures Appl.* (7), 4 (1918), 47–245.
- [12] KOUKAL, S., KRÍŽEK, M., POTŮČEK, R.: *Fourierovy trigonometrické řady a metoda konečných prvků v komplexním oboru*. Academia, Praha 2002.
- [13] LORENZ, E. N.: *Deterministic non-periodic flow*. *J. Atmos. Sci.* 20 (1963), 130–141.
- [14] ŠTĚPÁNEK, F.: *130 let divergentních trigonometrických řad (2. část)*. *PMFA* 49 (2004), 122–128.
- [15] ZAHN, C. T., ROSKIES, R. Z.: *Fourier description for plane closed curves*. *IEEE Trans. Comput. C-21* (1972), 269–281.
- [16] *Carleson receives 2006 Abel Prize*. *Notices Amer. Math. Soc.* 53 (2006), 679–680.
- [17] [www.abelprisen.no/en/prisvinnere/2006/lennartcarleson/henon.html](http://www.abelprisen.no/en/prisvinnere/2006/lennartcarleson/henon.html)

## Chtěli byste být mozkiem v baňce?

*Jiří Wiedermann, Praha*

**Abstrakt.** Moderní teorie poznávacích systémů pohlíží na tyto systémy jako na roboty, tj. autonomní výpočetní systémy, které jsou vybaveny čidly, jimiž se orientují v prostředí, a manipulátory, pomocí kterých se pohybují v prostředí a vykonávají v něm různé akce. Přesto zejména v kruzích počítačových teoretiků je opakovaně slyšet názory, že na poznávací proces je pořád možné pohlížet i jako na klasický problém zpracování (byť specifických) dat a že tudíž uvažování „těla“ není nezbytné pro pochopení podstaty poznávání. Ukážeme, že takto zjednodušený pohled opomíjí základní vlastnost poznávacích 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 poznávacích systémů zavedený autorem [12], [13], [14]. 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 renomovaného internetového magazínu EDGE zmiňuje přední americký neurovědec V. S. Ramachandran [8] neformální průzkum, který provádí mezi

---

Prof. RNDr. JIŘÍ WIEDERMANN, DrSc. (1948), Ústav informatiky AV ČR, v. v. i., Pod Vodárenskou věží 2, 182 07 Praha 8, e-mail: [jiri.wiedermann@cs.cas.cz](mailto:jiri.wiedermann@cs.cas.cz)

Předběžná verze tohoto článku byla přednesena na 6. česko-slovenské konferenci „Kognice a umělý život 2006“ v Třešti a zveřejněna ve sborníku příspěvků z této konference.

Tato práce vznikla v rámci výzkumného záměru AV0Z10300504 s částečnou podporou grantu 1ET1003004191E.

svými známými. Předmětem Ramachandranova zájmu je odpověď na 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 schopný myslet a cítit stejně jako v reálném životě.*“ Ramachandran dochází k závěru, že lpět na první alternativě je iracionální. Nás však dále nebude v prvé řadě 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 poznávací 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 poznávání 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 pohled je principiálně mylný. Je přílišným zjednodušením modelované skutečnosti, které ignoruje aktivní podíl poznávacího subjektu na výběru a dokonce i na vzniku dat vstupujících do systému. Ve 2. části ukážeme, že systém, pomocí kterého poznávací entity interagují s vnějším světem, tvoří jejich integrální součást a musí být vzat do úvahy v jakémkoliv realistickém modelu výpočetního poznávacího systému. To znamená, že je nutné uvažovat tzv. vtělené poznávací systémy, tj. systémy, jejichž součástí jsou senzorní a motorické jednotky. Dále ve 3. části naznačíme, že pokud má být takový systém schopen řešit složitější poznávací 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 poznávacího systému do míry postačující pro naše účely.

Poznávací (neboli 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) inteligence komplexem vzájemně propojených kognitivních procesů, jejichž projevy lidé používají ve svém životě i ve vzájemné mezilidské interakci.

Stoupenci komputacionalismu věří, že podstatnou část kognitivních procesů lze modelovat pomocí výpočtů. V rámci komputacionalismu pod pojmem „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 (senzorní) a motorické jednotky, které mu umožňují řešit daný poznávací problém. Jinými slovy, výpočetní kognitivní systém je vlastně robot — umělý organismus. Tento robot samozřejmě nemusí vypadat jako mechanická loutka — může být vybaven různými „orgány“, jakými jsou kamery, radary, ultrazvukové

senzory, tlaková čidla, teplotní čidla, receptory chemických látek, mikrofony, radiokomunikátory, manipulátory, chapadla, kola, pásy, nohy, křídla, ploutve, lodní šrouby, trysky, vrtule atd., které se jeví pro řešení daného kognitivního problému v daném prostředí jako nejvhodnější. Všechny shora uvedené „orgány“ nemusejí nutně být „mechanické“, ale mohou být třeba organického původu, např. uměle vypěstované pro daný účel. Robot dokonce nemusí být „souvislý“, může se skládat z více oddělených částí, které komunikují bezdrátově (to je např. případ robotického fotbalu, kde jednotliví roboti komunikují rádiem s řídicí jednotkou, která je napojena na kameru společnou pro všechny roboty jednoho „mužstva“). Robot také může být velmi malý (např. jako bakterie), vyrobený nanotechnologií anebo pomocí genového inženýrství a pole jeho působnosti může být třeba uvnitř lidského těla.

## 2. Motivační příklad

Věhlasný anglický matematik A. M. Turing si v roce 1950 jako první uvědomil, že mezi všemi artefakty jedině počítače mají potenciál „myslet“ (viz [11]), avšak nebyl schopen podat ani definici, ani „algoritmus“ myšlení. Namísto toho však navrhl test, který nese jeho jméno. Jeho cílem bylo zjistit, zda se počítač, s nímž 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. podle 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 plyne, 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 dalším, 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 sestrojít, a soustředíme se nyní na tok dat v celém systému. Z kamery, namířené 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 poznávací 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 jako celek 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 soustředíme se pouze na počítač, tak zřejmě pracuje jako digitální systém. Takže v principu by pro studium výpočetních procesů v navigátorovi 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 podvozku 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.

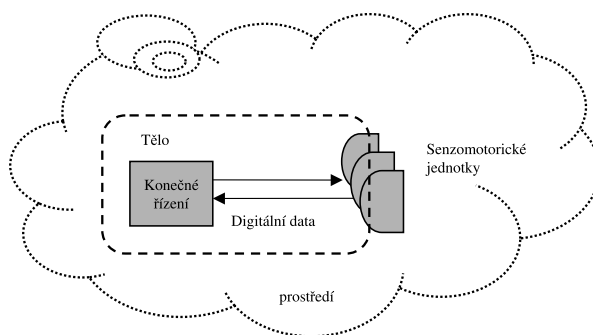
Zastavme se nyní u problému, jak by asi mohl náš simulátor reality fungovat. Těžko si lze představit jinou cestu než takovou, že simulátor by „v sobě“ obsahoval model světa, ve kterém se navigátor pohybuje. Poslušen instrukcí navigátorova „mozku“ (pro chybějící tělo) by simulátor „navigoval“ ve svém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

Další otázkou 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... Mimořádně, předchozí úvahy dávají odpověď na Ramachandranovo dilema, jestli je lepší mít mozek ve vlastním těle, anebo v kádince [8]. 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 data vhodnou změnou svého okolí prostřednictvím svých motorických jednotek. Příkladem takového chování u živých organismů je pachové značení cest u mravenců anebo psaní v případě lidí. Kognitivní systém tedy není pasivním spotřebitelem dat, která mu jsou předkládána. Není ani 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 nalézání svých vstupních a využívání 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. 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.



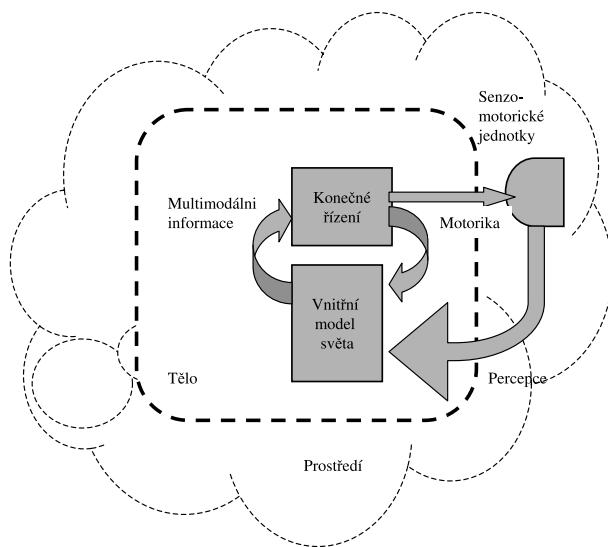
Obr. 1. Univerzální schéma vtěleného agenta.

### 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

„nositel“ sensorických a motorických jednotek agenta (obr. 1). Toto schéma je „univerzální“ v tom smyslu, že zachycuje všechny známé živé organismy, od těch nejjednodušších (např. bakterie) až po ty nejsložitější (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 předpokládaných umělých organismů, narazíme na jakýsi předěl mezi „nižšími“ organismy (které se zdají být ovládány 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“ organismy 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. 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. Navigátor+ se tudíž může ve svém vnitřním světě „virtuálně pohybovat“ — může si odzkoušet výsledek nějakých svých akcí, aniž by tyto akce realizoval. 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 „vcítění“ (empatie). Zde začíná schopnost porozumění důsledkům akcí jiných podobných entit, a také svým vlastním akcím. Zde se formuje i „model“ sama sebe. Odtud je 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 toho, 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 neznamena 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 Sapirovu-Whorfovu hypotézu ([10], [15]), 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 zařadit 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 obr. 2 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 je založena na skutečnosti, že agenti již dovedou navzájem komunikovat ve vyšším jazyku, tj. nikoliv v jazyku elementárních motorických příkazů, ale v jazyku „abstraktním“, v němž je relativně komplexní akce (posloupnost motorických příkazů) či abstraktní pojem nahrazen slovním vyjádřením či gesty. Úroveň jazyka je tím vyšší, čím je „bohatší“ (lze v něm komunikovat o více věcech) a čím více se oprošťuje 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 říkáme, že agenti mají vědomí. Všimněme si, že takového stavu nelze dosáhnout, aniž by agenti měli k dispozici vnitřní model světa současně se znalostmi o jeho fungování a o svém fungování v něm, a „nebyli udělání“ tak, že se dovedou učit. Také je dobré si uvědomit, že nevyžadujeme, aby agenti byli „stejněho konstrukčního typu“ (měli 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 např. imitaci.

Je zajímavé, že náš model kognitivního agenta přes všechnu svou jednoduchost vlastně zesiluje Sapirovu-Wharfovou hypotézu: nejen 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 tedy i vědomí) lze uvažovat o snížení závislosti rozvoje mentálních schopností na ztělesnění a situovanosti. Tím se dostáváme zpět k modelu mozku v baňce — takový mozek by snad mohl kontemplotovat o matematických problémech, avšak potěšení ze života by neměl :-)

#### 4. Vnitřní model světa

Poslední problém, který budeme řešit, je problém vnitřního modelu. Jak takový model může vypadat, jakým mechanismem se konstruuje?

Při hledání odpovědi na tyto otázky je třeba vycházet z toho, že agent nemá jiné možnosti získávání dat než prostřednictvím svých senzomotorických 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 orgány. Zde máme na mysli instrukce, které „hýbou“ sensorickými orgány (např. instrukce pro okulomotorické orgány). Sloučením souvisejících percepčních a motorických informací se dosahuje tzv. ukotvení percepčních informací — motorická informace je jakoby sémantikou příslušných percepčních informací a opačně [3].

Multimodální informace jsou vybírány a zapamatovány statistickým mechanismem na základě jejich četnosti. Podobně se pamatují 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 příslušné interní a externí vjemy (proprio- a exterocepce). Potřetí, v takto vytvořeném „modelu“ vlastními orgány poznaného světa je již integrován i model agenta — jsou zde „ukotveny“ jeho vlastní vjemy jeho vlastních akcí, vnímání jeho vlastního těla (pokud je patřičně vybaven příslušnými senzory), a vyskytují se zde samozřejmě vjemy odpovídající pozorováním ostatních agentů. Tento model je tedy šitý na míru jak agentovi, tak i světu, je závislý na vlastnostech jeho sensorických 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, jednak také přispívá k tvorbě multimodální informace odpovídající abstraktním konceptům. Abstraktní koncepty 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. Skladbu, tj. 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 senzory v stejném světě si také vytvoří různé vnitřní modely světa. Do třetice, agenti vybavení stejnými senzory, 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. [9]). V našem přístupu jsme ovšem funkci zrcadlových neuronů generalizovali na nejvyšší možnou míru (viz též [13], [14]). Práci zabývající se souvislostí zrcadlových neuronů s rozvojem řeči se v poslední době vyrojila celá řada, za všechny jmenujme např. [1], [2], [4] a další práce těchto autorů. Ramachandran ve vlivném článku [7] přirovnává objev zrcadlových neuronů k objevu struktury DNA. Pojmy a problematika vtělenosti a situovanosti jsou podrobně pojednány v učebnicích [5], [6]. Model uvedený v této práci je původní, ale samozřejmě vychází i z myšlenek všech již 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 podřídit 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.

## L i t e r a t u r a

- [1] ARBIB, M. A.: *The Mirror System Hypothesis: How did protolanguage evolve?* In: TALLERMAN, M., editor, *Language Origins: Perspectives on Evolution*. Oxford University Press 2005.
- [2] FELDMAN, J. A.: *From Molecule to Metaphor: A Neural Theory of Language*. The MIT Press, June 2006, 360 s.

- [3] HARNAD, S.: *The Symbol Grounding Problem*. *Physica D* 42 (1990), 335–346.
- [4] HURFORD, J. R.: *Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution*. In: KIMBROUGH, O., GRIEBEL, U., PLUNKETT, K. (eds.): *The Evolution of Communication Systems: A Comparative Approach*. The Vienna Series in Theoretical Biology, MIT Press Cambridge, MA, 2002.
- [5] PFEIFER, R., SCHEIER, C.: *Understanding Intelligence*. The MIT Press, Cambridge, Massachusetts, London, England 1999, 697 s.
- [6] PFEIFER, R., BONGARD, J., BROOKS, R., IWASAWA, S.: *How the body shapes the way we think — A new view of intelligence*. The MIT Press, 2006, 480 s.
- [7] RAMACHANDRAN, V. S.: *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)
- [8] RAMACHANDRAN, V. S.: *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)
- [9] RIZZOLATTI, G., FADIGA, L., GALLESE, V., FOGASSI, I.: *Premotor cortex and the recognition of motor actions*. *Cognitive Brain Research* 3 (1996), 131–141.
- [10] SAPIR, E.: *‘The Status of Linguistics as a Science’, 1929*. In SAPIR, E.: *Culture, Language and Personality* (ed. MANDELBAUM, D. G). Berkeley, CA: University of California Press 1959.
- [11] TURING, A.: *Computing machinery and intelligence*. *Mind*, vol. LIX, no. 236, October 1950, 433–460.
- [12] WIEDERMANN, J.: *Towards Algorithmic Explanation of Mind Evolution and Functioning* (Invited Talk). In: BRIM, L., GRUSKA, J., ZLATUŠKA, J. (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, 152–166.
- [13] WIEDERMANN, J.: *Mirror Neurons, Embodied Cognitive Agents and Imitation Learning*. In: *Computing and Informatics*, Vol. 22, no. 6 (2003), 545–559.
- [14] WIEDERMANN, J.: *HUGO: A Cognitive Architecture with an Incorporated World Model*. Proc. of the European Conference on Complex Systems ECCS ’06, Saïd Business School, Oxford University, 2006, viz také Technical report No. 966, Ústav informatiky AV ČR, 2006.
- [15] WHORF, B. L.: *‘Science and Linguistics’*. *Technology Review* 42 (6) (1940), 229–231, 247–248. Also in WHORF, B. L.: *Language, Thought and Reality* (ed. CARROLL, J. B.), Cambridge, MA: MIT Press 1956.