

UNIVERZITA KOMENSKÉHO V BRATISLAVE  
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY  
KATEDRA APLIKOVANEJ INFORMATIKY



---

VÝPOČTOVÉ MODELOVANIE  
EVOLÚCIE JAZYKA  
V MULTI-AGENTOVOM SYSTÉME

*Písomná práca k dizertačnej skúške*

MGR. ĽUDOVÍT MALINOVSKÝ

---

Školiteľ: doc. Ing. Igor Farkaš, PhD.

Bratislava, marec 2011



## Abstract

The paper reviews the research of origins of human language in more rudimentary communication forms in the process of evolution. Various ways of information transmission between minds are discussed, especially language and the theory of mind. Previous views of relation between language and simpler communication and their evolution are summarized. The second part of the work reviews current scientific approaches of study of these phenomena, and analyzes the concept of language games dominating in mathematical models, computational simulations and laboratory experiments. It describes recent computational models of language emergence dealing with various questions, from the role of the learning method, through neurological basis, to evolution. Summary of our thesis is that language originated gradually, during several phases. At first, the cooperative evolutionary pressure shaped the communication to correctly convey content, only later the competitive motivations became more dominant and communication exapted for unilaterally beneficial uses and the ability to pretend emerged. This paved the way to more complicated language pragmatics and later potentially to other phenomena that led to fully structured language. The proposed computational model should test this hypothesis.

## Abstrakt

Práca je prehľadom výskumu vzniku jazyka u ľudí z jednoduchších foriem komunikácie v procese evolúcie. Vymedzuje chápanie a vlastnosti rôznych druhov prenosu informácie z jednej mysle do druhej, najmä jazyka a teórie mysle. Stručne sa venuje doterajším pohľadom na vzťah jazyka a jednoduchších foriem komunikácie a ich evolúcie. Druhú časť práce tvorí stručný prehľad aktuálnych vedeckých prístupov skúmajúcich tieto fenomény a analýza konceptu jazykových hier dominujúcich v matematických modeloch, výpočtových simuláciách i laboratórnych experimentoch. Popisuje niektoré aktuálne výpočtové modely vzniku jazyka zamerané na rôzne konkrétne otázky, od vplyvu metódy učenia cez neurologickú podstatu až po evolúciu. Našou centrálnou tézou je, že jazyk vznikol vo viacfázovom procese, keď najskôr prevažoval tlak na evolúciu obsahovej stránky komunikácie užitočnej v kooperatívnom prostredí, no neskôr došlo k nárastu vplyvu kompetitívnych motivácií a exaptácii komunikácie na jednostranne výhodný prenos a kognitívnu schopnosť predstierania. Tým sa otvorila cesta vzniku zložitej jazykovej pragmatiky a neskôr potenciálne ďalším javom vedúcim k vysoko štruktúrovanému jazyku. Navrhovaný výpočtový model by mal túto hypotézu preskúmať.



# Obsah

<b>1</b>	<b>Komunikácia a evolúcia</b>	<b>7</b>
1.1	Komunikácia . . . . .	7
1.2	Emergencia a evolúcia . . . . .	10
1.3	Jazyk . . . . .	14
1.4	Teória mysle . . . . .	18
<b>2</b>	<b>Modelovanie emergencie jazyka</b>	<b>21</b>
2.1	Prístupy . . . . .	21
2.2	Komunikácia a hry . . . . .	23
2.3	Konkrétne modely . . . . .	27
2.3.1	Učenie a syntax (Kirby, 2002) . . . . .	27
2.3.2	Talking heads (Steels, 1998) . . . . .	29
2.3.3	Baldwinov efekt (Kvasnička a Pospíchal, 1999) . . . . .	30
2.3.4	Fungus eaters (Cangelosi a Parisi, 1998) . . . . .	31
2.3.5	Evolúcia imitácie (Borenstein a Ruppín, 2005) . . . . .	31
2.3.6	Sebapozorovanie (Bonaiuto a Arbib, 2010) . . . . .	33
2.3.7	Mentálny stav (Oztop a kol., 2005) . . . . .	35
<b>3</b>	<b>Projekt</b>	<b>37</b>
3.1	Základné tézy . . . . .	37
3.2	Vlastnosti a ciele modelu . . . . .	39
3.3	Návrh modelu . . . . .	40
3.4	Diskusia . . . . .	43



# Kapitola 1

## Komunikácia a evolúcia

### 1.1 Komunikácia

Ak chceme študovať vznik komunikácie, mali by sme si ju najprv definovať, aby sme mali kritérium, podľa ktorého môžeme vyhodnotiť, či v systéme ku komunikácii dochádza. V tejto oblasti nepanuje jednoznačná zhoda. Oliphant (1997) porovnáva rôzne definície komunikácie, ktoré vznikli v rôznych vedeckých kontextoch. Hlavným sporným bodom je podmienka intencionality, úmyselnosti vyslania a následného rozpoznanie úmyslu pri prijímaní informácie, ktorá sa vyskytuje najmä pri skúmaní ľudskej komunikácie. Zvieratám sme väčšinou menej ochotní priznať úmysel, často považujeme vysielané signály a reakcie na ne za inštinktívne, napriek tomu hovoríme o zvieracej komunikácii. Druhý sporný bod sa týka výhodnosti prenosu informácie. Zatiaľ čo niektoré definície sa uspokojia s tým, ak je prenos výhodný pre jednu (ktorúkoľvek, alebo konkrétnu) stranu, iné požadujú výhodnosť pre obe strany. Pre účely skúmania vzniku ľudskej komunikácie a toho, ako sa vyvinula zo zvieracej, je praktické akceptovať Oliphantove rozdelenie, ktoré nekladie podmienku explicitnej intencionality agenta, ale rozdeľuje prenos informácie podľa toho, pre koho je výhodný, nasledovne:

- **manipulácia** je prenos výhodný pre vysielateľa
- **exploatácia** je prenos výhodný pre prijímateľa
- **komunikácia** je prenos, ktorý je manipuláciou i exploatáciou súčasne

Na prvý pohľad existuje priepasť medzi zvieracími a ľudskými spôsobmi komunikácie. Kým ľudia používajú zložitú syntax, slovnú zásobu si neustále menia a rozširujú, odkazujú sa na abstraktné či neexistujúce koncepty,

Tabuľka 1.1: Typy komunikácie podľa Gärdenforsa

	<b>Jednoduché prvky</b>	<b>Kompozícia</b>	<b>Syntax</b>
<b>Signály</b>	zvieracie signály	včelie tance	-
<b>Symboly</b>	holistický jazyk	protojazyk	jazyk

zvieratá používajú obmedzený počet často vrozených signálov, ktoré nekombinujú do nových významov a komunikujú kvôli bezprostredným potrebám Gärdenfors (1995).

Gärdenfors (1996) klasifikuje druhy komunikácie podľa dvoch osí: oddelenosť reprezentácií a miera kompozicionality. Reprezentácia, teda spôsob, akým komunikačný prostriedok alebo stav mysle odkazuje na svoj význam, môže byť podľa neho dvoch druhov: navodená (cued) alebo oddelená (detached).

Navodená reprezentácia odkazuje na predmet alebo javy, ktorý sa nachádza niekde v blízkom priestorovom a časovom okolí jedinca. Aby ju jedinec mohol úspešne použiť, daný jav musí byť prítomný, musí byť možné naň zamerať pozornosť, alebo tak muselo byť veľmi nedávno a reprezentácia javu ostala v mysli jedinca ako priamy dôsledok kontaktu s javom.

Oddelená reprezentácia naopak znamená odkaz na javy či objekty mimo aktuálnej situácie, také, ktoré by do situácie len teoreticky mohli vstúpiť, prípadne ani to nie. Príkladom z prírody je situácia, keď hladný šimpanz narazí na termitisko a vyberie sa hľadať vhodnú paličku na lov termitov (Gärdenfors, 2004). Hlad a termitisko sú bezprostredne prítomné v jeho mysli, sú to navodené reprezentácie. Pre paličku vhodnú na lov termitov si však už musí vytvoriť oddelenú reprezentáciu, lebo do situácie vstúpi až vtedy, keď sa mu ju podarí nájsť, ak sa mu to vôbec podarí. Napriek tomu je schopný s ňou kalkulovať pri tvorbe plánu na zahnanie hladu (zrejme by o nej dokonca vedel aj komunikovať, ak by sme ho naučili symbol pre taký predmet).

Podľa Gärdenforsa (2004) majú niektoré zvieratá oddelené reprezentácie akcií a objektov, nie sú však schopné vytvoriť si oddelené reprezentácie cieľa, vždy len reagujú na existujúcu situáciu. Prostriedky na komunikáciu navodených reprezentácií Gärdenfors stotožňuje so signálmi, prostriedky na komunikáciu oddelených reprezentácií so symbolmi. Medzistupňom sú ikonické symboly, ktoré síce môžu odkazovať na niečo neprítomné v situácii, samy však túto vec nejakým spôsobom pripomínajú. Skutočnými symbolmi sú potom také, ktoré s odkazovaným javom nemajú vôbec nič spoločné, okrem zaužívaného štandardu, sú teda arbitrárne. Hranica medzi navodenými a oddelenými reprezentáciami nie je ostrá, sú to skôr dva extrémny kontinua.



Z hľadiska kompozicionality vníma tri úrovne: samostatné signály, jednoduchú kompozíciu a syntax. Samostatné signály nemajú štruktúru, nepoužívajú sa v kombinácii s inými signálmi, popisujú situáciu ako celok. Drvivá väčšina zvieracích signálov má tento charakter. Jednoduchá kompozícia už oddeľuje zložky situácie a umožňuje ich kombinovať novým spôsobom, nie je ich však veľký počet, nezáleží na ich poradí a často nie je možné vytvárať zložité nové významy, len kombinovať už existujúce. Na pochopenie významu nemusí záležať na poradí zložiek. Príkladom je včelí tanec, ktorý vyjadruje rôznymi svojimi aspektami smer, vzdialenosť a výdatnosť zdroja potravy, avšak nikdy nič iné, ani neumožňuje spýtať sa na chýbajúcu časť informácie a podobne. Niektoré ľudopopy, papagáje, či delfíny sú schopné v laboratórnych podmienkach produkovať či reagovať na takéto jednoduché kompozitné výroky, v prírode však tieto javy pozorujeme pomerne málo. Aj keď túto schopnosť majú, vystačia si väčšinou s holistickými signálmi. Jediný druh so zložitou syntaxou a oddelenými reprezentáciami je človek a jeho jazyk. Príklad syntaktickej komunikácie v kombinácii s čisto navodenými reprezentáciami nepoznáme.

Ďalším dôležitým aspektom komunikácie, ktorý vplýva na jej charakter, je komunikačný kanál, modalita, cez ktorú sa informácia prenáša. Zvuková modalita umožňuje veľmi efektívnu komunikáciu jedného ku mnohým (broadcast), a vďaka tomu, že každý človek má iný hlas, vieme aj len čisto z počutia rozpoznať zdroj. Naproti tomu znaková reč a väčšina gest vyžadujú, aby minimálne príjemca bol tvárou otočený smerom k vysielateľovi, pri dialógu obaja (Arbib, 2005; Corballis, 2009). Hodí sa viac na komunikáciu jedného s jedným. Okrem toho mnohé živočchy používajú na odovzdávanie správ chemickú (čuchovú, chuťovú) modalitu, ktorá zase umožňuje zanechať odkaz na jednom mieste na dlhšiu dobu. Ľudskou analógiou sú rôzne vyryté či namaľované značky a písomné odkazy. Podľa niektorých teoretikov sa ľudská schopnosť jazyka vyvinula najprv na účely komunikačnej gestikulácie, až neskôr došlo ku zmene modality na zvukovú (Arbib, 2005; Corballis, 2009).

Väčšina ľudí v praxi explicitný jazyk dopĺňa o tzv. neverbálnu komunikáciu. Umožňuje nám vcítiť sa do pozície partnera, pochopiť jeho postoj k tomu, čo hovorí, čo hovoríme my, prípadne komunikovať niektoré aspekty významu úplne bez slov, čiste výrazmi tváre, postojom tela, prozódiovou rečou a podobne. Čo však vieme určite, je, že ľudia postihnutí rôznymi formami autizmu majú v tejto oblasti zjavný deficit (Baron-Cohen a kol., 1985; Bašnáková a Rybár, 2010). To, čo im podľa Barona-Cohena chýba, je, že si nevedia vytvoriť teóriu mysle, nevedia prisudzovať iným ľuďom správne mentálne stavy, odlišné od ich vlastných, na základe neverbálnych náznakov a kontextu.

Teóriu mysle môžeme chápať ako nástroj na zdieľanie obsahu mysle, ktorý je alternatívny k jazyku. Nemá zrejme takú vyjadrovaciu silu, ale na rozdiel

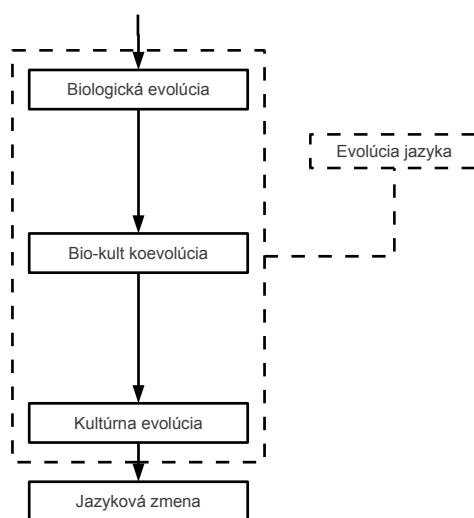
od jazyka sa deje automaticky a podvedome, a to pokiaľ ide o vysielanie i prijímanie, plné vedomé zvládnutie si vyžaduje kontrolu a uvedomovanie si všetkých vonkajších telesných prejavov, ktoré nemožno vnímať oddelene jeden od druhého. To znamená, že nám umožňuje komunikovať (obojstranne výhodne) veci, ktoré nevieme, alebo nechceme odkomunikovať jazykom a v istých situáciách ho môže efektívne dopĺňať. Navyše, ak jedna strana interakcie zvláda teóriu mysle lepšie ako druhá, môže to využiť na jednosmerne výhodný prenos informácie: manipuláciu (napríklad predstierať iný stav mysle než skutočný a spoliehať sa, že to druhá strana nerozpozná) alebo exploatáciu (rozpoznať obsah mysle, ktorý sa druhá strana neúspešne snažila zakryť).

## 1.2 Emergencia a evolúcia

Keď sa v kontexte multiagentových systémov hovorí o emergencii nejakého javu, máme na mysli, že sa následkom početných lokálnych interakcií medzi agentmi objaví v systéme nejaká globálna vlastnosť, napríklad spoločný komunikačný protokol (Wagner a kol., 2003). Špeciálnym prípadom emergencie, keď sa jav popri lokálnych interakciách vyvinul v procese prirodzeného výberu, je evolúcia. Prirodzený výber je proces, v ktorom sa vyhodnocuje schopnosť každého agenta nejakej generácie preniesť svoje vlastnosti na čo najväčšie percento agentov nasledujúcej generácie. Ak je agent schopný odovzdať aj vlastnosti získané až počas života, hovoríme o Lamarckovskej evolúcii, ak je schopný odovzdať iba tú časť svojej informácie, s ktorou sa narodil, hovoríme o Darwinovskej evolúcii.

Klasický neodarwinistický pohľad vidí biologickú evolúciu ako darwinovský proces, kde replikátory v podobe zdedených génov kódujú telesné prejavy jedincov, ktoré potom rozhodujú o ich reprodukčnom úspechu (Dawkins, 1976). Úspešní jedinci s čiastočnou nepresnosťou replikujú svoje gény, s ktorými sa narodili, efektívnejšie, a tým pádom sú samotné gény, respektíve ich funkčné navzájom sa podporujúce zoskupenia, úspešné. Jednotkou, ktorá v biologickom evolučnom procese súťaží, zdokonaľuje sa a v každom čase je jej produktom, sú teda na úrovni replikátorov genómy a na úrovni ich nositeľov druhy.

Memetici (Blackmore, 2000; Dawkins, 1976) rozširujú tento pohľad u ľudí o nové replikátory, mémy, a upozorňujú, že naše správanie v súčasnosti už zďaleka nie je dané len tým, s čím sme sa narodili, ale aj tým, čo sme sa naučili od svojho okolia počas života, a to až do tej miery, že môžeme túto kultúrnu memetickú evolúciu skúmať oddelene od biologickej. Mémy sa šíria také, aké si ich jedinci počas života pozmenili, ide teda o Lamarckovskú evolúciu. Na poli kultúrnej evolúcie súťažajú fungujúce zoskupenia mémov, tzv.



Obr. 1.1: Prechod od biologickej ku kultúrnej evolúcii (Scott-Phillips a Kirby, 2010).

memeplexy, ktoré sa prejavujú v podobe kultúr, náboženstiev, ideológií, jazykov, umeleckých prúdov a podobne. Jazyk je na jednej strane významným a účinným prostriedkom na kultúrne šírenie vzorcov správania a mentálnych obsahov, ktoré ich podmieňujú, teda nástrojom kultúrnej evolúcie, a zároveň je sám súčasťou kultúry, teda predmetom kultúrnej evolúcie.

Vznik komunikácie a jazyka znamenal významný prelom v evolučnom procese, ktorý sa zvykne porovnávať s prechodom od asexuálnej evolúcie k sexuálnej, alebo s prechodom od jednobuknových organizmov ku mnohobunkovým (Maynard-Smith a Szathmáry, 1997; Kirby, 2007). Keď hovoríme o evolúcii jazyka, myslíme tým proces jeho vzniku ako fenoménu, ktorý jednak umožnil a jednak podlieha kultúrnej evolúcii, pričom v istom období bol zrejme formovaný kultúrnymi i biologickými evolučnými procesmi súčasne - *biologicko-kultúrnou koevolúciou*. Po tom, čo sa vyvinuli všetky aspekty jazyka ako fenoménu od fonológie cez syntax až po sémantiku a pragmatiku, a dochádza už iba k súťaži a vývoju jeho konkrétnych podôb, hovoríme o jazykovej zmene (Scott-Phillips a Kirby, 2010).

Existujú rôzne názory na to, či na emergenciu komunikačného štandardu alebo schopnosti vytvoriť si ho treba viacgeneračný proces, alebo nám stačí jedna generácia. Jadrom diskurzu je protiklad vrodenej a získanej jazykovej zručnosti (Perfors, 2002). Klasický nativistický postoj je, že jazykové schopnosti zabezpečuje špecifický modul v mozgu, s ktorým sa narodíme, máme ho teda zdedený geneticky, po narodení už len doladíme niektoré parametre. Argumenty preň sú založené najmä na schopnosti detí vytvoriť si

syntax aj pri nedostatku vonkajších vstupov: jednak schopnosť malých detí naučiť sa jazyk od dospelých v relatívne krátkom čase, z malého množstva vstupov a bez dostatočnej spätnej väzby (poverty of stimulus), absencia niektorých druhov chýb, existencia kritického obdobia pre akvizíciu prvého jazyka (založená najmä na skúsenostiach s tzv. divokými deťmi), samovoľný vznik kreolských jazykov s plnou syntaxou u detí rodičov hovoriacich pidgin jazykom, či príklad samovoľného vzniku plnohodnotnej znakovej reči u hluchých detí v Nikaragui (Senghas a kol., 2004). Ďalšou dôležitou podporou je dvojité disociácia jazykovej schopnosti: existencia špecifickej jazykovej poruchy, ktorá ponecháva ostatné kognitívne schopnosti nedotknuté, a naopak všeobecnej poruchy inteligencie, ktorá obchádza jazyk (Williamsov syndróm). Za neurologické centrum jazyka sa považuje kombinácia Brockovej a Wernickeho oblasti, ktoré sú navzájom prepojené a zodpovedné za porozumenie a produkciu jazyka. Klasickou empirickou podporou sú v tomto prípade početné príklady Brockovej a Wernickeho afázie.

Nenativistická opozícia vidí jazyk ako vedľajší produkt našich všeobecných intelligenčných schopností. Väčšinu argumentov nativistov odmieta: dospelí inštinktívne hovoria s deťmi zjednodušeným a dôraznejším jazykom a opravujú ich, keď urobia chybu (čiže sú inštinktívne dobrými učiteľmi), a sú schopní naučiť sa druhý jazyk kedykoľvek počas života tak dobre, že stratia svoj pôvodný prízvuk. Divokých detí sme skúmali veľmi málo a mali poškodených oveľa viac schopností než len jazyk. Špecifická jazyková porucha podľa najnovších výskumov vyplýva z poruchy v spracovaní fonologickej štruktúry hovorenej reči, nie jazyka ako celku, a Williamsov syndróm zrejme postihuje jazykové schopnosti viac, než sme si mysleli. Neurologické výskumy jazykových centier zatiaľ neobjavili štruktúry, ktoré by vysvetľovali našu schopnosť jazyka, naopak, existujú homologické oblasti u "nižších" primátov s podobnými štruktúrami, ale inou funkciou (rozpoznávanie akcií, imitácia). Jazyk teda môže byť len výsledkom použitia týchto centier a ich výpočtových vlastností pre komunikačnú funkciu. Podľa nenativistov by sme teda mali aplikovať Occamovu britvu a teórie s prebytočnými neopodstatnenými predpokladmi odmietnuť.

Chomskeho vplyv na lingvistickú komunitu dlho brzdil diskusiu o evolúcii jazyka (Corballis, 2009), keďže Chomskeho postoj bol, že zvieracia komunikácia a ľudský jazyk sú založené na úplne odlišnom princípe, pričom zároveň odmietal, že by sa jazyk nutne musel vyvinúť v procese prirodzeného výberu. Opačný názor predostreli Pinker a Bloom (1990), pričom sa opierali o argument, že jediné racionálne vysvetlenie vzniku komplexných štruktúr je pomocou prirodzeného výberu a že jazyk je takým komplexným komunikačným systémom. Tým sa prelomili ľady a s teóriami sa roztrhlo vreco.

Všeobecne sa evolúcia jazyka vníma ako dvojstupňový proces (Perfors,

2002; Gärdenfors, 2004). Najprv sa musela vyvinúť jednoduchšia komunikácia, ktorá zahŕňala schopnosť reprezentovať objekty a javy v mysli i pomocou komunikačných jednotiek (symbolov, signálov) a primitívnym spôsobom ich komponovať. Tento tzv. protojazyk mal bohatú slovnú zásobu, no veľmi primitívnu syntax a prirovnáva sa k pidgin jazykom, reči dvojročného dieťaťa, či komunikačným schopnostiam niektorých tréňovaných zvierat. Niekedy sa hovorí v tejto súvislosti o evolúcii reprezentácie, či evolúcii komunikácie. Až následne došlo k evolúcii syntaxe, ktorá sa často skúma oddelene.

Toto rozdelenie je dôležité napríklad u Bickertona (1995), ktorý vidí evolúciu protojazyka ako postupný proces a vznik syntaxe ako náhly evolučný skok vyvolaný mutáciou, ktorá zmenila štruktúry v mozgu. Syntax sa vyvinula v priebehu niekoľko málo generácií niekedy pred 100000 až 40000 rokmi a viedla k oddeleniu Homo Sapiens od Homo Erectus. Bickerton sa opiera o fosílné nálezy z neskorého paleolitu (Perfors, 2002), ktoré dokazujú, že v tomto období došlo k rapídному rozvoju výroby zbraní a nástrojov, umeleckých a kultových predmetov, pričom tieto nálezy vykazujú regionálne odlišnosti naznačujúce vznik kultúrneho prenosu a teda i jazyka. Bickerton tiež vidí paralelu medzi ontogenézou a fylogenézou a porovnáva rapídny nástup syntaxe u ľudského rodu s rýchlym nástupom syntaxe (zopár mesiacov) u dvojročného dieťaťa.

Alternatívnou teóriou k biologickej evolúcii mozgových štruktúr je kultúrna evolúcia samotného jazyka (Deacon, 1997; Perfors, 2002). Jazyk ako predmet kultúrneho prenosu sám podlieha prirodzenému výberu a ľudia na komunikáciu uprednostňujú také prostriedky, ktoré sú ľahšie zvládnuteľné ich kognitívnym aparátom (Kirby, 2002). To je dôvod, prečo sa jazyk tak ľahko učíme. Rýchly nástup sa dá vysvetliť vyššou rýchlosťou Lamarckovskej kultúrnej evolúcie. Deaconova teória však úplne vynecháva vplyv biologického prirodzeného výberu.

Až na znakovú reč hluchonemých sú ľudské jazyky výlučne vokálne. To však nemusí nutne znamenať, že to tak bolo vždy (Corballis, 2009; Arbib, 2005). Teória, že hovorený jazyk sa vyvinul z gestikulácie je veľmi stará a nadväzuje na skúsenosti cestovateľov, že gestami sa dá dohovoriť prakticky všade. V nedávnej dobe získala novú podporu vďaka objavu zrkadliacich neurónov, ktoré reagujú pri rozpoznávaní i vykonávaní akcie, v oblasti kôry makakov homologickej s ľudskou Brocovou oblasťou (Rizzolatti a Arbib, 1998). Prisudzuje sa im podiel na schopnosti imitácie, ktorá je základom kultúrneho prenosu správania. Z imitácie manuálnych akcií sa vďaka pohybu po dvoch končatinách a uvoľneniu rúk vyvinuli primitívne komunikačné gestá, sprevádzané zrejme aj výrazmi tváre. Po zdokonalení vokálneho traktu narástol význam zvukovej modality pri percepcii a rozpoznávaní komunikačného aktu, čím sa otvorila cesta plne vokálnemu jazyku.

V neposlednom rade sa úvahy zaoberajú otázkou evolučných tlakov a podmienok, ktoré umožnili, alebo si priam vynútili vznik toho, či onoho aspektu jazyka (Gärdenfors, 2004). Keďže komunikácia má význam iba v situácii, keď majú obe strany interakcie nejaký spoločný záujem a ustanovenie štandardu sa oplatí iba, ak sa také situácie vyskytujú často, zdôrazňuje sa väčšinou nutnosť spolupráce vrámci skupiny ako významná evolučná podmienka. Zdokonaľovanie sa v oddelených reprezentáciách pre potreby komunikácie postupne viedlo k oddeleným reprezentáciám cieľov, schopnosti predstaviť si možné budúce či spomenúť na minulé potreby a sledovať fiktívne ciele, čo poskytlo evolučnú výhodu v podobe dokonalejšieho plánovania.

Súťaž medzi jedincami sa v kontexte evolúcie jazyka zvyčajne považuje za kontraproduktívnu, keďže jedna strana má pričasto motiváciu akúkoľvek dôveru a čestnú komunikáciu zneužiť a druhá má teda záujem nekomunikovať. Napriek tomu sa nedá predpokladať, že by sa predkovia ľudí ocitli v situácii, kde by sa súťaž vrámci druhu stratila. Vždy existovala ako v podobe súboja medzi skupinami tak v podobe boja o postavenie v skupine. Knight (1998) ponúka riešenie tejto otázky, podľa ktorého si skupiny slabších jedincov vyvinuli spôsoby ako sa kolektívne brániť pred zneužívaním zo strany silnejších tým, že kolektívne komunikovali nepravdivé obsahy voči okoliu. Vznikli tak akési rituály, kolektívna pretváarka, ktorú samotní členovia skupiny nepokladali za dôvod navzájom si nedôverovať, museli ju teda vedieť prehliadať, zároveň ju však museli vedieť odhaliť u outsiderov. Rituály ako nástroje na rozoznávanie dôveryhodnosti sa postupne vyvíjali a schopnosť nevinnej pretváarky umožnila vznik metafory, gramatikalizácie (premeny plnovýznamových slov na pomocné) a podobne, čo viedlo ku zvýšeniu komplexnosti a rôznorodosti komunikácie postupne až na úroveň dnešného jazyka.

### 1.3 Jazyk

Lingvistická tradícia rozdeľuje javy v jazyku do niekoľkých úrovní podľa toho, akých drobných častí jazyka sa týkajú. Fonológia sa zaoberá javmi na úrovni hlások, morfológia spôsobom tvorby slov, syntax pravidlami skladania slov do väčších celkov, sémantika vzťahom medzi slovami a významami a pragmatika cieľom komunikačného aktu v širšom kontexte. Každý jav sa potom skúma nezávisle, pričom sa predpokladá, že k pochopeniu reči dochádza jej postupným spracovaním od nižších úrovní po vyššie a k produkcii opačným procesom, od cieľa a významu, ku slovám a hláskam.

Kognitívna psychológia definuje nasledovné podmienky nutné pre to, aby sme jav mohli považovať za jazyk (Willingham, 2007):

- **komunikatívnosť** - umožňuje aspoň dvom jedincom komunikovať, vy-

jadriť sa a navzájom si a porozumieť.

- **arbitrárnosť** - neexistuje žiadna zjavná podobnosť, súvislosť medzi symbolmi a významami, ktoré reprezentujú, mapovanie medzi nimi by mohlo byť ľubovoľné iné a fungovať rovnako dobre.
- **štruktúrovanosť** - existuje súbor pravidiel, podľa ktorých sa symboly smú kombinovať na vyjadrenie rôznych významov. Odchýlka od pravidiel často znamená zmenu alebo stratu významu.
- **generatívnosť** - možnosť skladať symboly do väčších celkov majúcich význam neobmedzeným počtom spôsobov
- **dynamickosť** - podliehanie zmene v čase, vývoj.

Mozgové centrá zodpovedné za jazyk boli objavené pomerne dávno skúmaním mozgov pacientov so získanou poruchou reči, afáziou. Rozlišujeme Wernickeho a Brocovu afáziu. Wernickeho afázia postihuje porozumenie reči, prejavy siahajú podľa stupňa poškodenia od občasného používania nesprávnych slov až po slovný šalát. Pacienti nie sú schopní sami rozoznať, že použili nesprávny výraz a poruchu si často neuvedomujú. Brocova afázia postihuje produkciu reči, takže pacienti majú problém súvisle rozprávať a svoj problém si plne uvedomujú. Wernickeho a Brocova oblasť sú priamo prepojené oblúkovým zväzkom (arcuate fasciculus), ktorého poškodenia majú za následok ďalšie druhy afázie.

Napriek tomu, že neurálne koreláty jazyka poznáme už pomerne dlho, neobjavili sme doteraz žiadne konkrétne štruktúry, ktoré by riadili jednotlivé úrovne jazyka tak, ako ich rozlišuje lingvistika. Je otázne, či je toto rozdelenie vôbec nejako biologicky opodstatnené, prípadne, či je dôležité pri skúmaní evolúcie jazyka.

Objavom, ktorý v poslednej dobe vyvolal veľkú pozornosť, sú takzvané zrkadliace neuróny (mirror neurons). Ich definujúcou vlastnosťou je, že pália vtedy, keď jedinec nejakú konkrétnu akciu úmyselne vykonáva, bez ohľadu na to, kam sa pozerá, ale tiež keď túto akciu pozoruje v prostredí, bez ohľadu na to, či ju vykonáva on sám, alebo niekto iný. Pôvodne boli objavené v oblasti F5 kôry makakov (Pellegrino a kol., 1992), ktorá je vývojovým homológom Brocovej oblasti u ľudí. Ich funkcia u makakov sa dáva do súvislosti s rozpoznávaním akcií, najmä takých, ktoré sú zacielené na nejaký predmet, napríklad rôzne druhy úchopu. Umožňujú tým svojim majiteľom akoby rozpoznať stav mysle, úmysel iného jedinca. (Arbib, 2005) na základe tohto objavu sformuloval hypotézu, že práve evolučné objavenie sa zrkadliacich neurónov otvorilo cestu takým javom ako je imitácia, manuálny protojazyk zložený z gest, a neskôr ľudská reč.

Popri zrkadliacich neurónoch pre akcie sa v F5 nachádzajú aj tzv. kanonické neuróny, ktoré reagujú v situácii, keď jedinec nejakú akciu môže fyzicky vykonať. Sú zapojené v okruhu pre vykonávanie akcie, páliace kanonické neuróny akoby nám teda poskytovali repertoár akcií, ktoré v danom momente sme schopní vykonať (Arbib, 2005). Zrkadliace neuróny nie sú priamo zapojené v tomto okruhu, ale sú jeho nepriamou súčasťou. Mnohí, napríklad Miall (2003); Tani a kol. (2004) predpokladajú, že realizujú takzvané dopredné modely (Wolpert a Kawato, 1998), t.j. umožňujú nám predpovedať dôsledky akcie, prípadne dopĺňať si (predstavovať si) chýbajúcu časť spätnej väzby počas vykonávania akcie.

Neskôr boli objavené zrkadliace neuróny aj v iných oblastiach kôry makakov i u ľudí (viď napríklad Farkaš a kol., 2011). V iných oblastiach sa podieľajú na rozpoznávaní iných mentálnych stavov, napríklad v amygdale reagujú pri pociťovaní emócie ako aj pri pozorovaní iných jedincov prejavujúcich túto emóciu (Gallese a kol., 2004), pri netranzitivných akciách (nezameraných na cieľ, napríklad mávanie rukami), či pri akciách, ktoré nemusia mať na prvý pohľad zmysel, napríklad pantomíma. Ďalej sa zistilo, že niektoré zrkadliace neuróny reagujú na širší okruh akcií ako iné, hovoríme, že sú široko alebo úzko kongruentné (Rizzolatti a Sinigaglia, 2010). Zrkadliace neuróny pre konkrétne slová u ľudí zatiaľ objavené neboli.

Podľa niektorých nie sú zrkadliace neuróny ničím zvláštne (Hickok, 2008; Heyes, 2010; Hickok a Hauser, 2010). Schopnosť zrkadliť sa podľa nich nevyvinula evolučne, ale je prejavom obyčajného asociatívneho učenia na základe vykonávania a pozorovania akcií v kombinácii s hebbovským princípom. Otázka úlohy evolúcie a funkcie zrkadliacich neurónov v kontexte jazyka je stále stredobodom diskusie (Farkaš a kol., 2011).

Existuje relatívna zhoda v tom, aké vlastnosti musí kognitívny aparát organizmu mať, aby umožnil svojim nositeľom vyvinúť si, či používať jazyk, alebo aspoň jeho jednoduchších predchodcov (Christiansen a Kirby, 2003; Arbib, 2005):

- **Arbitrárnosť symbolov** – symboly nemajú nič spoločné s tým, čo denotujú, nie sú vrodené a dajú sa preučiť. Je to jedna z hlavných vlastností jazyka. Súvisí s prechodom od navodených reprezentácií ku oddeleným, od signálov k symbolom. Symbolové prístupy predpokladajú predspracovanie vstupu do podoby amodálnych symbolov, v ktorých prebieha myslenie. Perceptuálne ukotvené symboly sú naviazané na konkrétne modálne reprezentácie vstupov a akcií. (Barsalou, 1999) Prepojenie medzi vstupmi a výstupmi v neurónovej sieti sa tiež môže vyvinúť v princípe ľubovoľné, ak sa istý symbol (pomenovanie, akcia) bude vyskytovať často súčasne s nejakým javom (objektom, nekomu-



nikaçnou akciou), vytvorí sa prepojenie. Možno teda argumentovať, že mozog má túto schopnosť automaticky. Otázka je skôr, ako došlo k tomu, že sa ikonické signály, ktoré sú zrejme robustnejšie vo vyvolávaní správnej reprezentácie i ľahšie naučiteľné, zmenili na arbitrárne symboly, čo k tomu predkov ľudí viedlo a akú výhodu im to poskytlo.

- **Zdieľaná pozornosť** (joint attention) – schopnosť pochopiť, na aký objekt alebo jav má upriamenú pozornosť iný jedinec. Môžeme ju odstupňovať do troch úrovní podľa miery aktívnej snahy oboch agentov skoordinať pozornosť (Kwisthout a kol., 2008). Pri kontrolovanej zdieľanej pozornosti (checked attention) si agent iba skontroluje, či zameriava pozornosť na to isté, ako iný agent. Nasledovaná pozornosť (follow attention) znamená, že agent odhadne, na čo zameriava pozornosť iný agent, a usmerní svoju pozornosť tak isto. Usmerňovaná pozornosť (directed attention) je taká, pri ktorej agent aktívne navedie druhého, na čo má zacieliť pozornosť. Vyššie úrovne umožňujú efektívnejšiu akvizíciu slovnej zásoby. Evolučne zrejme súvisí s potrebou vykonávať spoločné akcie (joint action) pre zabezpečenie prežitia (Sebanz a Knoblich, 2009; Knoblich a Sebanz, 2008). Dáva sa do úzkej súvislosti aj s teóriou mysle.
- **Učenie pomocou imitácie** – schopnosť odkopírovať postupnosť akcií vedúcich v istej situácii k istému cieľu od iného agenta. Niekedy sa delí na jednoduchú a komplexnú imitáciu podľa toho, o aké zložité sekvencie akcií ide. Zadefinovať jasnú hranicu je však ťažké. Otvára cestu kultúrnemu prenosu informácie a vzorcov správania. Podiel na nej sa prisudzuje zrkadliacim neurónom (Arbib, 2005; Borenstein a Ruppín, 2004, 2005; Oztóp a kol., 2005, 2006). Predpokladom učenia pomocou imitácie je zrejme aspoň základná forma zdieľanej pozornosti.
- **Teória mysle** – vnímanie druhých agentov ako intencionálnych bytostí, vcítenie sa za účelom pochopenia ich motivácií a presvedčení a predpovedania ich akcií. Úzko súvisí so schopnosťou usudzovať o vnútorných stavoch mysle, emóciách, intenciách a cieľoch iných agentov na základe vonkajších náznakov (ich konania a kontextu), empatiou. V kontexte jazyka sa podieľa hlavne na pochopení jazykovej pragmatiky. Predpokladá sa, že jej poškodenie je príčinou väčšiny problémov u autistov (Baron-Cohen a kol., 1985). Tí často vykazujú problémy aj s akvizíciou jazyka. Je otázne, do akej miery možno prisúdiť schopnosť teórie mysle zvieratám, napriek tomu o komunikácii, i keď len veľmi základnej, môžeme hovoriť zrejme pomerne často (Gärdenfors, 2004). Podrobnejšie sa jej venujeme v nasledujúcej podkapitole.

- **Prevod medzi sekvenčnou a hierarchickou informáciou** – schopnosť rozpoznať hierarchickú gramatickú a sémantickú štruktúru vety a prevádzať ju na sekvenčnú symbolovú (zvukovú) a naopak. Súvisí so schopnosťou rekurzie, ktorá sa pozorovala už u istých komplikovaných manuálnych činnostiach goríl (Byrne, 2003). Niektorí výskumníci ju považujú za nadbytočnú podmienku a ukazujú, že explicitný prevod medzi sekvenčnou a hierarchickou reprezentáciou informácie nie je na rozpoznávanie a generovanie syntaxe potrebný. (Frank a Bod, in press)
- **Arbitrárne plánovanie** – schopnosť uprednostňovať fiktívne dlhodobé ciele pred pociťovanými krátkodobými popudmi, rozpoznávať možné budúce hrozby a príležitosti a pripravovať sa na ne vopred. Arbitrárne je preto, lebo stanovený cieľ nemusí nijako súvisieť s aktuálnou situáciou, môže byť úplne vymyslený, pre jedinca dokonca kontraproduktívny. Súvisí s myšlienkou oddelených reprezentácií cieľov. Niektorí ho považujú za schopnosť, ktorá umožnila vznik kultúry ako takej (Prudkov, 2005; Gärdenfors, 1995, 1996, 2004).

## 1.4 Teória mysle

Ide o predstavu, že pri interakcii s niektorými druhmi objektov sa nám oplatí predpokladať, že konajú na základe vnútorných stavov akejsi mysle, ktorá funguje podobne ako naša, ale je v inom stave (Premack a Woodruff, 1978). Nejde teda o teóriu vo vedeckom zmysle slova, ale o teóriu, ktorú si vytvára jedinec počas života na základe interakcií so svetom. Kvalitná teória mysle mu umožňuje rozpoznať stav mysle druhého jedinca na základe jeho prejavov a predpovedať jeho akciu.

Koncept teórie mysle navrhol Baron-Cohen a kol. (1985) ako vysvetlenie príčiny mnohých problémov u ľudí s rôznymi formami autizmu. Autisti majú problém rozlíšiť medzi stavom vlastnej a cudzej mysle, a keď už to zvládajú, majú problém prisudzovať iným mysliam správne stavy. Z toho potom plynú celý rad ďalších symptómov a problémov v rôznych sociálnych situáciách. Treba však podotknúť, že absencia teórie mysle nevysvetľuje všetky prejavy.

Medzi konkrétne prejavy, ktoré sú u autistov typické, patrí znížená schopnosť imaginácie, čo sa prejavuje napríklad nízkym záujmom hrať sa v detstve s inými ľuďmi hry s predstieraním (pretend play), t.j. také, kde dieťa alebo dospelý predstiera, že banán je napríklad telefón, hra na doktora a podobne (?basnakova2010). Mladý autista má problém vcítiť sa do pocitov človeka v situácii, ktorú sa druhé dieťa snaží predstierať. S tým súvisia ďalšie problémy autistov so správnym pochopením, používaním a rozlišovaním medzi koncep-

tami ako lož, metafora, irónia a podobne, so zvládnutím sociálnych rituálov, chápaním a rešpektovaním sociálnej hierarchie, vzťahov, intríg a podobne, teda všetkých situácií, kde niektorá strana interakcie predstiera, dáva vzhľadom na zaužívaný úprimný štandard navonok najavo iný stav, obsah mysle, než je jej skutočný skrytý, bez ohľadu na to, či je cieľom skutočný obsah mysle v konečnom dôsledku nejako odhaliť, alebo nie.

Významným znakom je atypický, pomalší až žiadny vývin reči. Jedinci s Aspergerovým syndrómom mávajú často problém s prozódiami, ako pri produkcii, tak pri chápaní, monotónny a dôsledne presný prejav, nevhodnú hlasitosť či rytmus, doslovné chápanie obrazných výrazov, problémy so striedaním sa v úlohe hovorca a poslucháča (skákanie do reči, dlhé pauzy), pochopením a uvedením kontextu nejakej informácie a ďalšími pragmatickými aspektami jazyka. Medzi ďalšie zaujímavé súvisiace prejavy patria neochota vytvárať očný kontakt, a tým pádom porušená schopnosť zdieľania pozornosti, celkovo poškodená neverbálna komunikácia (výraz tváre, poloha tela, gestá) a častá prítomnosť alexitímie (zníženej schopnosti chápať, spracúvať a popísať vlastné emócie), aj keď nie je jasné, či ide o dôsledok, alebo príčinu autizmu, alebo majú spoločnú príčinu (Knoblich a Sebanz, 2004).

Na testovanie, či nejaká osoba má teóriu mysle, sa používa tzv. test mylných presvedčení (false-belief test) (Bašnáková a Rybár, 2010). Podstatou je dostať subjekt do situácie, kedy reakcia na situáciu ukáže, či subjekt rozlišuje medzi vlastným a cudzím stavom mysle, alebo nie. Test existuje v dvoch verziách. Prvá verzia, "Sally-Anne test", spočíva v tom, že pozorujeme subjekt v situácii, kde sa udeje nasledovná postupnosť udalostí:

1. subjektu (väčšinou dieťa) predostrieme situáciu, kde si iné dieťa (Sally) schová nejakú hračku na miesto X.
2. Sally odíde z miestnosti a príde tam druhé dieťa (Anne). Nájde Sallyinu hračku a ukryje ju niekde inde (Y).
3. do miestnosti sa vráti Sally. Experimentátor sa snaží odhaliť (napríklad priamou otázkou), kde bude testované dieťa očakávať, že Sally bude svoju hračku hľadať. Ak subjekt odpovie X, usúdime, že chápe, že Sally nevie o presune hračky na miesto Y, a teda má teóriu mysle. Ak odpovie Y, zrejme si nevie oddeliť vlastnú reprezentáciu situácie od Sally-inej, nemá teda teóriu mysle.

Druhá verzia, nazývaná lentilkový test (Smarties test), alebo aj test vzhľadu a reality (appearance-reality test) zahŕňa menej osôb, ktorým treba prisudzovať stavy mysle:

1. subjektu ukážeme krabičku od lentiliiek, zahrkáme s ňou a spýtame sa, čo v nej je. Subjekt odpovie že lentilky.
2. otvoríme krabičku a ukážeme, že v krabičke sú v skutočnosti pastelky.
3. zavrieme krabičku s pastelkami naspäť a zavoláme ďalšiu osobu, ktorá doteraz nebola prítomná. Ukážeme jej krabičku a zahrkáme ňou. Spýtame sa subjektu, ako privolaná osoba odpovie na otázku, čo je v krabičke. Ak subjekt odpovie "lentilky", usúdime, že subjekt chápe, že privolaná osoba ešte nemá odkiaľ vedieť, že v krabičke je niečo iné, než sa zdá, teda má teóriu mysle. Ak odpovie "pastelky", tak ju nemá.

Všeobecná štruktúra by sa dala popísať:

1. Umožníme subjektu ustanoviť si reprezentáciu:  $Realita = CudziaMysel = X$
2. Pozmeníme realitu:  $Realita := Y$
3. Zistíme, čomu sa rovná  $CudziaMysel$

Bežné deti začínajú zvládať tento test vo veku okolo 4 rokov, autistické deti o niekoľko rokov neskôr. Pre porovnanie, deti s Downovým syndrómom, ktorý postihuje prakticky všetky aspekty inteligencie, s testom nemajú väčší problém než zdravé deti. Autistické deti nemajú problém zvládnuť pozmenenú verziu testu (camera test, photography test), v ktorej sa nepýtame na obsah mysle iného človeka, ale na obsah fotografie, ktorá bola urobená pred zmenou reality v bode 2.

Aby sa mohli lepšie vylúčiť iné dôvody, prečo subjekty test nemusia zvládať, napríklad nedostatočné pochopenie otázky v bode 3, a aby sa rozšírilo spektrum možných subjektov aj na veľmi malé deti a zvieratá, vznikli návrhy na alternatívne spôsoby odhaľovania očakávania subjektu. Príkladom je sledovať, kam subjekt automaticky zameria svoj zrak pri sledovaní a predpovedaní správania Sally. V tomto prípade zdravé deti úlohu zvládajú oveľa skôr. Tieto metódy sú však kritizované a na metódu prijateľnú širšou komunitou sa ešte len čaká.

Neurologická príčina porúch autistického spektra nebola zatiaľ presvedčivo vysvetlená. Po objave zrkadliacich neurónov sa vynorili okrem iných aj teórie o ich súvislosti s autizmom (Rizzolatti a Sinigaglia, 2010; Rizzolatti a Fabbri-Destro, 2008). Niektorí výskumníci považujú za základnú funkciu zrkadlenia schopnosť nahliadnuť do mysle iných jedincov, čo priamo súvisí s teóriou mysle. Niektoré neurologické výskumy im poskytujú nepriamu podporu. Napríklad v experimente (Rizzolatti a Sinigaglia, 2010), kde mali subjekty vykonávať a pozorovať dve podobné postupnosti akcií (natiahnuť sa pre

jedlo, chytiť ho a následne ho buď zjesť, alebo položiť do nádoby na pleci) sa zistilo, že autisti sa ako pri vykonávaní, tak pri pozorovaní pripravujú na otvorenie úst až v momente, keď sa jedlo blíži jednoznačne k ústam. Subjekty bez autizmu vykazovali elektrickú aktivitu v príslušných svaloch už počas fázy chytania jedla, aktivita opäť poklesla až následne v prípade, ak jedlo smerovalo na plece. Nešlo teda o výskum, ktorý by meral priamo aktivitu konkrétnych zrkadliacich neurónov. Ďalšie výskumy používajú štandardné zobrazovacie techniky (EEG, fMRI), v prípade ktorých skúmame vzory pálenie celých oblastí mozgu a hovoríme skôr o zrkadliacom systéme, než neurónoch (pozri Rizzolatti a Sinigaglia, 2010). Behaviorálne štúdie tieto hypotézy odmietajú. Okrem klasických zrkadliacich neurónov pre akcie sa autizmu týkajú aj neuróny v amygdale zodpovedné za zrkadlenie emócií. Keďže skúmanie pálenia konkrétnych neurónov je pri dnešnej technológii riziková operácia, na priamejšie dôkazy pre, či proti si budeme musieť ešte počkať.



# Kapitola 2

## Modelovanie emergencie jazyka

### 2.1 Prístupy

Rigoróznejšie prístupy k jazyku a jeho evolúcii než verbálne teoretizovanie, možno rozdeliť do dvoch kategórií: matematické modely a výpočtové modely. Ich aplikácie a motivácie majú často interdisciplinárny charakter. Formálnym matematickým modelom syntaxe a gramatiky je teória formálnych jazykov, ktorá bola inšpirovaná otázkami v lingvistike a silne ovplyvnila teoretickú informatiku. Matematický model prenosu informácie (čo je hlavný účel jazyka) poskytuje teória informácie. Vzťahmi medzi symbolmi, ich kombináciami a významami sa zaoberá formálna logika a sémantika. Jazyk ako komplexný dynamický systém skúmajú experti na teóriu chaosu a teóriu komplexných systémov (Logan, 2005). Skúmajú sa aj reprezentácie jazyka ako grafu, resp. siete, napríklad v súvislosti so sieťami malého sveta (Markošová, 2005). Matematické modely evolúcie sú založené na teórii hier a koncepte evolučne stabilných stratégií (Smith, 1982). Úlohu komunikácie v hrách explicitne skúmajú modely hier s možnosťou komunikácie pred prijatím rozhodnutia (cheap talk, pre-play communication) (Demichelis a Weibull, 2008).

Jazyk je tiež ústredným problémom umelej inteligencie od jej zrodu. Turingov test vo svojej pôvodnej podobe predpokladal, že umelý účastník testu zvláda plynulú komunikáciu v nejakom jazyku (Turing, 1950). Umelá inteligencia sa od počiatku snažila vytvoriť výpočtový model skutočného ľudského jazyka, čo vyústilo do vzniku samostatného odboru, výpočtovej lingvistiky. V tejto oblasti klasicky súťažila symbolová a subsymbolová paradigma. Symbolistický prístup vidí procesy v mysli, teda aj tie pri spracúvaní jazyka, ako ekvivalentné počítačovému výpočtu, v princípe ako transformáciu symbolových štruktúr. Pri modelovaní sa opiera najmä o poznatky z teórie formálnych jazykov a automatov a formálnej logiky a sémantiky, pričom často používa

logické programovanie na vyjadrenie procesov a stavov mysle a pravdepodobnostné metódy na odhad chýbajúcej, či predpoveď budúcej informácie. Subsymbologový, resp. konekcionistický prístup sa inšpiruje vlastnosťami skutočných štruktúr v mozgu. Neurónové siete umožňujú distribuovanú reprezentáciu konceptov a procesov. Miera plauzibilitnosti výpočtových neurálnych modelov je rôzna, nie vždy je prioritou. Z hľadiska teoretickej informatiky bola dokázaná ekvivalencia Turingovho stroja a istej podmnožiny rekurentných neurónových sietí ako modelov výpočtu. Pre každý turingov stroj existuje rekurentná neurónová sieť, ktorá ho dokáže realizovať. Iné neurónové siete sú funkčne ekvivalentné markovovským modelom, bayesovským klasifikátorom, a ďalším matematicky inšpirovaným modelom. Neurónové siete sú teda plnohodnotným modelom výpočtu, ktorý sa navyše svoj program dokáže s rôznou presnosťou naučiť na základe tréningových vstupov. V tomto zmysle sú neurónové siete heuristickou metódou, kde produktom je nepresný ale v praxi použiteľný program. Ani jeden z prístupov doteraz nevedol k modelu, ktorý by zvládol turingov test.

Procesom evolúcie sa inšpirovali evolučné algoritmy (Pelikan, 2009). Formalizujú koncepty ako replikátor (chromozóm, genóm), populácia, generácia, fitness, reprodukcia, mutácia a kríženie. Z hľadiska informatiky ide o paralelné prehľadávanie priestoru riešení, ktoré môže uviaznuť v lokálnom minime, ide teda o heuristickú metódu. Ak v evolučnom algoritme súťažia priamo riešenia problému zakódované v genóme, ide o genetický algoritmus. Ak sú v genóme zakódované celé programy, ktoré majú byť schopné nejaký problém riešiť, ide o genetické programovanie. Ďalšie modely, ktoré smerujú od púhej inšpirácie k biologickej plauzibilite, chápu zakódovaný program ako mentálny aparát agenta (či už symbolový, alebo subsymbologový), ktorého fitness sa vyhodnotí na základe úspešnosti v nejakom prostredí, ideálne takom, kde agenty s rôznymi mentálnymi aparátmi medzi sebou interagujú. Ak je v tomto prostredí navyše umožnená reprodukcia a smrť agentov takým spôsobom, že fitness je implicitná, hovoríme o umelom živote. Evolučné simulácie s autonómnymi agentmi sa považujú za vhodný model skutočnej evolúcie, ktorý je vzhľadom na to, že musí byť špecifikovaný tak, aby mohol prebehnúť na počítači a byť analyzovaný z informatického hľadiska, považovaný za dostatočne exaktný a rigorózný, a teda použiteľný na overenie hrubej plauzibilitnosti rôznych evolučných teórií (Ruppin, 2002). Multiagentové simulácie sa používajú aj na simulovanie procesov, v ktorých nie je prítomný prirodzený výber a evolúcia, aspoň nie nejakou explicitne.

Relatívne novým prístupom je skúmať vznik komunikačného štandardu v laboratóriu pri použití skutočných ľudí ako agentov (Scott-Phillips a Kirby, 2010). Tým sa v praxi odstráni potreba navrhovať model mysle a vyskúšajú sa schopnosti a limity skutočnej ľudskej mysle v prostredí podobne zjedno-



dušenom, ako v počítačových simuláciách. Otestujú sa tým tiež limity danej úlohy vo vzťahu ku skutočným ľuďom, a navyše sa odhalia stratégie, ktoré ľudia používajú na ich vyriešenie.

Keďže komunikácia zahŕňa interakciu minimálne dvoch jedincov, výpočtové simulácie, ktoré sa priamo zaoberajú vznikom komunikácie alebo syntaxe sú takmer vždy nejakým druhom multiagentovej simulácie. Možno ich rozdeliť podľa niekoľkých kritérií. Na jednej strane existujú modely, ktoré majú bližšie k symbolovej paradigme (Kirby, 2002; Christiansen a Kirby, 2003), prevažujú však modely, ktoré sa snažia o väčšiu biologickú plauzibilitu a kognitívny aparát je v nich reprezentovaný konekcionisticky (napríklad Bonaiuto a Arbib, 2010; Borenstein a Ruppín, 2005; Chaminade a kol., 2008; Kvasnička a Pospíchal, 1999; Steels, 1997). Môžu tak nadväzovať na výskum v oblasti neurolingvistiky. Z hľadiska toho, či ako na formovanie kognitívneho aparátu agentov vplýva evolúcia, môžeme rozdeliť modely na neevolučné, kde nejaký jav súvisiaci s jazykom emerguje čisto na základe lokálnych interakcií agentov jednej generácie, s biologickou evolúciou, s kultúrnou evolúciou, prípadne s biologicko-kultúrnou koevolúciou (porovnaj napríklad Borenstein a Ruppín, 2005; Kvasnička a Pospíchal, 1999; Kirby, 2002; Steels, 2005). Posledné kritérium posudzuje mieru situovanosti agentov v prostredí, teda to, či je výsledok interakcie a úspešnosť agenta závislá na vlastnostiach a interakcii s prostredím samotným, alebo je prostredie nepodstatné a ide len o samotných agentov (Wagner a kol., 2003). Situované experimenty potom môžeme rozdeliť na simulované v počítačovom prostredí, alebo vykonané s reálnymi fyzickými agentmi, väčšinou robotmi. Niektoré projekty sa snažia umožniť minimalizovať tento rozdiel a poskytujú ako fyzické roboty, tak ich verné počítačové simulácie, pričom obom stačí doplniť riadiaci aparát (Metta a kol., 2008).

## 2.2 Komunikácia a hry

V mnohých modeloch je základnou lokálnou interakciou medzi agentmi koncept jazykovej hry, z ktorej vznikne globálny jazykový jav (Steels, 1998; Cangelosi a Parisi, 1998; Belpaeme a Bleys, 2005). V nej jeden agent (hovorca) nejakým spôsobom komunikuje svoj vnútorný stav navonok, na čo musí zareagovať druhý agent (príjemca) nejakou správnou akciou spomedzi dostupných. Ak zareagoval príjemca správne, komunikácia bola úspešná a jeden, druhý, alebo oba agenty môžu byť odmenené, resp. ich reprezentácie nejakým spôsobom posilnené. Ak zareagoval nesprávne, odmena je nižšia až záporná, prípadne dochádza k zmene reprezentácií v prospech lepšieho úspechu v budúcnosti, učeníu.

Tabuľka 2.1: Koordinačná hra

	Akcia 1	Akcia 2
Akcia 1	1; 1	0; 0
Akcia 2	0; 0	1; 1

Konkrétne poznáme hru na hádanie (guessing game), kde hovorca zvolí objekt či význam prítomný nejakým spôsobom v prostredí a vyprodukuje jazykovú reprezentáciu tohto významu. Príjemca spracuje jazykový prejav hovorca a skúsi mu priradiť správny význam. Následne nájde význam v prostredí a poukáže naň nejakým nejazykovým prostriedkom (napríklad ukáže prstom). Ak ide o ten istý význam, ktorý mal na mysli hovorca, hra bola úspešná.

Zrkadlovou verziou je hra na pomenúvanie (naming game). Prvý agent poukáže na význam nejakým spôsobom a druhý ho skúsi pomenovať v jazyku. Ak by prvý agent význam pomenoval zhodne, hra bola úspešná. V opačnom prípade buď prvý agent svoj lexikón upraví smerom k lexikónu druhého hráča, alebo naopak koriguje druhého hráča, aby on upravil svoje mentálne reprezentácie, prípadne sa udeje oboje.

Z hľadiska teórie hier ide v oboch prípadoch o koordinačnú hru, teda takú, kde sú obaja hráči ako celok najviac odmenení v prípade, že zvolia zhodné stratégie. Množinou stratégií sú v prvom prípade (hra na hádanie) možné významy, v druhom prípade (hra na pomenúvanie) možné jazykové prejavy. Hráči sú odmenení, ak sa rozhodnú pre zhodnú stratégiu, u každého je však kódovaná iným spôsobom. Ešte zjavnejším príkladom je imitačná hra, kde prvý agent vykoná nejakú akciu a druhý ju musí čo najvernejšie napodobniť, tu je aj kódovanie rovnaké. Iné motivačné štruktúry

Čisto matematický výskum v oblasti teórie hier hľadá riešenia a vysvetľuje motivácie hráčov v podobe formálneho konceptu Nashových ekvilibrií, teda takých kombinácií stratégií, kde sa žiadnemu hráčovi zmena stratégie nevyplatí, prípadne hľadá z rôznych hľadísk dominantné alebo optimálne stratégie. Rozšírením o koncept evolúcie a prirodzeného výberu dostávame evolučnú teóriu hier, ktorá hľadá evolučne stabilné stratégie, čo sú také Nashove ekvilibriá, ktoré, keď sa raz v populácii dostatočne uchytili, odolajú invázii malého počtu jedincov s ľubovoľnou inou stratégiou (Smith, 1982). Úloha komunikácie v nich sa skúma pomocou formálnych modelov hier (Demichelis a Weibull, 2008), kde rozhodnutiu o stratégii predchádza možnosť hráčov komunikovať, rozprávať sa (pre-play communication), pričom táto komunikácia neovplyvňuje podobu výplatnej matice (cheap talk). Výskum sa

Tabuľka 2.2: Súboj pohlaví

	Akcia 1	Akcia 2
Akcia 1	5; 10	0; 0
Akcia 2	0; 0	10; 5

Tabuľka 2.3: Lov jeleňa

	Jeleň	Laň
Jeleň	2; 2	0; 1
Laň	1; 0	1; 1

podobne ako v simuláciách zameral vo významnej miere na koordinačné hry, napríklad hru nazývanú "súboj pohlaví" (battle of the sexes). Jej matica odmeňuje oboch hráčov, ak sa dohodnú na tej istej akcii, avšak pri jednej akcii je viac odmenený jeden hráč a pri druhej druhý. Ak sa nedohodnú, odmenena je pre každého horšia, než prípade ľubovoľnej dohody.

Zaujímavá je koordinačná hra známa ako lov jeleňa (stag hunt). Jej výplatná matica motivuje oboch hráčov ako celok vykonať tú istú akciu (loviť spoločne jeleňa), ale každý hráč riskuje, že ten druhý sa k nemu nepridá, a môže radšej uprednostniť individuálnu akciu (loviť laň osamote) a mať s istotou aspoň niečo. Ak budú obaja loviť laň, ulovia ju možno ľahšie (celkovo z nej získajú viac), ale budú sa o ňu musieť podeliť (individuálne menej). Spoluprácou na jeleňovi by získali viac. Tento model použil za základ svojej teórie evolúcie sociálnych štruktúr a spoločenskej zmluvy Skyrms (2004, 1996), ktorý najnovšie presunul svoj záujem na javy súvisiace s evolúciou komunikácie (Skyrms, 2010).

Najznámejšou takto študovanou hrou je väzňova dilema (prisoner's dilemma). Podobá sa na lov jeleňa, s tým rozdielom, že v prípade, že jeden hráč zradil, teda zvolil individualistickú stratégiu, získava individuálne viac, než keby spolupracoval (ako celok však stále získavajú menej). Pre každého hráča tak platí, že viac získa individualisticky, ako spoluprácou, bez ohľadu na to, ako sa rozhodne protivráč. Tým pádom sa racionálni egoistickí hráči navzájom podvedú, i keď by spoluprácou získali viac. Toto ekvilibrium nezmení ani pridanie možnosti komunikácie. V iterovanej verzii, to jest takej, kde podvedený hráč má možnosť odplatiť zradu v nasledovnej iterácii sa ukazuje ako najvýhodnejšia stratégia tit-for-tat, podľa ktorej začneme priateľsky, teda

Tabuľka 2.4: Väžňova dilema

	Spolupráca	Zrada
Spolupráca	5; 5	-10; 10
Zrada	10; -10	-5; -5

skúsime spolupracovať, a potom opakujeme predchádzajúcu akciu protihráča, oplácame mu zradu zradou a spoluprácu spoluprácou. Väžňova dilema si získala veľkú pozornosť, mnohí v nej hľadali matematický model altruizmu (Axelrod a Hamilton, 1981). Pravidelne sa organizujú turnaje s originálnymi i rôzne pozmenenými pravidlami, kde môžu záujemcovia navrhnúť jedného či viacerých ľubovoľne naprogramovaných hráčov a otestovať svoje stratégie v praxi.

Z uvedených dôvodov bola úloha komunikácie vo väžňovej dileme považovaná za nulovú. Ak má však tím možnosť navrhnúť do súťaže väčší počet hráčov a úspech tímu sa hodnotí podľa umiestnenia najlepšieho hráča tímu, ukazuje sa, že komunikácia môže hrať aj vo väžňovej dileme rolu (Rogers a kol., 2007). Hráči v jednom tíme môžu mať rozdelené úlohy na jedného vodcu, ktorý má za úlohu vyhrať pre tím víťazstvo, a ostatných radových členov, ktorí sa preňho maximálne obetujú. V kontakte vodca - radový člen teda vodca hrá individualisticky, zatiaľ čo člen spolupracuje. V kontakte dvoch radových členov z jedného tímu obaja spolupracujú. V kontakte vodca - cudzinec vodca hrá tit-for-tat, aby si uchránil čo najviac bodov, a v kontakte radový člen - cudzinec hrá člen agresívne, aby obral ostatné tímy o čo najviac bodov. Jediným problémom je rozpoznať, proti komu v súboji hráč stojí. Na to môžu slúžiť dve krátke vhodne zvolené dopredu dohodnuté sekvencie akcií na začiatku interakcie, pomocou ktorých spoľahlivo rozoznáme vodcu od radového člena a oboch od cudzinca. Tím má takto šancu, že vodca obsadí pozíciu blízko vrcholu (podľa toho, ako efektívne sú v realizácii nejakej podobnej stratégie iné tímy), za čo bude odmenený celý tím, napriek tomu, že jeho drivivá väčšina obsadí pozície veľmi hlboko. V prírode sa to podobá eusociálnemu hmyzu, kde novú generáciu vytvára len úzka elita, ktorú treba chrániť a kŕmiť, o čo sa stará neplodná väčšina kolónie.

Na väžňovu dilemu sa podobá aj hra jastrab-holubica (hawk-dove). Oproti väžňovej dileme je tu situácia, keď sa hráč (holubica) nechal využiť (jastrebom) preňho priaznivejšia, než keby sa s protihráčom stretol vo vzájomnej zrade. Kolektívne to dokonca môže byť výhodnejšie, než spolupracovať, čo je v tomto prípade skôr bezpečná, než víťazná voľba. Hráčom ide o to, dohodnúť sa, kto koho využije a nestretnúť sa pri tom omylom vo vzájomnej agresii.

Tabuľka 2.5: Jastrab - holubica

	Holubica	Jastrab
Holubica	5; 5	-5; 10
Jastrab	10; -5	-10; -10

Tabuľka 2.6: Diskoordináčna hra

	Akcia 1	Akcia 2
Akcia 1	1; -1	-1; 1
Akcia 2	-1; 1	1; -1

Hry tohto typu, kde účastníci sú kolektívne najviac odmenení v prípade, že zvolia protikladné, vzájomne sa akoby dopĺňajúce stratégie, sa nazývajú anti-koordinačné. Na koncept koordinačných a antikoordinačných hier nadväzuje koncept diskoordinačnej hry, v ktorej jeden hráč profituje z toho, keď si hráči stratégie skoordínujú, a druhý z toho, keď zvolia opačné stratégie. Dohoda v tomto prípade zjavne nie je možná.

## 2.3 Konkrétne modely

V tejto podkapitole sa pozrieme bližšie na niekoľko modelov, ktoré sú v kontexte cieľa práce nejakým spôsobom zaujímavé. Aj keď náš model nenadviaže priamo na všetky, dopĺňajú celkový pohľad na prístupy a výsledky v problematike. Z tabuľky možno vyčítať, s akým javom súvisiacim s jazykom a jeho emergenciou či evolúciou modely nejakým spôsobom súvisia. Bližšie sa zameriame najmä na model Borensteina a Ruppina (Borenstein a Ruppina, 2005), ktorý bol centrálnou inšpiráciou pre náš model popísaný v nasledujúcej kapitole.

### 2.3.1 Učenie a syntax (Kirby, 2002)

Kirby (2002) modeloval vznik syntaxe vo viacgeneračnom procese bez biologickej zmeny. Model je symbolový, významy sú logické výrazy a ich množina je preddefinovaná. Signálmi sú slová nad anglickou abecedou, generovanie signálov z významov zabezpečuje gramatika, v ktorej je použitie pravidla podmienené množinou významov, pri ktorých ho možno aplikovať.

Tabuľka 2.7: Vlastnosti skúmaných modelov

	Kirby	Steels	Kv.Posp.	Cang.	Bor.	Bon.	Oztop
vznik komunikácie	-	x	x	x	-	-	-
vznik syntaxe	x	x	x	x	-	-	-
sebapozorovanie	-	-	-	-	-	x	x
evolúcia	-	-	x	x	x	-	-
imitácia	-	-	-	-	x	-	x
teória mysle	-	-	-	-	-	-	x
zrkadliace neuróny	-	-	-	-	x	x	-
jazykové hry	x	x	x	x	x	-	-
situovanosť	-	x	-	x	-	x	-
telepatia	x	x	x	-	-	-	-

Medzigeneračný prenos sa modeloval v interakcii páru rodič - dieťa, pričom vrámci každej generácie existovali v skutočnosti len tieto dva agenty. Rodič bol vždy hovorca a odovzdával náhodné významy a im prislúchajúce signály v podobe reťazcov agentovi nasledujúcej generácie, ktorý ich používal ako učebné príklady. Ak náhodnému významu neprislúchal žiadny signál, vygeneroval sa špeciálnym algoritmom vhodný nový a naučili sa ho oba agenty, rodič aj dieťa. Učebný algoritmus v prvom rade skontroloval, či daná dvojicu patrí do nášho jazyka, a ak nie, pridal ju do gramatiky ako pravidlo. Potom upravil gramatiku tak, aby dokázala vygenerovať všetky doterajšie známe páry význam - signál, ale pre páry s podobnou štruktúrou významov a signálov mala spoločné pravidlá s neterminálmi. Tým je gramatika pripravená na zovšeobecnenie procesom subsumpcie, ktorý rozumne zjednotí pravidlá so zhodnými pravými stranami a zovšeobecní tak niektoré neterminály na ľavej strane. Gramatika potom začne generovať signály aj pre významy, ktoré predtým nestretla.

Príjemca sa v tomto modeli v podstate učí imitáciou. Prvý rodič začína s prázdnu gramatikou, všetky signály zo začiatku generuje nové. Tým vznikne jazyk s holistickými pomenovaniami pre komplikované situácie. Dieťa sa snaží doplniť si gramatiku tak, aby dokázala generovať všetky páry, ktoré dostala ako učebné vstupy (imituje správanie rodiča). Keďže dieťa nedostane ako

učebné vstupy všetky významy, ktoré má reprezentované rodič, málo používané holistické signály vypadnú z jazyka a neskôr ich nahradia iné. Pre niektoré podobné významy časom vzniknú podobné signály a môže dôjsť ku generalizácii a vzniku jednoduchej syntaxe. Kirby ďalej ukazuje, že po mnohých generáciách takto môžu vzniknúť aj zložitejšie gramatiky s rekurziou.

Za vznik syntaxe v tomto modeli je zodpovedný spôsob učenia a generalizácie. Kirby ukazuje, že kognitívny aparát agenta môže limitovať, aké štruktúry mapovania medzi symbolmi a významami sa budú efektívnejšie kopírovať z generácie na generáciu. Ak má raz agent kognitívny aparát s vhodnou metódou učenia a generalizácie, nie je ďalej nutný podiel biologickej evolúcie, aby sa za niekoľko generácií z ničoho vyvinul jazyk so syntaxou. Jazyky so syntaxou prechádzajú bottleneckom, ktorý tvorí Kirbyho metóda učenia, lepšie, než idiosykratické, holistické jazyky.

Model uvažuje telepatiu - priamy prístup dieťaťa do mysle rodiča, miesto toho, aby si dieťa muselo zamýšľaný význam odhadnúť samo, potenciálne nepresne. Model neuvažuje ani prenos a zmenu jazyka vrámci jednej generácie, ani žiadnu formu súťaže. Kirby sa nezaujíma o to, ako by mohla byť ním popísaná metóda učenia implementovaná v mozgu a ako by taká implementácia mohla vzniknúť, resp. aké iné učebné algoritmy, s akými vlastnosťami, by mohli viesť ku vzniku syntaxe. Existenciu úzkeho profilu pri učení považuje za dostatočnú pre vyššiu úspešnosť syntaktických jazykov. Model čiastočne podporuje Deaconovu teóriu evolúcie jazyka kultúrnym prenosom, aj keď vynecháva prirodzený výber.

### 2.3.2 Talking heads (Steels, 1998)

V robotickom modeli "rozprávajúcich sa hláv" (Talking heads) emerguje syntax vrámci jednej generácie následkom opakovaných jazykových hier medzi dvoma hlavami, ktoré sledujú dynamické deje v nezávislom robotickom ekosystéme (Steels, 1998). Hlavy sa nepohybujú a sú umiestnené na opačných koncoch ekosystému, vidia ho teda z rôznych uhlov. Hrajú sa imitačné, rozlišovacie a pomenúvacie hry. Imitačné vedú ku vzniku spoločného repertoára zvukových prejavov (z ktorých sa budú skladať jazykové výrazy, ak komunikácia ide zvukom), rozlišovacie k rozpoznávaniu odlišností v prostredí a samoorganizácii mapovania medzi priestorom vstupov a priestorom konceptov. Pomenúvacie hry prebiehajú nasledovne: kognitívny aparát hlavy spracuje zmenu vizuálneho vnemu v priebehu istej doby, teda pohyb robotov v ekosystéme, a rozpozná v ňom niekoľko konceptov. Jeden z nich si zvolí za tému rozhovoru a vygeneruje zodpovedajúci jazykový prejav. Ten v opísanom modeli síce ide cez sieť, ale v princípe by mohol ísť aj zvukom, ak by mali hlavy implementovaný modul na spracovanie zvuku (vtedy by bola

dôležitá imitačná hra). Druhá hlava skúsi dekodovať prijatý jazykový prejav a výsledok sa porovná so zvolenou témou rozhovoru, na základe čoho potom prebehne učenie u oboch agentov. Po čase si hlavy vyvinú jazyk so syntaxou. Čím dlhšie sú obdobia, po ktoré sledujú hlavy scénu medzi jednotlivými kolami jazykovej hry, tým komplikovanejšie javy sa na scéne vyskytujú a hlavy si vytvárajú komplikovanejšiu syntax.

Steels stanovuje hlavné princípy, ktoré by mal model spĺňať: postupný rast komplexity javov i jazyka; adaptívnosť jazykových hier, aby dochádzalo k zvyšovaniu úspešnosti; selekcionizmus, ktorý sa deje na úrovni syntaktických a významových štruktúr, pretrvávajú tie, ktoré vedú k úspechu v jazykových hrách a efektívnejšie sa prenášajú; formovanie hierarchických úrovní procesom, v ktorom sa nové štruktúry budujú zo starších, vytvorených počas predchádzajúcich hier; a samoorganizácia, ktorá je priebežná, reaguje na zmeny v prostredí a jazyku a sama ich spôsobuje, vďaka čomu sa systém neustále vyvíja.

Model je situovaný vo fyzickom prostredí, používa subsymbolové reprezentácie, no nutne na nich netrvá a nesnaží sa byť po tejto stránke biologicky plauzibilný. Na vyhodnotenie úspešnosti v jazykových hrách používa telepatiu a nepredpokladá biologickú evolúciu, ani medzigeneračný prenos, všetko sa deje vrámci jednej generácie. Čiastočne podporuje Deaconovu teóriu.

### 2.3.3 Baldwinov efekt (Kvasnička a Pospíchal, 1999)

Táto simulácia (Kvasnička a Pospíchal, 1999) skúma vznik komunikácie v procese zmiešanej genetickej a memetickej evolúcie. Gény aj mémy sú reprezentované v jednom chromozóme a kódujú potenciálne rekurentnú architektúru neurónovej siete, ktorá prevádza interné stavy agenta na sekvencie znakov a naopak. Gény, ktoré sa dedia darwinisticky, kódujú prepojenia medzi neurónmi, mémy predstavujú úvodnú učebnú sadu, ktorú agent absolvuje pred spustením jazykových hier a vyhodnocovania fitness. Časť chromozómu, v ktorej sú kódované mémy, sa počas života agenta mení a nasledovná generácia dedí zmenenú štartovaciu učebnú sadu (Lamarckovská evolúcia). Fitness sa vyhodnocuje v sérii komunikačných interakcií analogických k jazykovým hrám na hádanie či pomenúvanie, pričom sa postupne prejde celá množina možných interných stavov. Hovorca daný stav prekóduje do sekvencie znakov, prijímateľ tento reťazec naspäť prekóduje do interného stavu a následne sa porovnávajú interné stavy. Na základe toho môže dôjsť jednak k adaptácii neurónovej siete, k zmene memetického obsahu chromozómu, a k zvýšeniu fitness agenta. Po kombinácii každého agenta s každým sa na základe fitness vygeneruje nová generácia, pričom môže dôjsť k mutácii na chromozóme.

V modeli dochádza k tzv. Baldwinovmu efektu, keď nejaká dedičná in-



formácia, ktorá sa najprv prenáša memeticky, sa za istých okolností po dostatočnom počte generácií prekóduje do genetickej podoby. Model opäť vyhodnocuje úspech v hre a poskytuje učebné vstupy telepatickým prenosom informácie o interných stavoch. Učenie prebieha spätným šírením chýb. V modeli emerguje gramatika takého charakteru, že pravidelnosti vo vektoroch interných stavov sa odrážajú v pravidelnostiach v ich jazykových reprezentáciách.

### 2.3.4 Fungus eaters (Cangelosi a Parisi, 1998)

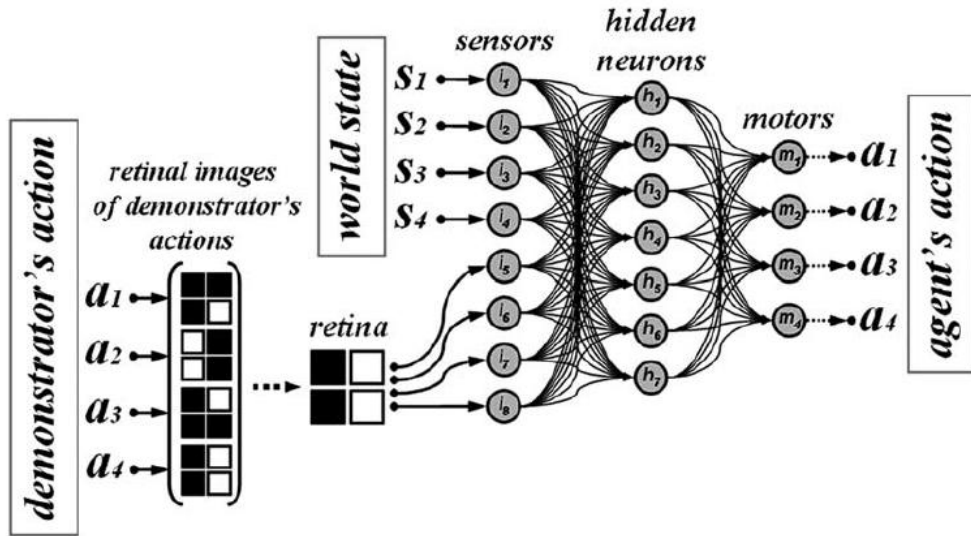
Model situuje skupinu 80 agentov riadených doprednými neurónovými sieťami do prostredia "sveta húb", kde je ich fitness vyhodnocovaná podľa schopnosti pozbierať za život čo najviac jedlých húb a vyhnúť sa jedovatým (Cangelosi a Parisi, 1998). Pri identifikácii húb si môžu agenti navzájom pomáhať a vymieňať informácie, musia si však vyvinúť komunikačný štandard. Jedinci majú pevný repertoár akcií, a vykonávajú tri druhy úloh (hier), pričom hovorca je vždy agentov rodič: počúvanie (analógia hry na hádanie, agent musí vykonať akciu, ktorú pomenoval hovorca), pomenúvanie (agent používa verbálny výstup hovorcu ako učebný vstup pre seba, samotný verbálny výstup nie je na vstupe prijímateľa) a imitáciu (totožné s pomenúvaním, len agent pred imitáciou dostane na vstup aj verbálny prejav hovorcu, ktorý má zopakovať).

V modeli emerguje základná kompozicionalita – agenti komunikujú pomocou dvojslovných viet. V simulácii agenti samovoľne vytvárajú kultúry, ktoré sa líšia vyvinutým štandardom. Autori pozorovali v modeli jav pripomínajúci Baldwinov efekt – genetickú asimiláciu toho, čo sa bolo pôvodne treba učiť. Interakciu kultúrnej variácie a Baldwinovho efektu dochádza k emergencii kompozicionality.

Tento model je veľmi komplexný, zahŕňa mnohé z javov, na ktoré sa chceme sústrediť. Imitácia je však do modelu vložená a priori, tento základ kultúrneho prenosu sa sám nemusí vyvinúť.

### 2.3.5 Evolúcia imitácie (Borenstein a Ruppin, 2005)

Borenstein a Ruppin (2003) sa venuje vplyvu učenia pomocou imitácie na evolučný proces. Navrhol evolučný model vývoja schopnosti učenia sa pomocou imitácie (Borenstein a Ruppin, 2004, 2005) u agentov riadených doprednou neurónovou sieťou s hebbovským učením. U časti z vyvinutých agentov sa objavili štruktúry s vlastnosťami zrkadliacich neurónov na skrytej vrstve.



Obr. 2.1: Borensteinov a Ruppertov (2003) model neurálnej architektúry agenta.

Život agenta sa skladal z kôl, úlohou agenta bolo v každom kole priradiť aktuálnemu stavu sveta na vstupe správnu akciu na výstupe. Niekedy nebol stav sveta viditeľný, vtedy mal ostať agent nečinný. Skóre, ktoré sa agentovi udelilo, sa počítalo ako 1 mínus stredná kvadratická odchýlka vygenerovaného a očakávaného výstupu. Okrem toho niekedy dostával agent na vstupe aj sensorickú reprezentáciu akcie demonštrátora. Tá bola vždy správna (ak bola viditeľná) a mala agentovi slúžiť ako vzor, od ktorého sa učí imitáciou. Počet možných stavov sveta bol 4, rovnako ako počet možných akcií. Nasledovný stav sveta nijako nezávisel od predchádzajúceho, ani od vykonanej akcie. Jeden beh života agenta pozostával z 500 kôl, pričom skóre sa počítalo iba za posledných 400 kôl, prvých 100 kôl mal agent na učenie. Sensorická reprezentácia akcie demonštrátora sa líšila od motorickej, zobrazenie medzi nimi bolo však pevné. Zobrazenie medzi stavmi sveta bolo v každom behu života agenta rôzne. Dôvodom bolo, aby sa toto mapovanie nemohlo vyvinúť evolučne. V každej generácii sa agent odskúšal na niekoľkých behoch, aby sa presnejšie odmerala schopnosť agenta zvládať úlohu.

Dopredná neurónová sieť agenta obsahovala jednu skrytú vrstvu so siedmimi neurónmi. Vstupná vrstva sa skladala z dvoch častí: štyroch neurónov, ktoré lokalistickým kódovaním reprezentovali stav sveta a štyroch neurónov s distribuovaným kódovaním akcie demonštrátora. Na výstupnej vrstve štyri neuróny opäť lokalistickým kódovaním reprezentovali motorickú akciu agenta. Šlo o klasickú doprednú sieť, s tým rozdielom, že adaptácia váh sa riadila hebbovským pravidlom. Váhy mali hodnoty v rozmedzí 0 až 1. Presná

podoba pravidla bola pre každé spojenie iná, zakódovaná v genóme agenta. Okrem klasického hebbovského pravidla mohlo ísť aj o jeho modifikácie: pre-synaptické a postsynaptické hebbovské pravidlo, podľa ktorého na posilnenie stačí aktivácia na jednej strane spojenia, prípadne kovariančné pravidlo, ktoré vyvolá posilnenie, ak sú aktivácie na oboch koncoch zhodné. Niektoré váhy sa nemuseli učiť vôbec, po celý život agenta boli rovnaké.

Každému spojeniu zodpovedali štyri gény, ktoré kódovali rôznu jej vlastnosť: počiatočnú váhu, znamienko váhy, rýchlosť učenia a verziu učebného pravidla. Do nasledujúcej generácie sa vždy skopírovala pätina najúspešnejších agentov. Zvyšok sa doplnil náhodne, proporčne voči fitness, pričom u malého počtu génov každého agenta došlo k mutácii. Keďže úvodná populácia mala náhodný genóm, nevedela ešte správne previesť na akciu vstup od demonštrátora. Vďaka tomu, že toto mapovanie bolo pevné, mohlo sa vyvinúť evolučne a zvýšiť úspešnosť agenta aspoň v tých prípadoch, keď vidí demonštrátora. Keď už bolo toto mapovanie rozšírené v populácii, mohla sa vyvinúť schopnosť pomocou korelácií previesť zobrazenie medzi demonštrátorom požadovanou akciou na správne mapovanie medzi stavmi sveta a akciou, ktoré sa nemohlo vyvinúť priamo evolučne. Tým sa v agentovi vyvinula schopnosť učiť sa pomocou imitácie.

Približne polovica výsledných simulácií bola neúspešná, agenty nevedeli imitovať a po celú dobu nevykonávali žiadnu akciu. V ďalších simuláciách, kde sa učenie imitáciou vyvinulo (po približne 2000 generáciách), sa ukázalo, že u niektorých takto vzniknutých agentov sa na skrytej vrstve objavili neuróny, ktoré silno pálili pri vykonávaní určitej akcie, teda vtedy, keď bol na vstupe stav, pre ktorý bola daná akcia správnu reakciou (bez ohľadu na to, či bolo vidieť aj demonštrátora), mierne, ak sa na vstupe vyskytla iba demonštrácia tejto akcie, ale nemala sa vykonať (a teda sa ani nevykonala), a slabo, ak vstup (a teda ani výstup) nemal s danou akciou nič spoločné. Autori prirovnali toto správanie k zrkadliacim neurónom.

Pozitívom modelu je, že používa biologicky plauzibilnú metódu učenia. Nedostatkom je jednak, že schopnosť imitácie a učenia pomocou imitácie sa nevyvinula vždy, jednak, že keď už sa vyvinula, nie vždy vznikli neuróny so zrkadliacim správaním (no napriek tomu agenty vedeli imitovať). Druhým nedostatkom je, že si agent musí vyvinúť mapovanie medzi senzorickej a motorickej podobou akcie evolučne, pričom plauzibilnejšie by bolo vyvinúť si ho pozorovaním senzorických dôsledkov vlastných akcií. Okrem toho je otázne, či vzniknuté správanie neurónov skutočne zodpovedá úlohe zrkadliacich neurónov, keďže tie v oblasti F5 sa v skutočnosti nepodieľajú priamo na realizácii akcie, skôr sa predpokladá, že realizujú dopredné modely, teda predpovedajú dôsledky akcií. Na realizácii akcie sa skôr podieľajú kanonické neuróny, tie však zase nepália pri pozorovaní iného agenta. Modelu možno

vytknúť aj jeho nesituovanosť. Vzťah medzi stavom sveta a akciou je veľmi abstraktný, akcia okrem zmeny fitness nemá v živote agenta žiadne iné dôsledky. Stav sveta v tomto modeli reprezentuje ako kontext limitujúci, ktoré akcie je možné vykonať, tak cieľ, ktorý určuje ktoré akcie je žiadúce vykonať, a spája tieto dve veličiny do jednej.

### 2.3.6 Sebapozorovanie (Bonaiuto a Arbib, 2010)

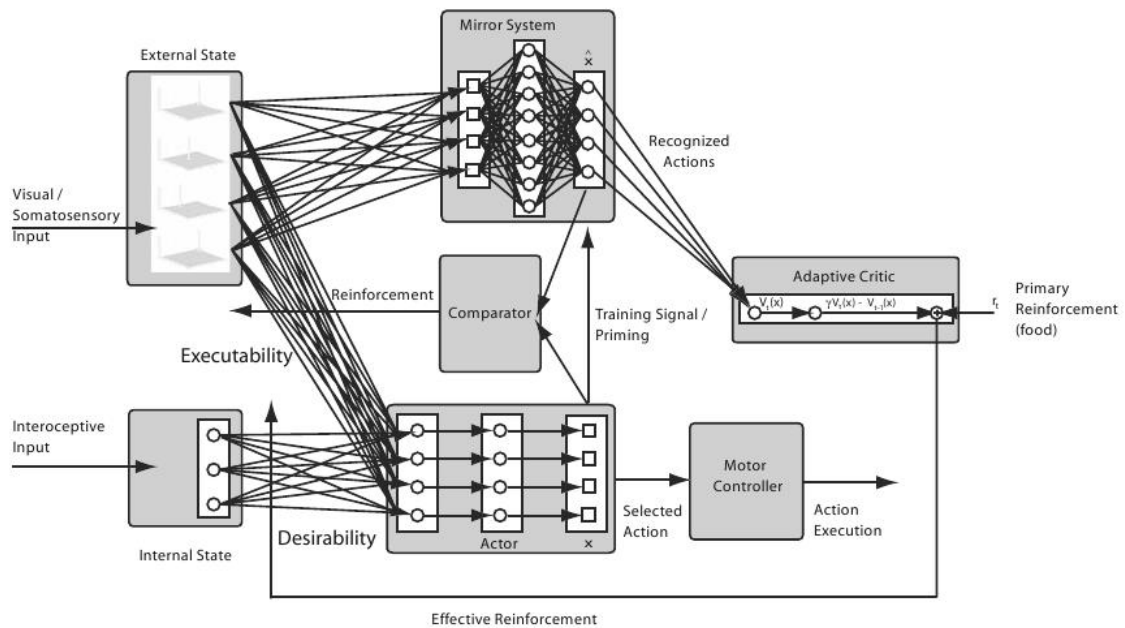
Bonaiuto a Arbib (2010) v rozšírenom modeli zrkadliaceho systému skúmali schopnosť architektúry, prednaučenej na vykonávanie jednoduchého programu pre krmenie sa mačky, efektívne reorganizovať po lézii. Vstup bol zložený z vizuálneho opisu situácie a vnútorného stavu agenta (hladu). Z vizuálneho vstupu sa vyhodnotila vykonateľnosť (executability) a z vnútorného vstupu žiadanosť (desirability) akcie. Tieto hodnoty sa vynásobili a zvolila sa akcia s najvyšším ohodnotením. Po jej vykonaní sa obnovil vizuálny vstup a zrkadliaci modul vyhodnotil, akú akciu pozoroval. Na základe porovnania mienenej a reálne pozorovanej akcie sa upravila vykonateľnosť akcie. Pomocou adaptívneho kritika a prípadného posilnenia z vonka sa upravila žiadanosť akcie. Takto pomocou sebapozorovania a učenia s posilňovaním nastala reorganizácia programu pre nové podmienky po lézii.

Tento model bol inšpirovaný inými modelmi zrkadliacich neurónov a skúma potenciálnu alternatívnu úlohu zrkadliacich neurónov v kombinácii so sebapozorovaním. Sebakontrolná funkcia umožňuje v agentovi preorganizovať si program pre vyššiu efektivitu v prípade zmeny nejakých významných podmienok, napríklad pri lézii. V modeli nejde o komunikáciu, ani o evolúciu, namiesto toho sa agent učí paradigmatom učenia s posilňovaním (reinforcement learning) (Sutton a Barto, 1998).

Model je situovaný v jednoduchom prostredí. Nedostatkom modelu je, že predpokladá predučenosť niektorých komponentov, a má jednorozmerný interný stav (hlad).

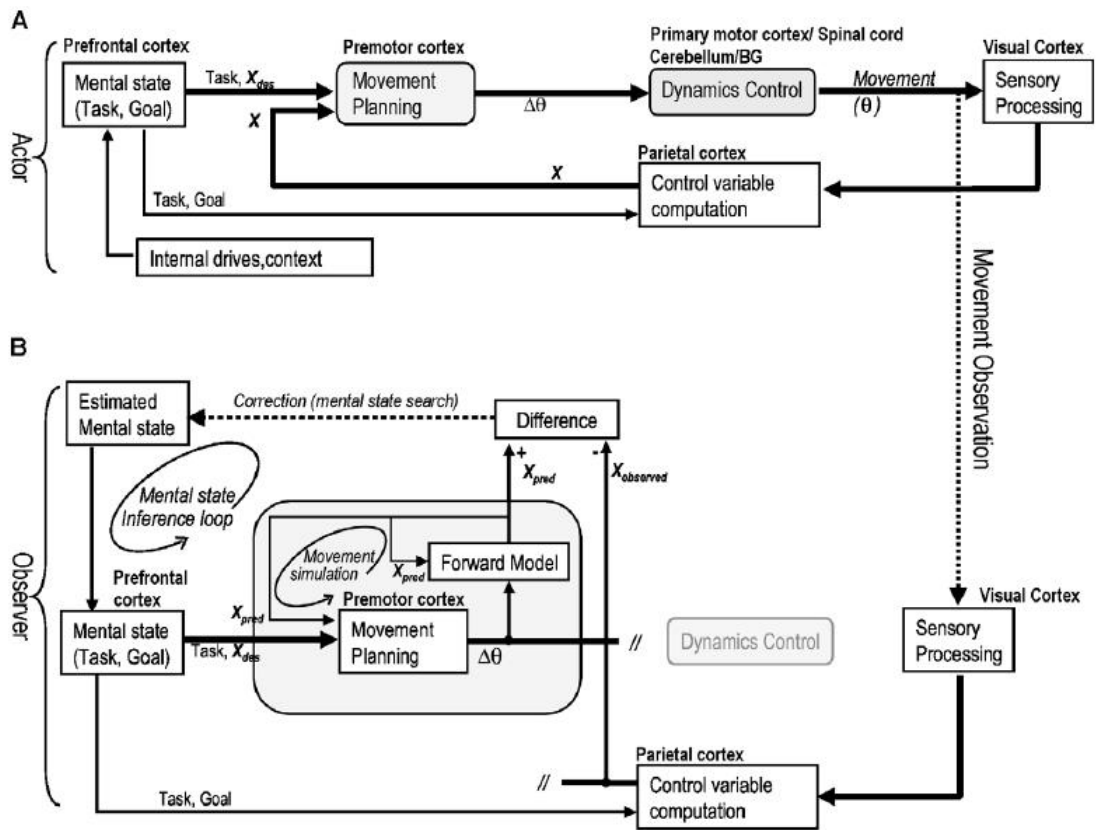
### 2.3.7 Mentálny stav (Oztop a kol., 2005)

Model inferencie mentálneho stavu (mental state inference) je odvodený od modelov slúžiacich na spojitú motorickú kontrolu, kde spätná väzba umožňovala predpovedať a priebežne korigovať pohyb efektora za nejakým účelom (Oztop a kol., 2005). Súčasťou modelu je aj okruh zodpovedný za predpovedanie dôsledkov akcií, realizuje teda dopredný model. Táto časť okruhu sa dáva do súvislosti so zrkadliacimi neurónmi. Ak prepojíme niektoré spojenia istým spôsobom (na obrázku), je agent s takýmto zapojením schopný



Obr. 2.2: Bonaiuto a Arbibov (2010) model neurálnej architektúry agenta.

imitovať pozorovanú akciu. Pri tomto zapojení sa agent snaží vytvoriť si reprezentáciu mentálneho stavu, cieľa druhého agenta. Z toho potom vyplýva úspešná imitácia. Schopnosť prisúdiť inému agentovi správny mentálny stav je významnou zložkou teórie mysle.



Obr. 2.3: Model inferencie mentálneho stavu (Oztop a kol., 2005).

# Kapitola 3

## Projekt

V našom výskume by sme sa chceli zaoberať vzťahom jazyka a teórie mysle ako dvoch fenoménov, ktoré sa evolučne zrejme do značnej miery opierajú o blízke, podobné preadaptácie, v procese svojho vývoja sa navzájom ovplyvňovali, a v procese výmeny informácie sa prejavujú rôznym spôsobom a v iných situáciach.

### 3.1 Základné tézy

Jazyk vnímame v prvom rade ako nástroj na úmyselné zdieľanie obsahu mysli. Ide o nástroj spolupráce, z ktorého majú prospech obe strany interakcie. Miera štruktúrovanosti jazyka bude odrážať mieru štruktúrovanosti komunikovaných významov a ich oddeliteľnosť od iných významov, ktoré komunikovať nechceme. Ak použitie jazyka viedlo k zvýšeniu kolektívnej fitness, dá sa považovať za úspešné.

Dobrý jazyk minimalizuje nejednoznačnosť, je úsporný ale i dostatočne redundantný, aby bol odolný voči šumu, umožňuje čo najefektívnejší a najpresnejší prenos informácie z mysle jedného agenta do druhého. Používame ho ako nástroj získavania a odovzdávania poznatkov, koordinácie úmyslov pri spoločnom vykonávaní komplikovaných akcií, atď. V princípe používame jazyk vždy vtedy, keď chceme úmyselne externalizovať nejaký aspekt nášho vnútorného stavu, aby sa dal internalizovať ľubovoľným iným agentom a mohli sme svoje akcie koordinovať v spoločnom záujme.

Teória mysle, teda schopnosť prisudzovať rozličným agentom presvedčenia, túžby, záujmy, úmysly a iné stavy mysle odlišné, než máme my sami, je tiež nástrojom na nahliadnutie do obsahu mysle iného agenta. Je to nástroj, ktorým robíme úsudky o mentálnych stavoch druhých na základe ich vonkajších prejavov, akcií v danom kontexte, a to aj keď sa ich nesnažia dá-

vať úmyselne najavo, ba naopak sa ich možno snažia skryť. Umožňuje nám oddeliť chápanie záujmov z nášho a cudzieho pohľadu, pochopiť, že aj ostatní majú teóriu mysle, a prispôbiť svoje prejavy, aby ich ostatní chápali a vysvetľovali si ich tak, ako my chceme.

Jedinec s dobrou teóriou mysle vie pochopiť aj motivácie iných v takých situáciách, v akých sa sám nikdy nenachádzal, a na základe toho sa orientovať v zložitých sociálnych vzťahoch v skupine, využívať a upravovať ich vo svoj prospech. Teória mysle ako prostriedok zdieľania vnútorných stavov teda zrejme funguje oboma smermi, podobne ako jazyk, ale na podvedomej, aspoň do istej miery neúmyselnej úrovni. Vtedy ide o nástroj zabezpečenia správnej spoločnej akcie v situácii, kedy to bude obojstranne výhodné. Na úmyselnej úrovni ide často o nástroj presadzovania vlastných záujmov na úkor okolia, o nástroj, ktorý nám otvára cestu k úmyselnému klamaniu a predstieraniu.

Jazyk a teóriu mysle vnímame ako dva navzájom sa dopĺňajúce nástroje prenosu informácie medzi agentmi, ktoré nám umožňujú orientovať sa v dvoch ortogonálnych aspektoch komunikácie: v sémantickom obsahu a v pragmatickom celi komunikácie. V prvých fázach evolúcie komunikácie pragmatická stránka komunikácie umožnila agentom rozoznať, že ide o úmysel komunikovať, zdieľať pozornosť a umožnila najmä rozvoj obsahovej stránky komunikácie pre zvládnutie rozličných kooperatívnych úloh. Neskôr, keď sa zvýraznil vplyv súťaže kultúr namiesto súťaže druhov, vytvoril sa tlak na zlepšovanie schopnosti orientácie pragmatike. Pragmatika zahŕňala rôzne ritualizované štandardy pozmenenia komunikácie oproti širšej štandardnej tak, aby bolo stále možné rozoznať obsahovú stránku, ale aby ju zároveň bolo možné interpretovať iným spôsobom, ako prejav príslušnosti k nejakej podskupine vrámci druhu, prejav dôvery, vyjadrenie postoja k vyslovenému obsahu, prípadne ako úplne nezávislý obsahový odkaz určený úzkej skupine agentov a podobne. Obom týmto javom sa prisudzujú aspoň z časti spoločná neurologická podstata – zrkadliace neuróny.

Pokiaľ ide o arbitrárnosť symbolov, či úmyselné sledovanie arbitrárnych dlhodobých cieľov, našou hypotézou je, že tieto javy úzko súvisia s teóriou mysle. Realizovať arbitrárne symboly pomocou neurónovej siete je možné a priori. Otázka je, čo v procese evolúcie spôsobilo, že sme od navodených signálov, indexickej a ikonickkej komunikácie prešli k arbitrárnym symbolom, ktoré nemajú k referentovi žiadny vzťah mimo toho, že sú prepojené existujúcim komunikačným štandardom.

Našou hypotézou je, že schopnosť predstierať, že nejaký symbol na chvíľu (napr. pre niekoho iného) znamená niečo iné než “v skutočnosti” (teda pre nás), mohla viesť k tomu, že takéto dočasné predstieranie sa stalo súčasťou štandardu a symboly sa tým postupne stali ľubovoľné. Schopnosť oddeliť vlastný cieľ od cieľa iného agenta nám analogicky umožňuje predstaviť si na



chvíľu, že náš cieľ je iný, než skutočný. Ak tento predstieraný cieľ sme schopní udržiavať dlhodobejšie a riadiť sa ním, akoby bol skutočný, dosiahli sme oddelené reprezentácie arbitrárnych cieľov. Arbitrárnosť symbolov a cieľov teda môžu byť produktom teórie mysle.

Pokiaľ ide o prechod od biologickej ku kultúrnej evolúcii, *našou centrálnou tézou* je, že rudimentárne spôsoby komunikácie, prvopočiatky teórie mysle a kultúrny prenos vzorcov správania, ktoré boli naraz umožnené alebo zefektívnené vznikom zrkadliacich neurónov, sa formovali v kompetitívno-kooperatívnom prostredí, kde na seba navzájom pôsobili, dopĺňali sa, urýchlňovali svoj vývoj a spätne zvyšovali komplexnosť prostredia. Vznikla tak pozitívna spätnoväzobná evolučná slučka, ktorá postupom času viedla k zanedbaniu vplyvu biologickej evolúcie, vzniku kolektívnych štandardov pretvárania, raketovému nástupu štruktúrovaného jazyka, ľudskej kultúry a civilizácie. Tento vývoj pokračuje na vyšších úrovniach dodnes, a na nižších, už relatívne ustanovených úrovniach sa prejavuje ako neustála zmena.

## 3.2 Vlastnosti a ciele modelu

Chceme vytvoriť výpočtový model pre testovanie tejto hypotézy založený na predchádzajúcich modeloch. V skratke si priblížime niekoľko z nich, ktoré by mohli byť inšpiráciou. Každý naznačuje cestu k nejakej časti takého modelu. Formulujeme niekoľko podmienok kladených na agenta, aby mu bolo umožnené geneticky si vyvinúť jazyk a teóriu mysle a spustiť tak kultúrnu evolúciu. Agent má možnosť:

- nejakým spôsobom s inými agentmi kooperovať (správať sa altruisticky, v záujme spoločného blaha potenciálne na úkor vlastného) ale aj odmietnuť spoluprácu (správať sa egoisticky, v záujme maximalizácie vlastného blaha). Obe stratégie by mali za istých okolností viesť k zvýšeniu fitness, ideálne by sa mal agent naučiť vedieť kedy spolupracovať a kedy nie.
- komunikovať, t.j. produkovať arbitrárne verejne vnímateľné akcie, vzory v prostredí, ktoré budú môcť iné agenty vnímať a vyhodnocovať, ale na začiatku by nemal byť ustanovený žiadny štandard, ani by nemalo byť umožnené, aby sa vyvinul geneticky.
- ovplyvňovať pravdepodobnosť interakcie s rôznymi agentmi, tak aby si mohol vytvoriť svoj sociálny okruh agentov, s interakciou ktorými si dokáže zvyšovať fitness lepšie než s inými. To by malo mať svoje

hranice, tak aby existovala aj možnosť akéhosi násilia, nedobrovoľnej nepriaznivej interakcie.

- pozorovať sám seba, teda vnímať svoje vlastné akcie podobne, ako ich vnímajú ostatné agenty v jeho okolí, čo by malo byť základom pre neskoršiu imitáciu.
- interagovať medzi-generačne, prechod na novú generáciu by teda mal byť postupný, mal by existovať akýsi súbeh generácií, minimálne dvoch, možno aj troch. Techniky kultúrneho prenosu by sa mali samy vyvinúť a podliehať evolučným tlakom, pokiaľ možno, nemalo by ísť o predprogramované uprednostňovanie rodičov ako imitačných vzorov, ani nič podobné.

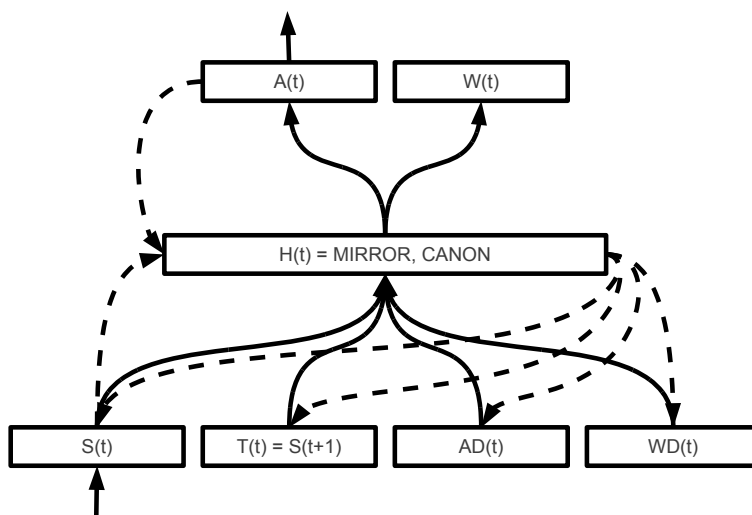
Po splnení týchto podmienok by sme mali mať evolučný model prostredia a agenta, kde môže dôjsť k evolúcii komunikácie, teórie mysle, sociálnych štruktúr a kultúrneho prenosu. Potom môžeme zistiť, za akých podmienok sa jazyk a teória mysle vyvíjajú nezávisle na sebe či spoločne, ako sa ovplyvňujú a či samy o sebe alebo spoločne podporujú postupný vznik kultúrnej evolúcie.

### 3.3 Návrh modelu

Základnou inšpiráciou pre nás bude model evolúcie učenia pomocou imitácie Borensteina a Ruppina (2004; 2005). V tejto podkapitole popíšeme v hrubých črtách ako by sme túto simuláciu chceli rozšíriť, tak aby spĺňala podmienky stanovené v predchádzajúcom odstavci. Zameriame sa na architektúru riadiacej neurónovej siete agenta, jednak na nejaké vlastnosti prostredia, v ktorom bude agent situovaný, najmä charakter interakcií.

Návrh neurálnej architektúry agenta je dopredná neurónová sieť s jednou skrytou vrstvou a hebbovským učením, podobne ako v pôvodnom modeli. Veličinu stav sveta nahradia dve veličiny stav sveta a cieľový stav a ich kombinácia bude určovať, akú akciu treba vykonať. Vstup od demonštrátora nahradí vizuálny a jazykový vnem. Vizuálny bude predstavovať senzorickú podobu akcie demonštrátora, ktorá má viesť z aktuálneho stavu do cieľového. Jazykový vnem je senzorickou podobou jazykového výstupu demonštrátora, ktorý sprevádza akciu. Môže byť ľubovoľný, nemá priamy dopad na zmenu stavu sveta.

Výstupná vrstva sa skladá z motorického a jazykového výstupu. Motorický kóduje akciu agenta, na základe ktorej sa vypočíta nasledovný stav simulácie. Jazykový výstup kóduje verbálnu reakciu agenta. Tá môže byť



Obr. 3.1: Navrhovaný model neurálnej architektúry agenta.

ľubovoľná a nemá priamy dopad na nasledovný stav simulácie. Ak však interakcia agentov pokračuje, prevedie sa na jazykový sensorický vstup druhého agenta a môže ovplyvniť jeho reakciu.

Okrem dopredných váh budú v modeli existovať spätné váhy od výstupnej vrstvy rozšírenej o aktuálny stav, cez skryté neuróny po vstupnú vrstvu. Tieto budú implementovať dopredný model, ktorý predpovedá dôsledky akcií agenta. Bezprostredne po vykonaní akcie dôjde k sebazozorovaniu. Zapoja sa spätné váhy, vytvorí sa predpoveď, a porovná sa so skutočnými dôsledkami jeho akcie. Nový stav sveta, nový cieľ, sensorická podoba vlastnej akcie a sensorická podoba vlastného jazykového výstupu sa použijú ako učebný vstup pre dopredný model. Nový stav sveta sa navyše porovná s aktuálnym cieľom a podľa toho sa upraví fitness agenta.

Dopredný beh siete sa bude učiť hebbovským učením s rôznymi učebnými pravidlami podobne ako u Borensteina a Ruppina. Spätný beh, čiže dopredný model, sa bude učiť hebbovským učiacim algoritmom upraveným na algoritmus učenia s učiteľom. Do úvahy pripadá pravdepodobne najmä algoritmus CHL (Contrastive hebbian learning), prípadne jeho obdoba GeneRec (Generalized Recirculation), ktorá vedie k symetrickým váham (O'Reilly a Munakata, 2000). V princípe ide v týchto algoritmoch o to, že spojenia v hebbovskej sieti sú obojsmerné, po doprednom behu pokračuje spätný beh. Na výstup sa privedú učebné hodnoty, rozšíria sa späť v opačnom smere cez

Tabuľka 3.1: Koordinačná hra s viacerými akciami

	Akcia 1	Akcia 2	Akcia 3	Akcia 4	Akcia 5
Akcia 1	1; 1	0; 0	0; 0	0; 0	0; 0
Akcia 2	0; 0	1; 1	0; 0	0; 0	0; 0
Akcia 3	0; 0	0; 0	1; 1	0; 0	0; 0
Akcia 4	0; 0	0; 0	0; 0	1; 1	0; 0
Akcia 5	0; 0	0; 0	0; 0	0; 0	1; 1

tie isté neuróny až na vstup, a upraví sa váha. Zmena váhy je úmerná rozdielu medzi doprednou a spätnou aktiváciou na danom neuróne a aktivácií vstupu. Alternatívou je, že by dopredný aj spätný beh siete používali tie isté váhy.

Na rozdiel od pôvodného modelu bude v úlohe demonštrátora vystupovať skutočný agent. S vysokou pravdepodobnosťou by to mal byť rodičovský agent, aby mohlo dochádzať k medzigeneračnému kultúrnemu prenosu informácie. Taktiež by mal agent častejšie vstupovať do interakcie s agentmi s blízkym genetickým aj kultúrnym pôvodom. Tieto pravdepodobnosti môžu byť reprezentované pomocou akejsi sociálnej vzdialenosti medzi agentmi, pričom pravdepodobnosť interakcie s agentom by rástla s poklesom vzdialenosti.

Ciele v interakciách, do ktorých budú agenty vstupovať, budú zo začiatku nastavené tak, aby šlo o koordinačné hry. Jeden, druhý, alebo oba agenty budú odmeňované za vykonanie zhodných akcií. Podobne ako v pôvodnom modeli budú s istou pravdepodobnosťou niektoré časti vstupu zakryté a bude si ich treba domyslieť. Podľa toho potom pôjde o rôzne druhy úlohy. Ak je zakrytá akcia demonštrátora, ale viditeľný stav a cieľ, treba vykonať akciu, ktorá vedie k cieľu. Ak nevidíme cieľ, ale vidíme akciu, demonštrátora, ide o úlohu imitovať. Ak z akcie demonštrátora vidíme iba jeho jazykový prejav, ale zakrytý je cieľ akcie, môžeme to chápať ako príkaz, resp. radu. Ak by sa hodnotila aj zhodnosť jazykových prejavov, šlo by o jazykové hry (hádanie a pomenúvanie).

Postupom času, keď už by sa dalo očakávať ustanovenie učenia pomocou imitácie a možno nejakých základných signálov pre úspešnejšiu spoluprácu, by sa hra začala komplikovať. Namiesto koordinačnej by začali častejšie hrať antikoordinačnú hru, kde by bolo dosiahnutie cieľa možné iba tak, že jeden agent vykoná jednu akciu a druhý nejakú konkrétnu inú druhú. Stále majú spoločný záujem v danej situácii, ale úloha teraz zahŕňa potrebu demonštrátora, resp. hovorcu naviesť prijímateľa na vykonanie správnej akcie, prípadne potrebu prijímateľa predpovedať, ktorú z dvoch potrebných akcií

Tabuľka 3.2: Antikoordinačná hra s viacerými akciami

	Akcia 1	Akcia 2	Akcia 3	Akcia 4	Akcia 5
Akcia 1	0; 0	1; 1	0; 0	0; 0	0; 0
Akcia 2	1; 1	0; 0	0; 0	0; 0	0; 0
Akcia 3	0; 0	0; 0	0; 0	1; 1	1; 1
Akcia 4	0; 0	0; 0	1; 1	0; 0	0; 0
Akcia 5	0; 0	0; 0	1; 1	0; 0	0; 0

vykoná hovorca a vedieť ho správne doplniť. Akcia sa dá porovnať s testom falošných presvedčení, pričom motivácia je toto falošné presvedčenie druhej strany nejako korigovať v spoločnom záujme.

V konečnej fáze by sa repertoár hier rozšíril o hry, kde by sa ciele interagujúcich agentov nejakým spôsobom nezhodovali. Agent by sa musel naučiť buď naviesť druhú stranu na akciu, ktorá je pre ňu nevýhodná, alebo rozpoznať predstieranie u druhej strany a nenechať sa naviesť na takú akciu. Niekedy by rozhodnutie o akcii mohlo závisieť akoby na miere dôvery voči druhej strane, o ktorej sa usúdi na základe nejakého aspektu jej prejavu. De facto by šlo o verziu testu falošných presvedčení s motiváciou klamať. V tom prípade by sa mohlo ukázať výhodné používať nejaké aspekty komunikácie na rozlišovanie medzi priateľmi a nepriateľmi.

V druhej aj tretej fáze by sa štandard vyvinutý v prvej fáze musel upraviť pre nové použitie, keďže by sa však vyskytovali stále aj interakcie koordinačného charakteru, musel by si zachovať aj svoju staršiu funkciu.

Od modelu si sľubujeme, že bude škálovateľný, a do istej miery odolný voči šumu. Je dôležité, aby si zachoval tú vlastnosť, že niektoré veci sa agenty nebudú môcť naučiť inak, než v procese interakcie.

### 3.4 Diskusia

Zdieľaná pozornosť je v našom modeli prítomná implicitne. Oba agenty interakcie automaticky dostávajú na vstupe ten istý stav sveta a senzorickú podobu akcie toho správneho agenta. Samotná zdieľaná pozornosť sa v agentoch teda nemusí vyvinúť. Všeobecnejší model by musel zahŕňať neurálnu kontrolu agenta nad tým, čomu venuje pozornosť, čo je celkom nová úloha.

Knight na vznik rituálov a pretvárinky, ktorá viedla podľa neho k ďalším javom v jazyku, predpokladá situáciu, kedy sa viacerí príbuzní slabší jedinci nejakým spôsobom spojili proti cudzincom. Aby sme niečo také si-

mulovali priamo, museli by interakcie mať možnosť zahŕňať viac agentov, čo by skomplikovalo mnohé aspekty modelu (neurónovú sieť, priebeh interakcie, štruktúry ohodnotenia v hrách). Ak však hry budú mať vyššie popísané vlastnosti, je možné, že budú tlačené k tomu vytvárať nejakým spôsobom koalície tak, či tak.

Rola sexuálneho rozmnožovania na priebeh evolúcie by mohla byť zaujímavou otázkou. V závislosti od toho, ako by sa určovalo, ktoré dva agenty sa stanú základom pre nového agenta (čisto podľa fitness, alebo aj podľa situácie v prostredí, alebo by bolo párenie sa priamo jednou z možných akcií), a či by boli agenty rôzneho pohlavia, alebo by sa mohli páriť ľubovoľne, by sa tým model skomplikoval viac, či menej. V kontexte téz a cieľov modelu sa však úloha sexuálneho rozmnožovania nezdá dôležitá.

Spôsob reprezentácie prostredia a miera situovanosti simulácie je zatiaľ otázná. Ideálne by bolo mať reprezentáciu prostredia a interakcií takú, že presun motivácií od koordinačných cez antikoordinačné až ku hrám s možnosťou zavádzať, prebehne samovoľne, čisto ako dôsledok evolúcie. Otázne je, ako dať agentom najavo, že hra má novú štruktúru, resp. z akých náznakov by to mohli spoľahlivo rozpoznať sami.

Emergencia neurónov so zrkadliacim správaním na skrytej vrstve by bola žiadúcou vlastnosťou, ktorú by mohol náš model od Borensteinovho a Ruppinoého modelu zdediť. Zaujímavé by bolo, ak by v modeli emergovali zrkadliace neuróny spoľahlivejšie než u pôvodného a ak by navyše emergovali rozoznateľné oddelené skupinky neurónov so zrkadliacim a kanonickým správaním. Mohlo by to prispieť k riešeniu debaty medzi zástancami vrodenej a získanej zrkadliacich neurónov.

# Literatúra

- Arbib, M. From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(02):105–124, 2005. ISSN 0140-525X.
- Axelrod, R. a Hamilton, W. The evolution of cooperation. *Science*, 211(4489):1390, 1981.
- Baron-Cohen, S., Leslie, A. a U., F. Does the autistic child have a 'theory of mind'? *Cognition*, 21(1):37–46, 1985.
- Barsalou, L. Perceptual symbol systems. *Behavioral and brain sciences*, 22(04):577–660, 1999. ISSN 0140-525X.
- Bašnáková, J. a Rybár, J. Teória mysle a sociálna kognícia. V *Sociálna inteligencia*, 151–169. Europa Publishing House, 2010.
- Belpaeme, T. a Bleys, J. Explaining universal color categories through a constrained acquisition process. *Adaptive Behavior*, 13(4):293, 2005. ISSN 1059-7123.
- Bickerton, D. *Language and Human Behavior*. University of Washington Press, Seattle, 1995.
- Blackmore, S. *The meme machine*. Oxford University Press, USA, 2000. ISBN 019286212X.
- Bonaiuto, J. a Arbib, M. Extending the mirror neuron system model, II: what did I just do? A new role for mirror neurons. *Biological cybernetics*, 102(4):341–359, 2010. ISSN 0340-1200.
- Borenstein, E. a Ruppín, E. Enhancing autonomous agents evolution with learning by imitation. *Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour*, 1(4):335–348, 2003.

- Borenstein, E. a Ruppin, E. Evolving Imitating Agents and the Emergence of a Neural Mirror System. V *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*. 2004.
- Borenstein, E. a Ruppin, E. The evolution of imitation and mirror neurons in adaptive agents. *Cognitive Systems Research*, 6(3):229–242, 2005. ISSN 1389-0417.
- Byrne, R. Imitation as behaviour parsing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431):529, 2003. ISSN 0962-8436.
- Cangelosi, A. a Parisi, D. The emergence of a 'language' in an evolving population of neural networks. *Connection Science*, 10(2):83–97, 1998. ISSN 0954-0091.
- Chaminade, T., Oztop, E., Cheng, G. a Kawato, M. From self-observation to imitation: Visuomotor association on a robotic hand. *Brain research bulletin*, 75(6):775–784, 2008. ISSN 0361-9230.
- Christiansen, M. a Kirby, S. Language evolution: Consensus and controversies. *Trends in Cognitive Sciences*, 7(7):300–307, 2003. ISSN 1364-6613.
- Corballis, M. C. The evolution of language. *Annals of the New York Academy of Sciences*, 1156:19–43, 2009.
- Dawkins, R. *The Selfish Gene*. Oxford University Press, New York City, 1976. ISBN 0-19-286092-5.
- Deacon, T. *The Symbolic Species*. Penguin, London, 1997.
- Demichelis, S. a Weibull, J. Language, Meaning, and Games: A Model of Communication, Coordination, and Evolution. *The American Economic Review*, 98(4):1292–1311, 2008. ISSN 0002-8282.
- Farkaš, I., Malý, M. a Rebrová, K. Mirror neurons – theoretical and computational issues. *Technical Report TR-2011-28*, 2011.
- Frank, S. a Bod, R. Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, in press.
- Gallese, V., Keysers, C. a Rizzolatti, G. A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8(9):396–403, 2004.



- Gärdenfors, P. Cued and detached representations in animal cognition. *Behavioural Processes*, 35(1-3):263–273, 1995. ISSN 0376-6357.
- Gärdenfors, P. Language and the Evolution of Cognition. V Rialle, V. a Fisette, D., redaktori, *Penser l'esprit: Des sciences de la cognition à une philosophie cognitive*, 151–172. Presses Universitaires de Grenoble, Grenoble, 1996.
- Gärdenfors, P. Cooperation and the evolution of symbolic communication. V Oller, K. a Griebel, U., redaktori, *The Evolution of Communication Systems: A Comparative Approach*, 237–256. MIT Press, 2004.
- Heyes, C. Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, 34(4):575–83, 2010. ISSN 1873-7528. URL <http://www.ncbi.nlm.nih.gov/pubmed/19914284>.
- Hickok, G. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 21(7):1229–43, 2008.
- Hickok, G. a Hauser, M. (Mis)understanding mirror neurons. *Current Biology*, 20(14):R593–4, 2010. ISSN 1879-0445.
- Kirby, S. Learning, bottlenecks and the evolution of recursive syntax. *Linguistic evolution through language acquisition*, 173–203, 2002.
- Kirby, S. The evolution of language. V Dunbar, R. a Barrett, L., redaktori, *The Evolution of Communication Systems: A Comparative Approach*, 669–681. MIT Press, 2007.
- Knight, C. Ritual/speech coevolution: a solution to the problem of deception. V Hurford, J. R., Studdert-Kennedy, M. a Knight, C., redaktori, *Approaches to the evolution of language: social and cognitive bases*, 68–91. Cambridge University Press, Cambridge, 1998.
- Knoblich, G. a Sebanz, N. Brief report: cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *Journal of Autism and Developmental Disorders*, 34(2):229–235, 2004.
- Knoblich, G. a Sebanz, N. Evolving intentions for social interaction: from entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1499):2021, 2008. ISSN 0962-8436.
- Kvasnička, V. a Pospíchal, J. An emergence of coordinated communication in populations of agents. *Artificial Life*, 5(4):319–342, 1999. ISSN 1064-5462.

- Kwisthout, J., Vogt, P., Haselager, P. a Dijkstra, T. Joint attention and language evolution. *Connection Science*, 20(2):155–171, 2008. ISSN 0954-0091.
- Logan, R. The Extended Mind Model of the Origin of Language and Culture. V Gontier, N., Van Bendegem, J. a Aerts, D., redaktori, *Evolutionary Epistemology, Language and Culture*, diel 5. Springer, 2005.
- Markošová, M. Jazyk ako sieť malého sveta. V Rybár, J., Kvasnička, V. a Farkaš, I., redaktori, *Jazyk a kognícia*. 2005.
- Maynard-Smith, J. a Szathmáry, E. *The Major Transitions in Evolution*. Oxford University Press, New York City, 1997.
- Metta, G., Sandini, G., Vernon, D., Natale, L. a Nori, F. The iCub humanoid robot: an open platform for research in embodied cognition. V *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, 50–56. ACM, 2008.
- Miall, M. Connecting mirror neurons and forward models. *NeuroReport*, 14(17):2135–2137, 2003.
- Oliphant, M. *Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language*. Doktorská práca, University of California, San Diego, 1997.
- O'Reilly, R. a Munakata, Y. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. The MIT Press, Cambridge, MA, 2000.
- Oztop, E., Kawato, M. a Arbib, M. Mirror neurons and imitation: A computationally guided review. *Neural Networks*, 19(3):254–271, 2006. ISSN 0893-6080.
- Oztop, E., Wolpert, D. a Kawato, M. Mental state inference using visual control parameters. *Cognitive Brain Research*, 22(2):129–151, 2005. ISSN 0926-6410.
- Pelikan, M. Evolučné algoritmy. V Kvasnička, V., Pospíchal, J., Kozák, Š., Návrat, P. a Paroulek, P., redaktori, *Umelá inteligencia a kognitívna veda I*, 335–353. 2009.
- Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. a Rizzolatti, G. Understanding motor events: a neurophysiological study. *Experimental brain research*, 91(1):176–180, 1992.

- Perfors, A. Simulated Evolution of Language: A Review of the Field. *Journal of Artificial Societies and Social Simulation*, 5(2), 2002.
- Pinker, S. a Bloom, P. Natural language and natural selection. *Behavioral and Brain Sciences*, 13:707–784, 1990.
- Premack, D. G. a Woodruff, G. Does the chimpanzee have a theory of mind? *Behavioural and Brain Sciences*, 35(1-3):515–526, 1978. ISSN 0376-6357.
- Prudkov, P. Motivation rather than imitation determined the appearance of language. *Behavioral and Brain Sciences*, 28(02):142–143, 2005. ISSN 0140-525X.
- Rizzolatti, G. a Arbib, M. Language within our grasp. *Trends in neurosciences*, 21(5):188–194, 1998.
- Rizzolatti, G. a Fabbri-Destro, M. The mirror system and its role in social cognition. *Current opinion in neurobiology*, 18(2):179–184, 2008. ISSN 0959-4388.
- Rizzolatti, G. a Sinigaglia, C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11(4):264–274, 2010. ISSN 1471-003X.
- Rogers, A., Dash, R., Ramchurn, S., Vytelingum, P. a Jennings, N. Coordinating team players within a noisy Iterated Prisoner’s Dilemma tournament. *Theoretical Computer Science*, 377(1-3):243–259, 2007. ISSN 0304-3975.
- Ruppin, E. Evolutionary autonomous agents: A neuroscience perspective. *Nature Reviews Neuroscience*, 3(2):132–141, 2002. ISSN 1471-003X.
- Scott-Phillips, T. a Kirby, S. Language evolution in the laboratory. *Trends in Cognitive Sciences*, 2010. ISSN 1364-6613.
- Sebanz, N. a Knoblich, G. Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1(2):353–367, 2009. ISSN 1756-8765.
- Senghas, A., Kita, S. a Ozyurek, A. Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305(5691):1779, 2004.
- Skyrms, B. *Evolution of the social contract*. Cambridge Univ Pr, 1996. ISBN 0521555833.

- Skyrms, B. *The stag hunt and the evolution of social structure*. Cambridge Univ Pr, 2004. ISBN 0521533929.
- Skyrms, B. *Signals: evolution, learning, & information*. Oxford Univ Pr, 2010. ISBN 0199582947.
- Smith, J. *Evolution and the Theory of Games*. Cambridge Univ Pr, 1982. ISBN 0521288843.
- Steels, L. The synthetic modeling of language origins. *Evolution of communication*, 1(1):1–34, 1997.
- Steels, L. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103(1-2):133–156, 1998. ISSN 0004-3702.
- Steels, L. The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection Science*, 17(3):213–230, 2005. ISSN 0954-0091.
- Sutton, R. a Barto, A. *Reinforcement learning: An introduction*. The MIT press, 1998. ISBN 0262193981.
- Tani, J., Ito, M. a Sugita, Y. Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. *Neural Networks*, 17(8-9):1273–1289, 2004.
- Turing, A. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950. ISSN 0026-4423.
- Wagner, K., Reggia, J., Uriagereka, J. a Wilkinson, G. Progress in the Simulation of Emergent Communication and Language. *Adaptive Behavior*, 11(1):37–69, 2003.
- Willingham, D. T. *Cognition: The thinking animal (3rd ed.)*. Pearson/Allyn & Bacon, Upper Saddle River, NJ, 2007.
- Wolpert, D. a Kawato, M. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7-8):1317–1329, 1998.