Filozofická fakulta Masarykovy Univerzity v Brně
Psychologický ústav
Studijní rok 2004/2005

DIPLOMOVÁ PRÁCE

ZPŮSOBY SIMULACE INTELIGENCE

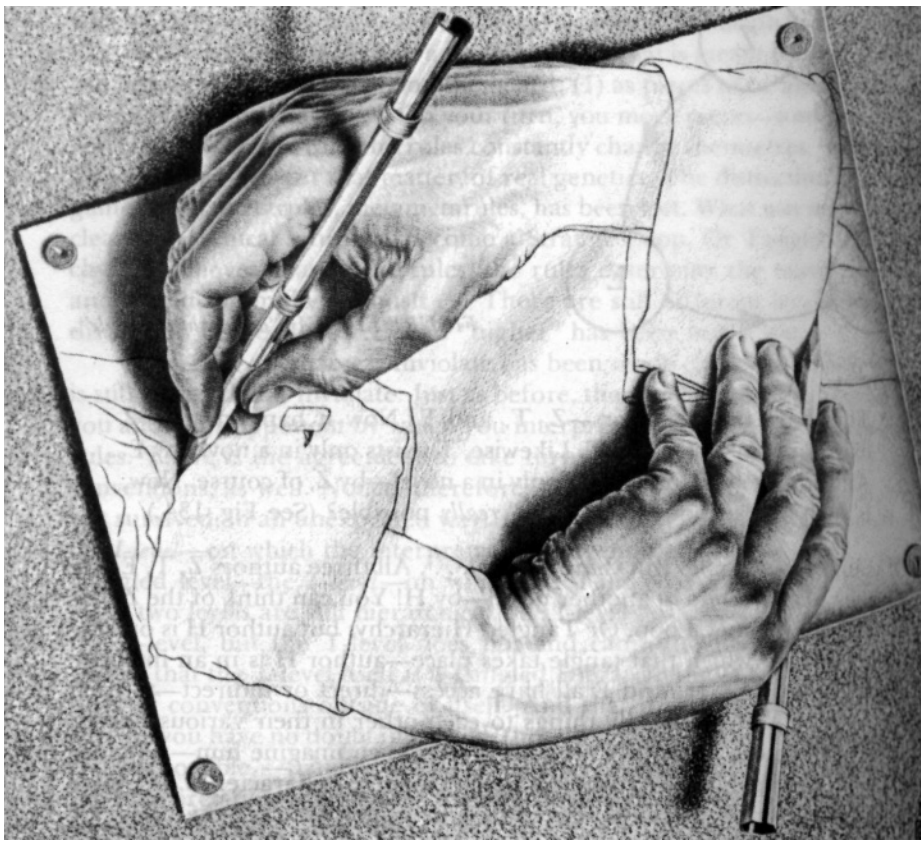
Michal Vavrečka

Vedoucí diplomové práce: Mgr. Helena Klimusová

Brno 2004

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a uvedl v ní veškerou literaturu a jiné prameny, které jsem použil

V Brně 30.11. 2001 .....Podpis



Motto:
Ví, že snadný znamená - neuskutečnitelný,
že je dvojramenná, řeka a směry v ní.
Ví, že lidský znamená – nenapodobitelný,
řeka je neměnná, jen voda plyne v ní.

Už jsme doma, Řeka

Poděkování

za korektury: Petra Balabánová, Jana Dvořáčková, Markéta Dvořáčková, Karel Gregor, Pavel Kroupa, Filip Rozsíval, Hikmet Salihová, Tereza Šarmanová, Danka Urbancová, Zuzana Vavrečková

za grafiku a textové úpravy: Libor Alexa, Jirka Mudrák
za sazbu a zlom: Ondra Kroupa
za počítač: Tomáš Jiřík
za trpělivost při čtení: Mgr.Helena Klimusová, PhDr. Tomáš Urbánek PhD.
za všechno: rodičům

OBSAH

8 Paměťové systémy a reprezentace znalostí

8.1 CAM

8.2 Memory surface

8.3 Mentální reprezentace

8.4 Rámce

8.5 Bayesianské sítě

9 Tvorba významu

9.1 Jazyk

9.2 Ukotvení symbolů

9.3 Kontext

Závěr = Sémantika

ÚVOD

Důvodů k napsání této práce je několik. Daly by zde vyjmenovat samostatně, ale to by z nich činilo odděleně stojící témata, což může být účelné pro získání přehledu a vytvoření struktury, ale také vytvářet dojem izolovanosti. Pokud bychom však hledali slovo, které by zastřešovalo témata, objevující se v této práci, asi nejvýstižněji znějí výrazy jako hledání vazeb a propojení.

A nejedná se pouze o propojení psychologie s ostatními obory, které se vyjadřují k problematice lidské mysli, ale také o propojení jednotlivých oborů mezi sebou a hledání společných „překrytů" a možností vysvětlení jedněch druhými. Přesné vymezení hranic mezi obory ztrácí smysl, pokud teoretické, metodologické či technické možnosti neumožňují jednomu z nich plnohodnotné pochopení zkoumaného jevu, kterým není nic menšího, než lidská mysl, potažmo schopnost existovat v prostředí „účelným" způsobem – inteligence.

Odvětví, které si za svůj cíl tuto oblast vytyčilo, má již své konkrétní jméno. Můžeme jej nazývat kognitivní věda či kognitivní vědy. Vědní obor integrující v sobě poznatky z mnoha oblastí bádání a výzkumu, snažící se vytvořit jednotnou teorii (popřípadě její aplikaci v praxi), umožňující pochopení zmíněné problematiky v celé její šíři.

Obr.1 *Vědní obory konstituující předmět kognitivních věd.*

Psychologie je oborem, který může k dané problematice mnohé říci. A jelikož je psychologie mým studijním oborem, pokusil jsem se své téma zpracovat s ohledem na psychologii, ale v kontextu oborů, které mohou poskytnout cenné informace při snaze o vystižení zkoumaného jevu. Výchozím bodem byly poznatky, které jsem studiem ve škole nezískal. Lépe řečeno, hledání poznatků doplňujících mezery pro nalezení širší souvislosti při tvorbě své práce. Propojení znalostí do jednoho celku.

Takový styl práce se setkává s obtížemi. Každý z oborů má svou vlastní terminologii a také specifickou formu vyjadřování, což může někdy působit dojmem jisté svébytnosti a samostatnosti. Také množství poznatků a informací získaných jednotlivými obory během jejich historie činí možnosti dostatečného pochopení jednotlivých oblastí obtížnými. I přesto mi připadá takový styl práce zajímavý a potřebný.

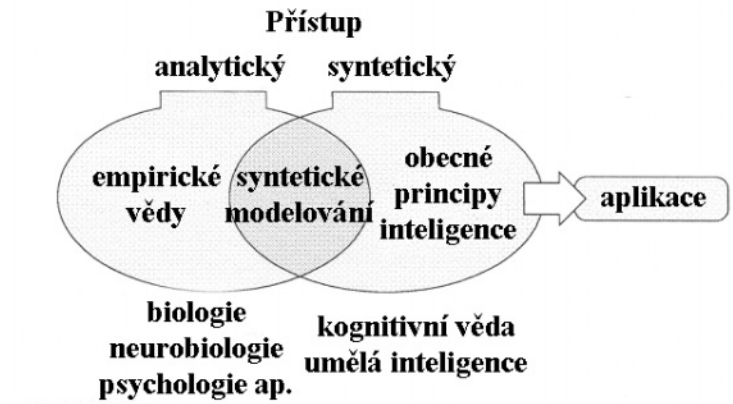
Samotná práce si klade za cíl bližší orientaci v oblastech, které se zabývají problematikou napodobování inteligentního chování, popřípadě možnostmi tvorby umělých inteligentních systémů. První část je věnována vymezování inteligentního systému z hlediska poznatků různých oborů. Jedná se o základní přehled, jelikož rozsah této práce neumožňuje přesnější rozbor konkrétních problému. Následující kapitoly se věnují popisu jednotlivých přístupů při simulaci inteligence a jejich omezení plynoucích z použité architektury, či způsobů jejího využití v konkrétních aplikacích. Závěrečná kapitola je věnována tvorbě významů, která se ukazuje jako nejslabší místo současných systému pokoušejících se o napodobování inteligence. Samotný závěr pak tvoří pomyslná otevřená vrátka pro možnost další práce s využitím poznatků, získaných při tvorbě této diplomové práce.

1 INTELIGENCE

Je obtížné začínat první kapitolu práce termínem, který je jejím cílem a shrnutím.Vyžaduje znalost kapitol následujících. Přesto je nutné zmínit pojem inteligence hned na počátku. V této práci se jím budeme zabývat v širším kontextu, než jak jej chápe psychologie. Inteligence není studována pouze jako lidská schopnost adaptivně se chovat a prožívat, ale jako obecná schopnost libovolného organismu jak živého, tak umělého.

Pokusy o definici lidské inteligence mají dlouhou historii, která je spojena s rozpory, zda lze spojit faktory tvořící inteligenci do jednoho obecného (lépe řečeno existuje pouze jediný faktor), či ponechat členění na jednotlivé oblasti. Otázku je možno zodpovědět až tehdy, budeme-li znát mechanismy, které jsou zodpovědné za adaptivní chování a operaci s reprezentacemi a veškeré ostatní mechanismy tvořící schopnost inteligentního jednání. V současné době je odpověď na tyto otázky neznámá. To by znamenalo konec práce hned v začátku, a proto zkusme pokračovat.

Pokud se přikláníme k názoru, že je třeba v oblasti kognitivních věd, a tedy i při výzkumu inteligence, identifikovat nejmenší univerzální jednotku zodpovědnou za adaptivní formu reagování, přikláníme se k názoru výzkumníků, kteří vidí inteligenci jako jedinečný „g" faktor, tedy inteligenci, přizpůsobující se prostředí a konkrétněji typu úlohy, před který je mysl či mozek postaven. V oblasti simulace tomuto trendu odpovídají neuronové sítě s univerzální jednotkou neuronem, popřípadě Turingův univerzální stroj (jako komputační zařízení schopné vypočítat libovolný typ úlohy) na straně druhé. Univerzálnost na úrovni komplexního inteligentního systému (umělého) je zatím v praxi neuskutečnitelná, ale jako výchozí teoretické principy lze zmíněné přístupy akceptovat. Chceme-li nazírat inteligenci jako soubor více modulů, přičemž každý je specializovaný na jeden typ úlohy (což nevylučuje, že specializované mechanismy mohou být tvořeny základní univerzální jednotkou), je takto inteligence chápána v simulačních přístupech založených na modularitě, subsumpci apod. Zmíněné oblasti simulace tvoří hlavní část této práce a v následujících kapitolách se s nimi seznámíme podrobněji.



Obr. 2 *Přístupy ke studiu inteligence*

Jiný typ dělení přístupů je založen na metodě, kterou používáme při výzkumu inteligence. Přístupy ke studiu inteligence se tak dají rozdělit na analytické a syntetické. Analytický přístup je blízký biologii, neurobiologii a vědám, které stavějí na empirických poznatcích. Hlavním úkolem je identifikovat jednotlivé komponenty systému a rozpoznat jeho strukturu a funkce. Syntetický přístup se naopak snaží sestrojit umělý systém, který by měl vlastnosti zkoumaného. Přístupy jsou k sobě ve vztahu komplementárním. Současná kognitivní věda používá princip, který je spojením obou předchozích přístupů a směřuje do aplikované oblasti. Nazývá se syntetické modelování a jeho hlavním mottem je „pochopení skrze vytváření", tedy kompromis, který vychází z omezených znalostí, které o inteligentních systémech víme, jež jsou aplikovány při tvorbě modelů těchto systémů. Chybějící znalosti se snažíme překonat metodou pokusů a omylů se zpětnou vazbou (Pfeifer&Scheier, 2001,s.154). Konkrétní aplikací syntetického modelování je např. práce profesora de Garise, který se snaží syntetizovat poznatky z oblasti neuronových sítí i klasické komputace. Svůj přístup nazývá *Brain building*. Základem architektury jsou specializované počítače využívající architektury neuronových sítí, které jsou poté propojovány do vyšších funkčních celků jako stavebnice. Taková syntéza se velmi blíží oblasti paralelních počítačů.

S možnostmi využití psychologických přístupů pro oblast simulace inteligence se můžeme setkat i ve starších pracích, u kterých bychom to nepředpokládali. Paralely s komputačním přístupem se objevují například v Piagetově Psychologii inteligence. V době, kdy se oblast programování teprve formovala, dokázal Piaget v lidském chování a prožívání (navíc obohaceném o faktor ontogeneze, který komputační přístup příliš nereflektuje) identifikovat principy, které lze přirovnat k počítačovým pro-

gramům. Vidí je jako posloupnosti operací, složené ze sekvenční spolupráce operátorů (algoritmy), které dohromady tvoří grupy (program, podprogram). V behaviorismu by byly části sekvence nazývány operanty.
Piaget ale rozlišuje mezi jednotlivými grupami (sekvencemi). Existují zvratné a nezvratné (Piaget, 1998). Z hlediska ontogeneze dochází nejdříve ke tvorbě nezvratných, až poté ke grupám zvratným, se kterými může být manipulováno nezávisle na prostředí (například akauzálně, reverzibilně). Tento princip se používá jen v některých oblastech komputace. Převažuje spíše oblast neuronových sítí, například metodou *backpropagation* (viz Konekcionismus), sloužící převážně k mechanické korekci průběhu zpracování informace. Pro Piageta je zvratnost jedním ze základních předpokladů pro vznik inteligentního chování. Při tvorbě inteligentních systémů je nutné, aby si inteligence při stoupajícím počtu operací zachovávala svoji „zvratnou pohyblivost" (Piaget, 1998).
Pro vznik inteligentního jednání je podle něj také třeba součinnosti dvou procesů. **Primární činnost** je interakce subjekt objekt, skrz které je možno vidět inteligenci jako strukturaci objektu či objektů subjektem. **Sekundární činnost** je vztah subjektu ke své činnosti, tedy emoce, dynamika, motivace. Oba procesy fungují spolu dohromady a nedělitelně (Piaget, 1998). V Piagetově práci již tedy nalezneme požadavek vědomí, intencionality a potažmo subjektivity pro konstituci inteligentního chování, které budeme muset v této práci často opomíjet, což je způsobeno povahou zmiňovaných přístupů.

Z novějších teorií stojí za zmínku např. Greenspanova definice inteligence, která je svou abstraktností blízká spíše filosofii mysli (pro oblast konkrétní simulace je zatím vzdálenou metou). Greenspan vidí inteligenci jako schopnost produkovat intence a ideje a schopnost dostat je do logického a analytického rámce (Greenspan, 1996). Bohužel se nevyjadřuje o mechanismech, které by tyto schopnosti měly produkovat.
Přestože se různé způsoby definice liší, v základě hovoří o inteligenci jako schopnosti či kapacitě. Většinou je problematická otázka toho, zda-li je inteligence determinována okolím či nikoliv. To se projevuje již v samotném vymezování pojmu. Například Terman hovoří o schopnosti abstraktně myslet, zatímco Peterson o biologickém mechanismu. V mnoha výzkumech bývá úloha prostředí často přehlížena (Pfeifer&Scheier, 2001).
Možné vymezení procesů, kterými se inteligentní systém projevuje, můžeme nalézt v Lugerově přehledové publikaci.

1. Reaktivita jako schopnost systému vytvářet k příčinám prostředí, následky v podobě prožívání či chování. Limitujícím faktorem (stejně jako u adaptace) je pak citlivost systému vůči prostředí. Citlivost je následně brána jako míra kvality a intenzity.
2. Diskriminace a generalizace jako mechanismy, které patří v oblasti simulace k nejobtížněji napodobitelným.
3. Komplexita jako schopnost adaptace organismu je odvislá od organizační komplexnosti.
4. Adaptace a učení. Hodnocení adaptace je uvedeno výše. Definice učení je stejně obtížná jako samotný pojem inteligence.
(Luger, 1994)

1.1 Pojem inteligence

V laické mluvě jsou slova jako vnímání myšlení, vědomí, úmysl, vůle, inteligence a chtění používána tak, že není třeba mít tyto výrazy přesně definovány. Lidé, kteří spolu komunikují, je mají subjektivně ukotveny, neboť souvisejí s jejich prožíváním. Pokud se výše zmíněná slova stanou oblastí vědeckého zkoumání, nastává problém, jelikož je téměř nemožné je definovat v celé jejich šíři. Souvisí to s mírou abstrakce, která tato slova mají a také s jejich subjektivitou. Posledním problematickým místem je, že procesy, které jsou jejich součástí, jsou z části nevědomé a pro současnou vědu neprozkoumané. Z čehož plyne, že pokud použijeme daná slova ve formě výroků, nemůžeme je ověřovat ani dokazovat. Můžeme o nich hovořit pouze s odvoláním se na **intersubjektivní shodu** (Havel, 2001).
Definice inteligence se také setkává s obtížemi, vyplývajícími z hierarchického uspořádání pojmů, které potřebujeme k jejímu vysvětlení. Při definování pole působnosti se bohužel neobejdeme bez pojmů myšlení, svět, reprezentace, prostředí, popřípadě generalizace, kategorizace, abstrakce apod. Přesné vymezení těchto pojmů je bez určité míry redukce nemožné. Pokud jsou pojmy hierarchicky uspořádány a nám se nedaří pojem postavený hierarchicky níže, definovat s přesností, která je dostačující, pojem hierarchicky výše bude již tuto nepřesnost obsahovat s tím, že díky jeho komplexnosti bude u něj nepřesnost vzrůstat. Tak se dostáváme až k pojmům jako je inteligence (a níže zmíněné vědomí) a neurčitost ve významu se kumuluje do podoby, která jej neumožňuje používat jako platnou definici.
Snaha o definici inteligence je nesnadná i proto, že se zatím musíme obejít bez pojmu vědomí, které se jeví jako nedílný předpoklad pro postulování inteligence. Dostáváme

se tak do obtíží, kdy se pojem inteligence snažíme vysvětlit pomocí elementárních principů,bez zastřešujícího pojmu vědomí. Přístupy popisované v této práci, jsou většinou založeny právě na elementárních principech a snad i díky tomu je vědomí věnována jen malá část textu práce.

Výzkumy v oblasti vědomí ukazují, že není jednoduché nalézt model či způsob vysvětlení, který by byl schopen tento fenomén uchopit. Velmi zjednodušeně můžeme vědomí vyjádřit Sutherlandovým výrokem."Vědomí je fascinující, ale prchavý jev; je nemožné určit, čím je, co dělá, ani proč vzniklo. Nebylo o něm napsáno nic, co by stálo za čtení" (Sutherland,1989,s.81).

V rovině spekulace by bylo možné obejít výše zmíněné argumenty tím, že hierarchické umístění pojmu vědomí není přesně dané, a že se tedy vědomí vyskytuje už u základních prvků živých inteligentních systémů (neuronů).

1.2 Modely inteligence

Z hlediska modelování zaznamenala psychologie mnoho pokusů o tvorbu modelu, který by byl schopen vysvětlit inteligentní chování v celé jeho komplexitě. Sternberg ve své práci shrnuje některé dosavadní psychologické modely inteligence a třídí je následujícím způsobem (Sternberg, 2000,s. 141).

Hierarchické modely inteligence	Kontextuální modely	Modely komplexních systémů
Jejich nevýhodou je, že nerozlišují rozdíl mezi genetickými faktory a faktory výchovy.	Jsou kulturně podmíněné, přikládají kontextu klíčovou úlohu. Problematická zůstává definice kontextu.	Zaměřené na fyziologické a kognitivní komponenty inteligence. Jedná se propojení předchozích systémů.

Tab. 1 *Modely inteligence*

1.3 Statistické a biologické měření inteligence

Dosavadní práce, které se zabývají měřením inteligence, používají jako nejčastější metodu měření formu testu. Hodnotícím kritériem je obvykle relativní vztažná soustava, jejíž průměrná hodnota je odvozena od nejčastějšího výskytu míry inteligence (například ve formě IQ) v populaci. V takovém případě je ale inteligence již předpokládána u zkoumaného organismu a nedozvídáme se tedy nic o vnitřních mechanismech. Umožní nám statistické srovnání vysouzených schopností. Pokud je test členěn do jednotlivých subtestů, dokáže navíc rozlišit poměr mezi jednotlivými dovednostmi, či typy úlohy, které jsou typické pro lidské jedince (například prostorová představivost, matematické operace apod.). Podle třídění z první kapitoly se jedná o analytický přístup se snahou potvrdit platnost teoretického modelu.

V případě zkoumání inteligence z hlediska biologického či neurobiologického patří mezi nejčastěji používané metody:

1. Neurální výkonnost – Měření křivky mozkové aktivity při jednoduchých stimulech. Používá se evokovaného signálu a následně se křivka analyzuje.

2. Neurální adaptabilita – Například jak rychle poznají ZO rozdíl mezi čarami podobné délky (ztotožnitelná spíše s procesy kategorizace či diskriminace)

3. Metabolismus cukru v mozku

4. Podle rychlosti přenosu vzruchu – Při měření rychlosti přenosu vzruchu v periferních nervech je přímá úměra mezi nárůstem rychlosti a inteligence. V mozkových centrech je tento výzkum zatím v počátcích.

(Sternberg, 2000)

Nevýhodou technik je, že jsou redukcionistické a z jejich výsledku se nedá odvodit mnoho o detailním fungování systémů. Změnou oproti předchozí metodě člověk/společnost je porovnávání člověk/kvalita biologického substrátu.

2 INTELIGENTNÍ SYSTÉM

2.1 Filosofie mysli

2.1.1 Teorie mysli (TOM)

Teorie mysli (TOM) tvoří zastřešující pojem ve výzkumech týkajících se inteligence a mysli obecně. Základní otázky této oblasti vycházejí z filosofie, která je nejstarší vědní disciplinou, zabývající se fungováním lidské psychiky. Díky tomu je míra abstrakce a obecnosti teorie mysli vysoká, je nutné doplňovat a konkretizovat jednotlivé části pomocí specifičtějších vědních oborů jako biologie či psychologie. Samotný pojem lidské mysli tvoří v hierarchii pojmů týkajících se mentálních schopností jedince pomyslný vrchol. Otázky po struktuře a funkci lidské mysli (ty doložitelné) mají za sebou víc než dva tisíce let trvající historii. Právě otázky povahy lidské mysli slouží k vymezení rámce zkoumání, ve kterém se pohybujeme a vedly k ustanovení specializovaných vědních oborů, jejichž cílem je základní východiska (nejčastěji ve formě otázek) zpřesňovat do takové míry, která by umožnila dostačující pochopení a následně případnou tvorbu mysli umělé.

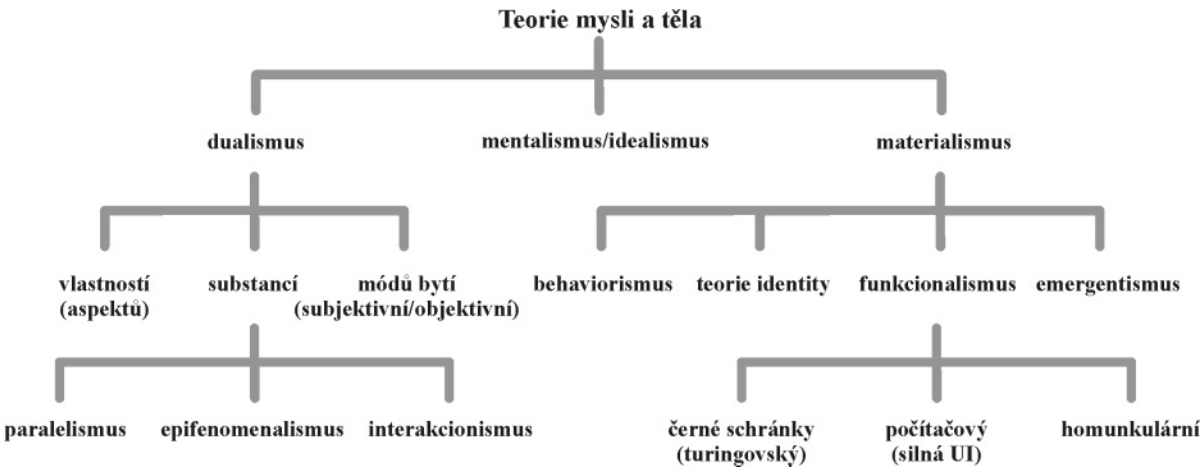
Aristotelova empirická orientace připravila podmínky pro monismus, teorii, která popisuje duši a tělo jako jeden prvek a celou realitu jako jednotu (Sternberg, 2001). Vyskytovaly se i teorie dualistické, ale jejich exaktní popis se objevuje až ve středověku a je spojován se jménem René Descartes.

Descartovu koncepci dualismu nazývají moderní filosofové substanční dualismus (ve 20. století Simon přichází s pragmatickým dualismem (Sternberg, 2001): věří, že mysl závisí na procesech mozku, ale toto spojení je tak komplexní, a víme o něm tak málo, že má větší smysl postulovat mentální zákony, které jsou na mozku nezávislé).

Kant redefinoval problematiku těla a duše. Začal mezi hledat nimi **vztah**. Předchozí zájem byl soustředěn pouze na to, jak mysl kontroluje tělo. Místo hledání duality či jednoty vytvořil základní schopnosti, neboli mentální síly: vnímání, porozumění a myšlení. Vnímání je blízké tělu, myšlení mysli a pochopení je propojuje dohromady. Také se vyjádřil k přístupu zkoumání mysli. Navrhoval sjednocení mezi racionalismem a empirismem. Empirické zkušenosti nazýval **aposteriorní**, tedy získané až po samotném zážitku. Ty, které jsou nezávislé na empirii, jsou nazývány **apriorní**. Celková zkušenost je tedy syntézou těchto dvou, tedy syntézou vrozeného a získaného(Sternberg, 2001).

Přístupy filosofů k lidské mysli slouží jako dobrá ukázka neustálé (ale i neuspořádané)

tendence zpřesňování komponent a mechanismů, které jsou hybateli lidské mysli. Množství přístupů, které se vyjadřovaly k tomuto fenoménu, je velmi různorodé a nejlépe asi názorovou pestrost vyjádří následující diagram (Havel, 2001).



Obr. 3 *Dělení přístupů ke studu mysli*

Argumentační možnosti daných směrů však také souvisí s empirickou rovinou a stavem poznání. Právě ty mají vliv na omezování platnosti jednotlivých přístupů a identifikaci jejich limit při aplikaci. Bohužel ani současné znalosti nám neumožňují odpovědět na základní otázku, kterou si kladl již Aristoteles. Tedy zda je za vznik mysli zodpovědná pouze duše nebo je nutné i fyzické tělo, popřípadě je nutná přítomnost obou. A jestliže je platné třetí tvrzení, jaký je vztah mezi těmito entitami, hierarchický, rovnocenný či komplementární? Nejedná se ani tolik o záležitosti metafyzické. Spíše je mysl fenoménem tak komplexním, že ani současné technologie nevládnou kapacitou a možnostmi získat, zpracovat a interpretovat data, která vypovídají o práci a schopnostech lidské mysli (mozku).

O možnostech zkoumání mysli z hlediska současných vědních oborů se vyjadřuje například Gazzaniga. Tři vědní obory se vyjadřují k problematice mysli, ale každý vysvětluje pouze určitou část předmětu výzkumu. Psychologie a filosofie se bez dostatečného empirického podkladu z chování snaží odvodit teorii, která by byla univerzální. Není ale schopna své poznatky biologicky ukotvit a zbývá jí pouhá statistika jako měřítko úspěchu. Neurovědy svou mravenčí prací rozdělily mozek na atomy, poznaly jeho strukturu, ale selhávají při syntéze částí v celek. Umělá inteligence se snaží zákony logiky a matematiky simulovat procesy na hardwaru, který je strukturou

i funkcí nepodobný lidskému mozku (Gazzaniga&Mangun, 1998).

Jiného názoru je Johnson-Laird. Psychologie (studium programů) může být uskutečňována nezávisle na neurofyziologii (studium strojů a strojového kódu). Neurofyziologický substrát musí poskytovat fyziologický základ pro procesy mysli a také zajistit dostatečnou výpočetní sílu pro rekurzivní funkce, jejichž fyzikální základ neklade žádné omezení pro vzorce myšlení (Johnson-Laird, 1980).

Zajímavým příspěvkem do diskuse o způsobech směřování kognitivních věd je Newellův článek z počátku 70. let, který se jmenoval „Nemůžeš hrát 20 Otázek s přírodou a vyhrát". 20 Otázek je dětská hra, ve které musí jeden hráč získat podle odpovědi druhého požadovanou informaci . Newell jí využil jako metaforu pro otázky vědců na povahu přírodních dějů. 20 otázek, které položil, se týkalo implicitních znalostí, které produkují dichotomie prostupující současnou kognitivní vědou. Například zda je interní reprezentace propoziční nebo obrazová? Vyvoláváme zpětně informace z dlouhodobé nebo krátkodobé paměti? Je pozornost organizována prostorově nebo objektově? Podle něj odpovědi na tyto otázky, pokud bychom je znali, v sobě obsahují nedostačující informace pro následná zkoumání. Suma těchto odpovědí nedokáže pomoci tvorbě jednotné teorie kognice. Psychologie není schopna vyhrát s přírodou tuto hru. Proto je třeba použít alternativní postupy. Analytický přístup je pro něj nedostačující a navrhuje přechod k syntetickému. Newell navrhuje vyjít z teorie kognitivní architektury, použít dostupné znalosti o funkci lidského kognitivního systému a poté za použití výše zmíněné teorie integrovat poznatky do oblasti výzkumu. Nutnost empirického výzkumu patří k základním metodám, které Newell zastával. (Sternberg, 1999).

2.2 Psychologie

2.2.1 Horká a studená metodologie

Dělení základních metod v oblasti Teorie mysli (aplikovaněji pak tvorby kognitivních architektur) uvádí Sedláková ve své přehledové publikaci. Jsou to:

1. Konstrukce mentálních modelů

2. Nápodoba (simulace) kognitivního obrazu zprostředkovaného chováním

3. Kombinace 1 a 2, modularita

(Sedláková, 2004).

Již v počátku je nutné rozšířit použití těchto metod na celou komplexitu kognitivního aparátu. Kognitivní psychologie jako odvětví psychologie vycházející z pojetí mysli coby informaci zpracovávajícího systému používá pro své zkoumání mnoho metod. Ve svých ranných fázích se však vyznačovala tím, že nebrala v potaz roli emocí a motivace. Je to dáno jejich nesnadnou definicí a obtížným vymezením právě v kontextu informačního přenosu a zpracování. Existují přístupy, které se snaží operovat s modely obsahující prvky typu *Desire* a *Belief*, ale i ty se ukázaly jako nedostačující pro tvorbu komplexních modelů mysli. Proto se objevuje požadavek autorů na tvorbu metodologie, která se nazývá horká (Sedláková, 2004). Studenou metodologií je nazýván právě předchozí přístup, který obsahuje pouze zpracování informace na základě pravidel logiky, či tvorbou složitějších operátorů, to vše pouze v syntaktické rovině. Horká metodologie se snaží začlenit do procesu také emoční podkres, motivaci a metakognici (cokoliv to slovo znamená).

Thagard ve své nové koncepci teorie mysli zvané *CRUM* tvrdí, že emoce mají podstatný význam při operacích jako je hodnocení, zpětná vazba a rozhodování. Bez emocí jsou procesy simulovatelné jen na základní, sériové úrovni. Se vzrůstající komplexností se nepřítomnost těchto markerů projeví neschopností „postihnou situaci". Při komplexnějších dějích totiž musí docházet k autoregulaci, určitému usměrňování operací, ve smyslu ovlivňování paralelně probíhajících procesů. Emoce usměrňují propustnost architektury (Thagard, 2001).

2.2.2 Kognitivní Architektury

Současný výzkum kognitivního modelování je založen na budování a následném testování kognitivních architektur. Kognitivní architektury jsou přístupem k analýze a tvorbě modelů systémů (či jeho částí). Architektura je tvořena pevně danou skupinou výpočetních mechanismů a prostředků, které tvoří podklad pro lidskou kognici. Architektury jsou většinou vytvářeny na základě předpokladů či teoretických východisek, takže jejich podoba je velmi komplexní a může být tvořena více přístupy a kombinací rozličných mechanismů (Harnad, 1990). V psychologii je tento přístup oblíben, jelikož dokáže ze získaných empirických dat určitého fenoménu vytvořit model, který identifikuje strukturu a mechanismy, které jej tvoří. V určitých případech se podaří identifikovat neurální koreláty, které jsou podobné vysouzeným teoretickým modelům. Toto kriterium však není bráno jako nezbytné pro jejich úspěšné fungovaní. Je to dáno postupem „shora-dolů". Nevýhodou oproti komputačnímu přístupu či neu-

ronovým sítím je právě jejich neukotvenost (architektury neobsahují popis svého fungování vedoucí až k základním jednotkám – „atomům"), respektive jejich atomy jsou ztotožnitelné se základními moduly.

Pokud hovoříme o teoriích zpracování informací myslí, je nutné rozlišovat mezi teoriemi architektur a teoriemi výpočetními. Teorie architektur hovoří o procesech mysli, jako funkční kapacita STM a LTM. Ty by měly být podpořeny neurologickými výzkumy. Výpočetní teorie odhalující algoritmy jsou mnohem obtížněji řešitelné. (Sternberg, 1999). Kognitivní architektury jsou v rovině simulace spíše identifikovatelné s výpočetními teoriemi (nejedná se přímo o požadavek „čisté" matematické formalizace, ale pouze algoritmizovatelnost). Nemusí být aplikovatelné na konkrétní typ hardwaru či wetwaru, protože v rovině základních prvků, tvořící jejich jednotlivé části, již nejsou definovány. Při jejich simulaci se používá nejčastěji právě počítačů. Činnost architektury je napodobována užitím komputačního hardwaru (v této úrovni není definována), a všechny speciální funkce jsou simulovány softwarově, s omezeními z toho plynoucími.

Díky tomu je kognitivním architekturám v této práci věnován jen malý prostor. Možnost posunu v oblasti tvorby kognitivních architektur je pravděpodobná pouze v případě, že by se poznatky z oblasti neurobiologie a psychologie dostaly do takového stádia, kdyby se poznatky o funkcích mechanismů začaly překrývat a doplňovat, což by vedlo k ukotvení psychologických poznatků na biologický substrát. Současný stav připomíná práci matematiků, kteří znají výsledek a snaží se skládat jednotlivé operátory a čísla dohromady tak, aby tvořili výsledek, aniž by věděli, čemu tyto symboly odpovídají.

2.2.3 Kognice

2.2.3.1 Kategorizace – tvorba konceptu

Nyní se budeme zabývat pojmem, který zastupuje jeden z nejsilnějších mechanismů konstituující inteligentní systém. U člověka je vypracován k dokonalosti, leč bohužel dosud pouze odhadujeme, které procesy jsou za něj zodpovědné. V oblasti simulace je kýženou metou, která by znamenala přechod do oblasti „chápajících" umělých systémů. Nejlepších výsledků dosahujeme použitím neuronových sítích. Z hlediska možností a schopností kategorizačního aparátu jde teprve o začátek.

Introspektivní pokusy o zkoumání kategorizace selhávají na tom, že cílem jejího bádání jsou procesy, které jsou nevědomé a neverbální, což už v začátku neumožňu-

je možnost, jak dané fenomény uchopit (Sternberg, 2001).

V psychologii se setkáme spíše než s tvorbou kategorií či tříd s pojmem konceptu, který uvádí tento proces do psychologického kontextu a ztotožňuje jej s mentální reprezentací. Koncept je nazírán jako mentální reprezentace třídy objektů (kategorií) (Sternberg, 2001).

Mezi základní přístupy ke studiu konceptu patří:

1. **Přístup na základě podobnosti** – vysvětluje pomocí stupně podobnosti ke známému
2. **Přístup na základě vysvětlovací báze** – operuje s předchozí znalostí

Klasický koncept je tvořen základními vlastnostmi, které konstituují jasné ohraničení kategorie. Vlastnosti musí být samostatně nezbytné a společně dostačující, aby obsáhly danou kategorii. Klasický pohled zastává názor, že struktura kategorií má být hierarchická. Podkategorie musí mít všechny vlastnosti nadřazených kategorií (Sternberg, 2001). Naopak to ale nemusí fungovat. V nadřazených kategoriích se některé vlastnosti nižší kategorie ztrácejí. Jsou vypuštěny, protože nadřazené kategorie jsou obecnější a zahrnují více předmětů. Další námitky proti klasickému konceptu jsou shrnuty v následujících bodech:

Prvky dané kategorie většinou nezapadají pouze do ní, ale jsou „fuzzy", tedy že podle svých vlastností náleží do více kategorií.

Objekty nebývají často tak přesně definované, aby mohly být jasně přiřazeny do určité kategorie.

Objekt podřadné kategorie (orel) bývá ztotožňován více se základní kategorií (pták) než s nadřazenou (zvíře), protože s ní sdílí více vlastností.

(Sternberg, 2001)

Klasický přístup není schopen vysvětlit **efekt typickosti .** Tedy že lidé rychleji přiřadí do kategorie jeho typického člena, než atypického (Sternberg, 2001).

Přístupy založené na podobnosti

Nedostatky v klasickém modelu kategorizace vedly ke tvorbě nových teorií. Patří mezi ně přístup založený na podobnosti, do kterého patří pravděpodobnostní přístup a přístup na základě exempláře.

Pravděpodobnostní přístup nepoužívá vlastnosti či podmínky, které jsou **nutné** nebo **dostačující**. Kategorizuje tím, že sdílí určitý podíl vlastností s objekty, které

náleží do stejné třídy. Tím se vyhýbá efektu typičnosti. Základním konceptem pro tuto teorii je termín **rodové podobnosti** (Wittgenstein, 1953). Jedná se o strukturu, která obsahuje množinu příkladů, přičemž každý má alespoň jednu shodnou vlastnost s jedním nebo více příklady dané množiny. Takže úroveň **typičnosti** je podmíněna právě **rodovou podobností.**

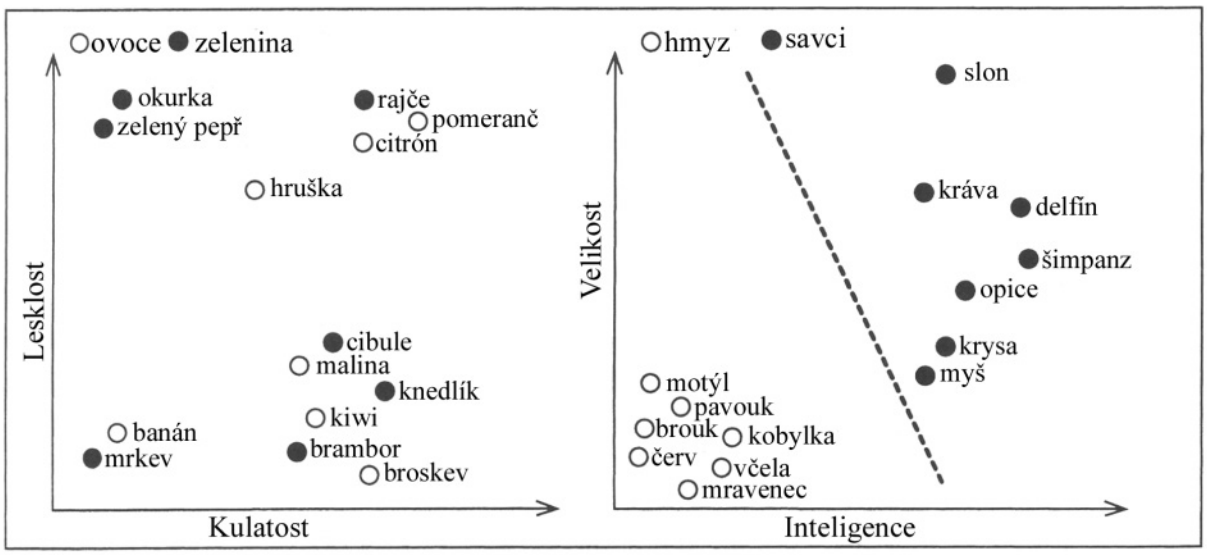
Podskupinou pravděpodobnostního přístupu tvoří **teorie prototypu** (Posner&Keele, 1968). Lidé si vytvářejí reprezentaci centrální tendence kategorie pomocí prototypu. Prototyp v sobě zahrnuje nejčastější vlastnosti, které danou třídu definují. Je použitelný pro svou jednoduchost, tedy že si jej můžeme představit jako jeden hybridní objekt, jehož vlastnosti mohou nabývat omezených hodnot.

Důležitým prvkem při tvorbě kategorií je váha dané vlastnosti, přičemž rozhodující je, zda vlastnost zvyšuje koherenci dané třídy. U člověka jsou váhy, podle kterých přiřazuje objekty kategoriím velmi subjektivní a obtížně zkoumatelné (Sternberg, 2001).

Lineární separovatelnost

Kategorie jsou **lineárně separovatelné,** pokud můžeme roztřídit objekty tak, že po znázornění podle míry vlastnosti (vynášeno v N-dimenzionálním prostoru podle N vlastností) vzniknou seskupení objektů dle vlastností, které lze od sebe oddělit přímkou, rovinou (podle počtu dimenzí prostoru).

Nejčastěji se používá dvou vlastností. Objekty jsou vynášeny podle míry dané vlastnosti, a pokud jsou lineárně separovatelné, můžeme jednotlivé skupiny oddělit přímkou.



Obr. 4 *Lineární separovatelnost*

Možnosti třídění jsou ovlivněny podle vlastností použitých jako osy grafu. V podstatě jde o hledání korelace mezi vlastnostmi. Umožní nám mechanické rozčlenění jednotlivých objektů pomocí kvantifikované míry vlastnosti. Pokud ale je objekt speciálním případem třídy (tučňák – pták, který nelétá), nejsme schopni využít tento způsob jako automatický kategorizátor a musíme provést korekce pomocí dalších mechanismů

Dalším typem je **přístup založený na exemplářích**. Nové objekty jsou porovnávány s exemplářem uloženým v paměti, který zastupuje určitou kategorii. Oproti teorii prototypu je zde v paměti uložen konkrétní model. Prototyp nabízel stupně volnosti ve vyjádření vlastnosti (abstrahoval tedy od reality), zatímco exemplář je jasně daný a definovaný. Jediná abstrakce se provádí, pokud je hledána míra shody nového objektu s exemplářem.
Tento přístup umožňuje vysvětlení kontextu v procesu kategorizace za pomoci míry aktivace exempláře. V různých situacích (v různém kontextu) je váha exempláře menší nebo větší (Sternberg, 1999). Pokud se ale podíváme na prototypový přístup, je zde tato vlastnost obsažena také. Prototyp je abstraktní exemplář a je mu umožněno, aby jeho vlastnosti nabývaly hodnot v daném rozmezí.

Hlavní kritika **přístupů založených na podobnosti** jde shrnout do následujících bodů:
1. Jednotlivé přístupy jsou obtížně rozlišitelné.
2. Nedokáží vysvětlit proces kategorizace a interference v takové šíři, jak probíhá u člověka.
3. Ignorují jinou informaci než podobnost, i když by mohla být důležitá pro proces kategorizace (Sternberg, 1999). Problematicky se také jeví dynamická povaha procesu generalizace (ve formě funkce), kterou nejsou modely založené na podobnosti schopné akceptovat.

Druhým teoretickým zpracováním procesu kategorizace jsou **přístupy založené na explanaci (vysvětlovací báze).** Jejich základem je vyloučení předchozího přístupu (posuzování podobnosti) v procesu klasifikace (kategorizace). Hovoří se v nich o potřebě získat vysvětlovací strukturu při kategorizaci z nějaké obecné teorie (Sternberg, 2001).
U umělých systémů to znamená, že jsou kategorie a proces kategorizace dány

člověkem, který jej navrhl (svět vnímaný systémem je tříděn do těchto předdefinovaných kategorií). Pokud se v prostředí objeví nový prvek, který nepatří do žádné z kategorií, systém s ním nedokáže interagovat, není ho schopen identifikovat (Pfeifer&Scheier, 2001). Kategorizaci tedy předchází uvažování v abstraktní rovině, které se snaží nalézt teorii (u umělých systémů to ještě není možné), která by byla schopna odvodit chybějící kategorii. Omezení vyplývá právě z abstraktní roviny operací. V takovém případě musí probíhat kategorizace na vědomé úrovni a nelze ji považovat za automatický proces.

Ve zkratce uvádím základní přístupy založené na vysvětlovací bázi:

1. Přístup založený na teorii
2. Psychologický esencialismus
3. Idealizované kognitivní modely
4. Kauzální modely

Jaké jsou výhody a nevýhody dvou kořenových přístupů? Explanační přístupy jsou jednodušší pro analyzování a testování, protože jsou diskrétní, logické a explicitní. Přístupy založené na podobnosti jsou daleko odolnější proti chybám a flexibilněji použitelné, protože je výpočet rozšířen na celý systém. Také mohou zpracovávat zcela nové podněty, jelikož jsou schopné generalizace na základě podobnosti (Sloman&Rips, 1998).

Výzkumy v oblasti kategorizace se posouvají v několika směrech. Jednou z nich je výzkum kategorizace způsobem řešení problému, přičemž reprezentace problému se mění v závislosti na zkušenostech systému a jeho interakci se strukturou problému (Ross, 1996). Druhou možností je zkoumat kategorizaci při aplikaci v oblasti neuronových sítí. Používá se metody nekontrolovaného učení, řízené pouze zpětnou vazbou systému (viz Konekcionismus).

Kategoriální reprezentace

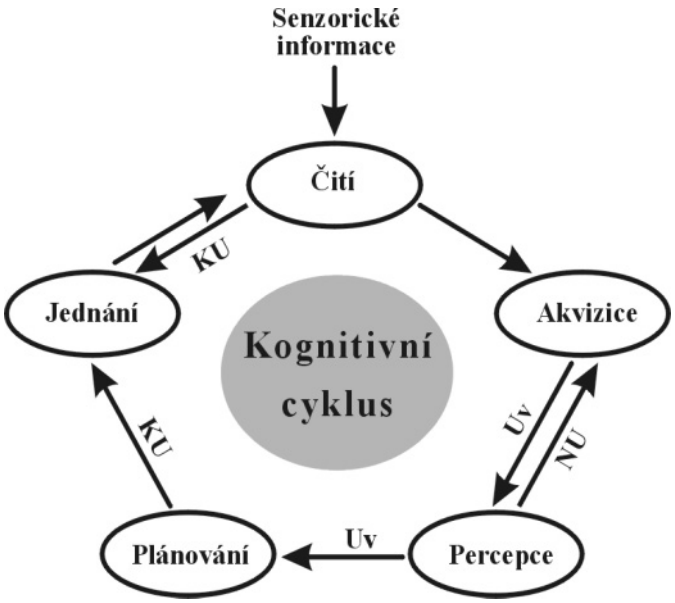
V případě, že pohlížíme na kategorizaci z pohledu automatického mechanismu, patří mezi základní procesy diskriminace (rozlišování) a identifikace. Diskriminaci je možno popsat jako posouzení, zda jsou dva vstupy podobné nebo ne a pokud jsou rozdílné, tak v jaké míře. Diskriminace je relativním posouzením založeným na schopnosti hodnotit odděleně objekty (vstupy) a rozpoznat míru podobnosti. Identifikace je schopnost přiřadit unikátní odpověď – „jméno" – třídě vstupů, které budou následně brány jako ekvivalentní a invariantní. Identifikace je proto soudem absolutním, jelikož

posuzuje podle kategorie, která je pevně dána.

Pokud si vezmeme vzájemný vztah mezi těmito procesy, pak je diskriminace nezávislá na identifikaci. Je možno rozlišovat objekty bez nutnosti vědět, do které kategorie patří. Proces identifikace nám umožňuje odhalit právě invariantní vlastnosti. Jejich aplikací (tedy použitím detektoru na invariantní vlastnosti) získáváme kategoriální reprezentace. Lépe řečeno, **kategoriální reprezentace** je teoretický soubor vlastností, které mohou mít takové hodnoty, že jsou při analýze detektorem invariantních vlastností akceptovány a přiřazeny do dané kategorie (Harnad, 1990). Tento pojem zavádí Harnad jako jednu z podmínek pro tvorbu systému, které by dokázaly sémanticky ukotvit význam symbolů, jež zpracovávají. V dřívějších pracích se novému pojmu nejvíce podobá kategorizace založená na prototypu.

2.2.3.2 Vnímání

Oproti klasickému Neisserovskému cyklu vnímání, který se skládal ze 3 částí, se můžeme setkat s modifikacemi, které operují s pětiprvkovým modelem. Ten slouží lépe pro vysvětlení některých procesů spojených s kognicí a také s kontrolovaným (*supervised*) či nekontrolovaným (*unsupervised*) učením, se kterým se podrobněji seznámíme v kapitole o neuronových sítích. Přesnější popis jednotlivých částí modelu je uveden v následujících odstavcích (Konar, 1999).



Obr. 5 *Cyklus vnímání: KU - kontrolované učení, NU - nekontrolované učení, Uv - Uvažování*

1. Čití (v klasické terminologii odpovídá spíše pojmu vnímání) – tohoto termínu se používá ve významu recepce a transformace signálu do měřitelné formy. V širším významu (vnímání) se k tomuto procesu přiřazuje také předzpracování informace (odstranění šumu, vydělení důležitých prvků jako tvar, barva) a uložení do krátkodobé paměti. To neplatí pro Tveterův model paměti, který předpokládá přímý vstup senzorů (po transdukci) do dlouhodobé paměti

2. Akvizice (osvojení) – porovnává obsahy v krátkodobé paměti (STM) s informacemi, které jsou permanentně uloženy v dlouhodobé paměti (LTM), která může být novými informacemi modifikována. Tato úprava obsahu se nazývá vylepšení znalosti a je spojována s nekontrolovaným *(unsupervised)* učením (NU) , tedy když systém nepotřebuje interpretátora prostředí, ale dokáže sám rozeznat změny mezi svými interními obsahy a prostředím, porovnat je a posoudit, zda-li by mělo dojít k vylepšení znalosti. Zmíněné procesy bývají často nevědomé.

3. Vnímání (v klasické terminologii spíše myšlení) – jde o proces konstrukce *„high-level"* znalostí pomocí informací získaných na *„low-level"* úrovních a jejích následná organizace a kategorizace do strukturální podoby tak, aby mohly být efektivně používány pro operace s interními reprezentacemi v *„high-level"* úrovni. Jedná se opět o automatický proces. Nejlépe se při modelování používá sémantických sítí.

4. Plánování – v sobě obsahuje posloupnost akcí, které z počátečního stavu jsou schopny dosáhnout stavu cílového. Hlavním úkolem je nalézt a identifikovat patřičné znalosti či jejich části a ty následně aplikovat při řešení problému (tvorbou kroků). Důležité je rozlišení mezi plánováním a uvažováním, což jsou odlišné termíny. Uvažování může pokračovat i v průběhu vykonávání akcí (souběžně), zatímco plánování je proces naplánování kroků akcí, přičemž jejich vykonávání probíhá až následně. Daný rozdíl spíše ustanovuje hierarchii těchto termínů, tedy že plánování je jen určitou formou uvažování.

5. Jednání – v této fázi se jedná o vykonání plánovaných posloupností pomocí *„low-level"* mechanismů. Posloupnost jejich vykonávání je dána úrovní plánu. V této úrovni je možné aplikovat postupy kontrolovaného učení (KU), díky označení míry efektivity a její rozčlenění na jednotlivé úseky plánu, čímž poskytujeme systému zpětnou vazbu (Konar, 1999). Zohlednění typů učení a jejich možností v procesu kognice či zpracování informace inteligentním systémem, činí tento model nejlépe využitelný právě v oblasti neuronových sítí.

2.2.3.3 Myšlení

V mnoha přístupech bývá slovo „myšlení" spojováno s pojmem vědomá mysl. To odpovídá nazíráni inteligence, jak ji viděl Terman a v podstatě vede až ke tvrzení René Descarta „Cogito ergo sum". V oblasti simulace se klasický pojem myšlení téměř nevyskytuje. Simulace procesu myšlení v sobě obsahuje tak různorodé mechanismy, které jsou studovány samostatně a často nedochází k integraci do jediného pojmu myšlení. V kognitivních vědách je myšlení děleno na jednotlivé podúlohy (např. řešení problémů, rozhodování, uvažování, popřípadě ještě přesněji na generalizaci, kategorizaci apod.). V oblasti simulace dochází k podobné strukturaci. Myšlení je tedy pro aplikovanou oblast příliš silným pojmem, čemuž odpovídá i názorová pestrost na vymezení daného pojmu.

Hofstadter tvrdí, že mezi *low a high level* procesy (myšlení) leží tolik vrstev, že tyto dvě vrstvy nemají možnost vědět, co se děje na té druhé (Hofstadter, 1999).
Ulam vidí myšlení (zavřené) jako iterativní proces se vzorcem růstu (Hofstadter, 1999). Myšlení je možnost fragmentace a restrukturace kauzálního kontinua v reprezentované formě.

Myšlení a kreativitu lze považovat jako určitou formu deformace, schopnosti pozměnit vnímanou skutečnost tak, aby odpovídala záměru, který si vytýčil organismus pro její přeměnu (Boden, 1988). V podstatě je lze přirovnat k Freudovým obranným mechanismům jako regrese, sublimace, týkajících se oblasti vnímání sebe sama (reflexi vlastního vnímání a prožívání). Deformace ale může být přítomna již při pouhém vnímání (zde mu více odpovídá pojem centrace vnímání) nebo při vybavování či mentální operaci s reprezentacemi.

2.3 Biologie

2.3.1 Biologické versus umělé systémy

Obtížně se rozhoduje také v problematice základního substrátu (nejjednodušší prvek architektury). Otázka zní, zda-li musí být tento substrát biologické povahy (Pfeifer&Scheier, 2001). Dlouhou dobu panoval názor, že „pravá inteligence" se vyskytuje pouze u biologických mozků (které jsou tvořeny z prvků živé přírody). Na druhou stranu nebyl objeven důkaz, který by říkal, že není možné principielně zvládnout

na nebiologickém (umělém) substrátu to, co lze provést na biologickém. Jediný zřetelný rozdíl je ve způsobu napodobování přírody (užití prvků neživé přírody). Například model neuronu, který se používá dodnes, je redukcí neuronu biologického. U umělého neuronu nejsou obsaženy mechanismy známé u biologických jako metabolismus, tvorba neurotransmiterů, transport živin, závislost na okolním prostředí (příjem a výdej látek). Dalším nedostatkem umělých neuronů může být například jejich neschopnost napodobit práci transmiterů v extrémních (či omezených) podmínkách ve smyslu simulace komplexních systémů. Nevyskytují se pokusy vytvořit (komplexní) systém, který pracuje s nedostatečným množstvím transmiterů apod. Námitkou může být, že nám stačí simulovat neurony v optimálních podmínkách a zkoumání jejich činnosti v abnormálních podmínkách již není potřebné, jelikož výsledky lze za optimálních podmínek extrapolovat. Do jaké míry jsou námitky proti redukci biologického neuronu opodstatněné se ukáže při budoucích simulacích .

Nabízí se otázka, jak by fungoval biologický neuron, pokud bychom mu zajistili dostatečnou výživu, ale žádnou možnost provádět činnost. Pokud by tento stav znamenal pro něj smrt, lze předpokládat, že činnost je nutnou podmínkou jeho života. uměle vytvořených neuronů (na nebiologické bázi) taková vlastnost neexistuje. Již v základní úrovni jednotlivého neuronu se můžeme setkat s rozdílem, který v dalších důsledcích může vést k problémům se simulací. Jde o připomenutí klasického rozdílu mezi živým a neživým. Důsledek se promítá do funkční roviny, tedy oblasti, o kterou má simulace velký zájem, ale přesto nebere při svých vyvozováních zmíněnou rozdílnost v potaz.

Zajímavé jsou otázky ohledně podobnosti biologické smrti a „vypnutí" hardwaru . Hofstadter nazírá myšlení (software) jako reprezentaci reality v hardwaru mozku. Pokud ale organismus zemře, tak hardware zůstává, ale reprezentace zmizí. Vysvětlením může být, že pokud hardwaru chybí napájení, stává se nefunkčním a není ani schopen reprezentace v sobě udržet (Hofstadter, 1999). U lidského mozku ale nemůžeme přesně hovořit o softwaru, který přežije smrt hardwaru (wetwaru). Je to dáno povahou architektury. V případě neuronových sítí je softwarem (dle prezentacionistů) prostředí, které je zpracováváno v organismu pomocí vah neuronových spojení a jejich funkci (formou distribuovaných reprezentací). Dalo by se říci, že s hardwarovou smrtí mizí kopie softwaru (reprezentace), ale software (prostředí) přežívá, protože je na organismu značně nezávislý.

Existuje určitá souvztažnost mezi biologickou komplexitou a jednoduchostí algoritmů. Čím více chceme pomocí simulace dosáhnout plausibilnějšího modelu, tím více

výpočetní kapacity potřebujeme. Při složitých modelech však nutně musíme abstrahovat a redukovat. Otázkou ale zůstává, jestli při této abstrakci nepomíjíme vlastnosti, které jsou právě nezbytné pro tvorbu inteligentního chování. Například u neuronových síti nejsme schopni napodobit biologické sítě tak, aby obsahovaly v dostatečné míře vlastnosti jako generalizace, robustnost a paralelismus (Pfeifer&Scheier, 2001), což jsou omezení spíše technologické povahy.

2.4 Umělá inteligence

Umělá inteligence je empirická věda, která se zabývá výzkumem a napodobováním inteligentních projevů. Nejčastěji pomocí abstrakce a modelování inteligentních projevů mimo médium lidské mysli. Inteligentními projevy podle Feigenbauma rozumíme např.: učení, řešení problémů, porozumění jazyku, uvažování. Marvin Minsky, jehož definice je považována za nejobecnější a nejuznávanější, definuje umělou inteligenci jako vědu, která se zabývá tím, jak přinutit stroje, aby vykazovaly takové chování , jaké by v případě člověka vyžadovalo užití inteligence. Minského definice vychází Turingova imitačního testu: „Umělá inteligence je věda o vytváření strojů nebo systémů, které budou při řešení určitého úkolu užívat takového postupu, který – kdyby ho dělal člověk – bychom považovali za projev jeho inteligence" (Minsky, 1967,s.21). Definice je klasickým příkladem antropocentrismu v oblasti UI. Jako absolutní hodnota při porovnávání inteligentních systémů je brán člověk, stroje jsou s ním poměřovány, stejně jako u Turingova testu. S tím jsme se setkali v oblasti měření IQ u lidí. Hodí se k vyjádření míry výkonnosti a efektivity. Pokud je ale pozorovatel (člověk) také mírou inteligence, nastává problém, jestliže stroj přesáhne lidské schopnosti. Ztrácíme pak možnost kvalitativně vyjádřit míru strojové inteligence. Jestliže nová inteligence vykazuje větší možnosti a schopnosti než člověk, těžko lze lidskými vyjadřovacími prostředky novou kvalitu zachytit. Minského definice se také příliš nevyjadřuje o mechanismech či struktuře inteligentních systémů (předmětu zkoumání).

Jiným pokusem o vymezení oboru je Kotkova definice. „Umělá inteligence je vlastnost člověkem uměle vytvořených systémů, vyznačujících se schopností rozpoznávat předměty, jevy a situace, analyzovat vztahy mezi nimi a tak vytvářet **vnitřní modely světa,** ve kterých tyto systémy existují, a na tomto základě pak přijímat účelná rozhodnutí, za pomoci schopnosti předvídat důsledky těchto rozhodnutí a objevovat nové zákonitosti mezi různými modely nebo jejich skupinami" (Kotek a kol., 1983,s. 42). Oproti před-

chozí, nacházíme v Kotkově definici zdůrazněnou potřebu reprezentace, nutnou vytvoření obrazu okolí a následnou tvorbu a práci s modely založenými na reprezentacích. Autor se však spíše snaží definovat cíl oboru (umělý inteligentní systém) než vědní obor UI. Také lze polemizovat s tvrzením, že inteligence je vlastnost systému. Jde o schopnost. Právě vymezením předmětu jako zkoumání tvorby a aplikace reprezentací světa při jeho interakci s okolím, se autor hlásí k proudu reprezentacionistů. V oblasti umělé inteligence ale můžeme nalézt i zastánce prezentacionistů. Například v pracích Brookse či Gibsona se objevují názory, že inteligentní systém nepotřebuje pro svou činnost vytvářet vnitřní reprezentace okolního prostředí. Více informací o této problematice je uvedeno v kapitolách o mentálních reprezentacích a agentovém přístupu.

Tradiční paradigma v umělé inteligenci se v literatuře označuje různými přívlastky, jako logicko-symbolické, symbolicko-reprezentační, algoritmické, komputacionistické. Tyto přívlastky dostatečně charakterizují tradiční umělou inteligenci a odlišují ji od novějších směrů, které mimo jiné vedou k hlubší otázce po relevanci použité architektury počítače (či obecně technického systému) vhodné k realizaci kognitivních procesů.

Zmíněná otázka se stala aktuální hlavně v posledních dvaceti letech, kdy se objevily alternativní paradigmata v umělé inteligenci: konekcionismus (neuronové sítě) a distribuovaná umělá inteligence (multiagentní systémy) (Havel, 2001). Bližší rozbor relevance a limit architektur je proveden v následujících kapitolách.

Většina konkrétních aplikací umělé inteligence má jeden podstatný znak. Tím je odklon od tvorby obecného univerzálního systému způsobený technickými a znalostními obtížemi při napodobování. V historii umělé inteligence se z tohoto důvodu experimenty následně zaměřovaly vždy na některou specifickou (někdy velmi specifickou) kognitivní schopnost. V současnosti se ale znovu objevuje potřeba, aby se místo usilování po dokonalosti ve specifických úlohách zaměřila integrovaná umělá inteligence na dílčí propojení metod do jednoho celku. Samotný pojem inteligence v sobě totiž zahrnuje právě všestrannost jako podstatnou charakteristiku (Havel, 2001). O integrované umělé inteligenci lze dnes hovořit jen v souvislosti s konstrukcí inteligentních robotů.. Na robota můžeme pohlížet jako na prostorově situovaný a vtělený *(embodied)* systém, který je v kontaktu s reálným, nikoliv virtuálním prostředím (Havel, 2001). Nutno podotknout, že robotika má kromě aplikačního také badatelský význam i svá filosofická východiska. Vtělená kognitivní věda je probírána v kapitolách věnovaných přístupu založeném na agentech.

2.4.1 Metody UI

V oblasti metod umělé inteligence se setkáváme s pojmem silný a slabý hned dvakrát. Jsou to zaprvé - silná a slabá umělá inteligence a za druhé - silné a slabé metody umělé inteligence. I když se všechny čtyři zmíněné pojmy objevují ve spojitosti s metodami UI, jejich přesné vymezení je posouvá do poněkud jiných oblastí.
Hovoříme-li o silné a slabé umělé inteligenci, ocitneme se rázem v rovině nejvyšší obecnosti teorie. Základní otázka, která rozděluje silnou a slabou inteligenci je: „Co potřebujeme k napodobení inteligence?".
Silná umělá inteligence vychází z přesvědčení, že myšlení je jen a pouze zpracování informace, které může být úspěšně duplikováno v systému se správně realizovanou funkcionální sítí (tj. v systému se správnou komputační strukturou). Zpracování informací je algoritmizovatelné, a tudíž zpracovatelné univerzálním Turingovým strojem (viz Turingův stroj). Silná umělá inteligence je tedy jakousi derivací funkcionalistické teorie mysli, která se soustřeďuje na možnost (prostřednictvím vhodného programu) realizovat mentální stavy v digitálním počítači (Pícha, 2001). Extrémní zastánci přístupu hovoří o možnosti simulovat lidskou mysl pomocí pivních kelímků propojených provázky. Tvrzení silné UI jsou výrazem víry ve funkcionalismus a možnosti převedení okolního světa do výpočtů.

Jedním z hlavních směrů, který ovlivnil kognitivistické paradigma, je **funkcionalismus**. Jeho hlavní důraz není kladen na architekturu a strukturu systému, ale na to, aby systém byl schopen vykonávat požadované operace (Pfeifer&Scheier, 2001). Základní rozdíl mezi strukturalismem a funkcionalismem je, že první zmíněný zkoumá člověka jako pasivní bytost, která je vystavena působení prostředí, přijímaného skrz smysly. Funkcionalismus naopak vnímá jedince jako aktivního účastníka vnímání a konání (Sternberg, 2001).

Searle klade **slabou umělou inteligenci** do opozice k silné. Podle slabé umělé inteligence spočívá základní hodnota počítačů při studiu mysli v tom, že představují velmi užitečný, ale vždy pouze nástroj v rukou člověka. Počítače dokáží řešit jen specifické problémy. Definice slabé UI je z velké části postavena na popření možností, které nabízí silná UI.

Diskusí kolem silné a slabé umělé inteligence je nepřeberné množství. Vymezení tématu se objevuje v mnoha modifikovaných variantách. Bohužel odpověď na otázku, jaké jsou nutné podmínky pro architekturu systému schopného napodobit inteligentní chování (výše zmíněné „Co potřebujeme k napodobení inteligence?"), je předčasně položena. Na empirická ověření, která odpoví na položenou otázku, zatím nejsme dostatečně vybaveni. Legitimizuje ji oblast filosofie mysli, ze které se rekrutují autoři vyjadřující se k danému tématu.

Snadnější je odpověď na otázku ze začátku kapitoly, zda je silná a slabá umělá inteligence metodou UI. Výše zmíněná fakta ukazují, že slovo metoda by v tomto případě bylo zavádějící pro svou konkrétnost.

Přesuňme se nyní k problematice silných a slabých metod UI. Newell a Simon definovali dvě metody, kterými se dá simulovat zpracování informace. **Slabá metoda** říká, že funkci inteligentního systému stačí pouze pár výkonných mechanismů, které zpracovávají příchozí informace. Není potřeba tvořit interní reprezentace o těchto procesech. Systémy mají velmi slabé požadavky na předchozí znalost při řešení úkolu, důležitější je vždy samotný algoritmus. Tento způsob vycházel ze Simonových názorů, že důležitou komponentou při tvorbě inteligentních systémů je právě funkce operátorů a že interní reprezentace (paměťový sklad) nesouvisí s nárůstem inteligence. Silný přístup naopak vychází z důrazu na znalost a na její sílu. Kvalitní forma interní reprezentace je zde nutnou podmínkou (Hogan, 1998). **Silná metoda** v sobě obsahuje slabou metodu a navíc také znalostní bázi, která v sobě obsahuje reprezentace příchozích informací. Reprezentace jsou ukládány tak, aby byly postiženy vztahy mezi informacemi podle pravidel předem naprogramovaných mechanismů (Luger, 1994).

I v tomto případě se nám nedaří přesně naplnit pojem metoda umělé inteligence. Oproti příliš obecnému předchozímu případu, je zde problém opačný. Jedná se o specifické použití pro určitou oblast UI. Přesněji řečeno oblast produkčních, či expertních systémů, které jsou zmíněné v pozdějších kapitolách.

2.5 Informace

Pokud budeme chtít hovořit o architekturách, pomocí kterých se vědci snaží napodobovat lidské myšlení, je nutné předem zmínit klíčový termín informace. V kognitivních vědách se hovoří o zpracování informace, která má již konkrétní formu nebo alespoň vymezený rámec, v němž se pohybuje (stává se kódem). Podívejme se ale na pojem informace v obecné rovině.

V roce 1948 publikoval Claude Shannon společně s matematikem Warrenem Weaverem článek „A mathematical theory of communication". Někteří historikové vědy tuto práci nazývají „Magna charta informačního věku". Ukazuje se v ní, že k exaktnímu zkoumání informace je potřeba abstrahovat od její sémantické stránky a omezit se na stránku syntaktickou, která je statistickými prostředky snadněji popsatelná. Informace spočívá v odstranění neurčitosti. Při vyjádření míry odstraněné neurčitosti dospěl Shannon k formálně stejnému vztahu, který koncem 19. století odvodil Ludwig Boltzmann pro entropii. Následné konsekvence vedou ke zjištění, že zdánlivě nehmotná informace je pevně vázána na fyzikální svět hmoty a energie a že každý přenos či záznam informace vyžaduje disipaci jisté energie, a tedy vzrůst termodynamické entropie (Vysoký, 2004).

Podle toho, ve kterém vědním oboru nebo ve které oblasti lidské činnosti se používá, jsou aplikovány specifické přístupy ke zkoumání informace a jsou k dispozici různé způsoby jejího definování. Autor publikace z oblasti teorie informace Norbert Wiener se vyjadřuje ke zmíněnému pojmu poněkud metaforicky: „Mechanický mozek neprodukuje myšlení „jako játra žluč", jak si mysleli první materialisté, stejně jako jej neprodukuje ve formě energie jako svaly aktivitu. Informace je informace, ne hmota nebo energie (Wiener, 1947,s. 89). V samotné práci pak (stejně jako Shannon) hovoří o informaci jako o opaku entropie (Pstružina, 1998). Přesnější definice informace je možno nalézt v mnoha vědních oborech. Vyjadřují názorovou pestrost a také specifická pole využití pro dané obory.

Filozofické pojetí informace:

- Vlastnost hmotné reality být uspořádán a její schopnost uspořádávat (forma exis tence hmoty vedle prostoru, času a pohybu).
- Význam přiřazený obrazům, údajům a z nich utvořeným lidským celkům. Informace představuje míru uspořádanosti systémů na rozdíl od entropie, tj. míry neuspořádanosti.

Komunikační pojetí informace

- Objektivní obsah komunikace mezi souvisejícími hmotnými objekty, projevující se změnou stavu těchto objektů.

Kybernetické pojetí informace

- Název pro obsah toho, co se vymění s vnějším světem, když se mu přizpůsobu jeme a působíme na něj svým přizpůsobováním. Proces přijímání a využívání informace je procesem našeho přizpůsobování k nahodilostem vnějšího prostředí a aktivního života v tomto prostředí.
- Proces, kdy určitý systém předává jinému systému pomocí signálů zprávu, která nějakým způsobem mění stav přijímacího systému.

Matematický přístup k informaci

- Energetická veličina, jejíž hodnota je úměrná zmenšení entropie systému.
- Poznatek, který omezuje nebo odstraňuje nejistotu týkající se výskytu určitého jevu z dané množiny možných jevů.
- Obsah zprávy, který je definován jako záporný dvojkový logaritmus její pravděpodobnosti.

(Kučerová, 2002)

V nejobecnějším slova smyslu se informací chápe údaj o reálném prostředí, o jeho stavu a procesech v něm probíhajících. Informace snižuje nebo odstraňuje neurčitost systému (např. příjemce informace); množství informace je dáno rozdílem mezi stavem neurčitosti systému (entropie), kterou měl systém před přijetím informace a stavem neurčitosti, která se přijetím informace odstranila. V tomto smyslu může být informace považována jak za vlastnost organizované hmoty vyjadřující její hloubkovou strukturu (varietu), tak za produkt poznání fixovaný ve znakové podobě v informačních nosičích. V informační vědě se informací rozumí především sdělení, komunikovatelný poznatek, který má význam pro příjemce nebo údaj usnadňující volbu mezi alternativními rozhodovacími možnostmi. Významné pro informační vědu je také pojetí informace jako psychofyziologického jevu a procesu, tedy jako součásti lidského vědomí. V exaktní vědě se např. za informaci považuje sdělení, které vyhovuje přísným kritériím logiky či příslušné vědy. V oblasti výpočetní techniky se za informaci považuje kvantitativní vyjádření obsahu zprávy. Za jednotku informace se ve výpočetní technice považuje rozhodnutí mezi dvěma alternativami (0, 1) a vyjadřuje se jednotkou nazvanou bit (Jonák, 2000).

2.6 Nejmenší jednotky

Jednou z priorit, které si všimnete při studiu simulace inteligence, je snaha nalézt co nejmenší jednotku, která by byla tak univerzální, že by dokázala nést či zpracovávat libovolný typ kódu či informace (základní stavební prvky architektur). U dvou základních přístupů k simulaci inteligence (komputace a konekcionismus) můžeme základní jednotky dobře identifikovat. U komputačního přístupu jsou jako základní reprezentační jednotky používány symboly (nejlépe ve formě 1 a 0) a jako nejjednoduššího a univerzálního mechanismu pro zpracování Turingova stroje, který je schopen zpracovat libovolnou vypočitatelnou úlohu. Na druhé straně je nejmenší výpočetní jednotkou neuron a nejmenší jednotkou reprezentace váha propojení.
Tím je dána architektura. Můžeme se ale zeptat, co mají nejmenší jednotky zastupovat? Jaký je nejmenší atom prostředí, z jehož kompozice dokážeme vytvořit obraz (reprezentaci) prostředí? V rozličných přístupech se setkáváme právě s různým přiřazováním části prostředí těmto atomům. Jako atomy jsou v teoriích používány čísla, znaky, slova, věty, části obrazů, čáry, pravidla, objekty, geometrická primitiva apod., které jsou následně v reprezentované formě modifikovány, zpracovávány, transformovány.
Problém přichází v okamžiku, kdy chceme po systému, aby dokázal zpracovat informaci z prostředí, která je menší než nejmenší „reprezentační" atom daného systému (obsahuje prvky, které nedokáže reprezentační systém zachytit). Dostáváme se do fáze, která by se dala v logice nazvat snaha o rozdělení axiomu na ještě menší celky, což nelze (viz Gödel). Z hlediska hodnocení se přibližujeme k otázkám obecné rozlišovací schopnosti systému. Klíčovou úlohu sehrává, zda je systém schopen zpracovat libovolnou úlohu – systém je obecný nebo pouze úlohu, pro kterou má vhodný formální (či neformální) aparát - specifický systém.
Výhodou obecného systému *(general domain systém nebo general purpose system)* je právě schopnost akomodace na prostředí a tvorbu arbitrárních nejmenších jednotek, které jsou adekvátní typu úlohy, před níž je postaven. Jeho fungování tedy nevychází z axiomů a zákonů, které jsou neměnné, ale jeho aparát (zde je již obtížné říci, zda formální aparát) dokáže své „axiomy" rozšiřovat (sporné je, zda modifikovat). V případě modifikace formálního aparátu bychom se dostali na hranice schopnosti formálních logických systémů, jelikož by došlo k jeho zhroucení a neplatnosti jím vyvozovaných závěrů. Omezení ukazují formální logický systém jako nevhodný pro tvorbu systémů, které dokáží pracovat s libovolnou úlohou. Což nás může vést k závěru,

že obecný inteligentní systém, nemůže být postaven pouze na základech logiky.

V předchozím odstavci jsme ale spojili dohromady dvě témata. Jedním je reprezentační schopnost inteligentního systému (oblast převodu prostředí na adekvátní formu reprezentace) a druhým je výpočetní (obecněji „informace zpracovávající") schopnost, tedy oblast práce s reprezentacemi. Oblasti se navzájem překrývají, a proto o nich bylo hovořeno jako o celku. Pokud se vrátíme k obecným systémům, schopným řešit libovolný typ úlohy a také mající schopnost převádět informace z prostředí na adekvátní formu reprezentace (sloužící při řešení úlohy), budeme těžko hledat příklady v oblasti simulace. Jediným důkazem, že existuje takový obecný systém je člověk. I když není v jeho rozlišovacích schopnostech vidět atom (zde již ve smyslu jednotka hmoty), přesto to dokázal.

Nejasnosti z opačného konce souvisí s procesy generalizace, abstrakce či kategorizace. Jde o snahu vytvořit systém, který pracuje s již zmíněnými primitivy - atomy, se kterými jsou prováděny operace umožňující převod původních informací v novou (výše zmíněnými procesy). Je předpokládáno, že v nové úrovni, tedy ve výsledků operací, vytvoříme informace, které budou kvalitativně nové, což je obtížně definovatelné a také sémantické, což je většinou jen vysněným přáním v komputačním přístupu. Bohužel jsou atomy prostředí reprezentovány čistě syntakticky, a proto v nové rovině vzniká pouze takový obsah, který nabývá svého významu interpretací tvůrce. Jedná se o snahu získat obsah i pokud pracujeme na nižší rovině syntakticky, přičemž obsah apriorně předpokládáme (viz Ukotvení symbolů). Extremním vyjádřením legitimizujícím syntaktický přístup je Haugelandovo tvrzení, že pokud se postaráme o syntax, sémantika se o sebe postará sama (Crane, 2002).

V této kapitole jsme zmínili omezení, která vymezují hranice simulovatelného zpracování informací. Možnost vytvoření jednotné teorie je limitována současnou úrovní poznání v oblasti kognitivních věd.

V aplikovaných oblastech modelování (pro konkrétní existující modely) se používá jako kritérium hodnocení schopnosti modelu napodobit svůj originál *analýza citlivosti*. Což můžeme vidět jako abstraktní hodnocení stávajícího modelu a jeho vztahu k jevům, které se snaží modelovat. Jedná se o vytvoření relativního vztažného systému, ve kterém je fixním referenčním kritériem originál modelovaného jevu. Analýza citlivosti nemusí být kvantifikována ve smyslu vytvoření číselných škál reprezentujících míru podobnosti modelu s originálem. Častěji se využívá poměrových škál, vyjádřených pomocí nezbytných, dostatečných a nepodstatných podmínek, za kterých model dokáže napodobovat dané jevy. Neexistuje přesný algoritmický

popis, pomocí něhož můžeme analýzu citlivosti provést, jelikož k posouzení citlivosti je třeba nutné pochopení originální situace, znamenající nejen dekompozici na jednotlivé elementy, ale i uchování vztahů mezi elementy a kontextová znalost zkoumaného jevu vzhledem k prostředí (Sternberg, 1999). Díky tomu, že současná umělá inteligence je velmi vzdálená termínům význam či pochopení, je při tomto druhu analýzy nutný lidský (inteligentní) pozorovatel, který je schopen takové míry abstrakce, pochopení a vhledu, že dokáže posoudit kategorie nutnosti, dostatečnosti a nepodstatnosti. Nejznámějším zástupcem analýzy citlivosti pro umělé inteligentní systémy je Turingův test.

V oblasti neuronových sítí nám při analýze citlivosti může pomoci mechanismus zpětného šíření, tedy že výstupní vrstva vysílá informace o výsledku do skrytých a vstupních vrstev *(backpropagation)*, pomocí čehož je systém schopný získat informaci o míře své efektivity. To je ale příliš málo na to, aby taková síť sama dokázala z informace odvodit míru své komplexní citlivosti, která vyžaduje znalost prostředí, modelu, výstupů modelu a také limity, ve kterých se model během své činnosti pohybuje.

2.7 Paralelní versus sériové zpracování

Jinou podstatnou úlohou kognitivních věd je odhalit, jaké jsou vnitřní limity, které omezují lidský mozek při procesu myšlení (Sternberg, 1999). Při volbě způsobu nazírání této problematiky můžeme použít jako kritérium, zda mozek v průběhu zpracování informace pracuje ve všech krocích paralelně, či nikoliv.

Například Saul Sternberg se pomocí testu reakčního času snaží dokázat, že krátkodobá paměť zpracovává informace sériově (při administraci testů se zvyšujícím se počtem položek vzrůstá reakční čas v závislosti na jejich množství), což potvrzuje myšlenku o úžině vědomí (zde spíše úžinu paměti). Druhým příkladem (který je více kulturně podmíněn a vypovídá spíše o anticipaci ve vnímání) je Word superiority test. Sternberg v něm prováděl pokusy s krátkou prezentací 4 slabičných shluků (např. RACK, KARC, XXAX,). Pokud tyto stimuly tvořily slova, pokusné osoby je lépe odhadovaly. (Gazzaniga&Mangun, 1998).

Již jednoduchá úloha, jakou je násobení vícemístných čísel v hlavě, nám ukáže, že systém je schopen takovou úlohu provést, ale musí při ní využít svou krátkodobou paměť, ve které jsou částečné mezivýpočty ukládány, ale také musí být odstraněny pro další výpočty. Tato omezení se nazývají kapacitní (Sternberg, 1999). V systému pracujícím masivně paralelně, tvoří operační paměť pomyslné zúžení, ve kterém je nutné

převést paralelní procesy do sériové, sekvenční podoby. Jestliže zaujímáme zmíněné stanovisko (vjemy z okolí jsou po transdukci smyslovými orgány uloženy do krátkodobé paměti) docházíme ke zjištění, že vzniká značně neefektivní způsob zpracování, jelikož následné operace s obsahy krátkodobé paměti probíhají paralelně (musely by být opět převedeny do paralelní formy) a pak znovu do sériové podoby pro práci efektorů. Přikláníme-li se ale k Tverského modelu paměti, proběhne převod pouze jednou. Paralelní informace jsou uloženy rovnou do dlouhodobé paměti a z nich si jediným převodem vybírá krátkodobá paměť ty části informace, které budou použity pro operace v sériové podobě, jejichž výsledky lze posléze přímo směřovat na efektory. Jedná se ovšem o přílišné zobecnění na všechny senzory i efektory, což nás opět posouvá na hranici spekulace. Řešení předpokládá vyjasnění úlohy jazyka (zde pravděpodobně mentálního) v procesu zpracování, tedy zda-li je možné jej zpracovávat paralelně.

Mnoho vědců zabývajících se UI se domnívá, že hlavními příčinami je společně s již zmíněnými také nedostatečná robustnost a neschopnost provádět generalizaci v reálném čase. Dalšími nedostatky jsou problémy s vytvářením rámců při zpracovávání reality, problém s ukotvením symbolů, a také nedostatky v oblasti vtělenosti a situovanosti umělého systému (Pfeifer&Scheier, 2001). Jaký podíl na tom mají samotné architektury současných systémů? A jaký podíl způsob jejich využívání? Následný rozbor způsobů simulace inteligence se pokusí na některé otázky odpovědět.

3 KLASICKÝ PŘÍSTUP

3.1. Předpoklady

3.1.1 Logika

Z lidského hlediska jsou logika a logické systémy nejčastěji zastřešovány pojmem racionalita či racionální chování. Ovšem vztah těchto dvou pojmů není dán tak přesně a neexistuje mezi ním hierarchické uspořádání (Sternberg, 1999). Na formální logice je založeno mnoho oborů umělé inteligence, např. řešení problémů, automatické dokazování, produkční systémy. Historie vývoje logiky se datuje již od starého Řecka a probíhá do dnešních dní. Současně vznikají stále nové systémy, které se snaží zdokonalit nevýhody předchozích. Některá omezení logiky jsou ale takového charakteru, že popírají logický systém jako univerzální princip o vyvozovaní a ověřování výroků (viz Nejmenší jednotky, Gödel). Pro orientaci uvádím klíčové postavy a jejich přínos oboru v následujícím přehledu:

tradiční logika

- Aristoteles (5. st. před. n.l.) je zakladatelem logiky jakožto nástroje (ř. organon) poznání a uvažování; aristotelský sylogismus (fragment predikátové logiky)
- stoikové – výroková logika
- středověk – scholastikové rozvíjeli aristotelský sylogismus (logika byla součástí tzv. trivia)

moderní logika

- předchůdce Gottfried Wilhelm Leibnitz (pokud vznikne mezi filosofy spor, tak si své argumenty zapíší a „spočítají", který argument je korektní a jehož závěr je platný); též Bernard Bolzano (první definice vyplývání);
- bezprostřední předchůdci: George Boole (booleova algebra), či Charles Dodgeson (Lewis Caroll – Alenka v říši divů), Charles Sanders Peirce;
- konec 19. st., zakladatelé zejména Gottlob Frege (logicismus: snaha doložit, že všechna matematika je odvozena z logiky), Bertrand Russell (rovněž logicismus, dále významné uplatňování ve filosofii); vybudování predikátové logiky;
- predikátová logika uplatňována při zkoumání základů matematiky: Alonzo Church, Kurt Gödel, aj.; matematická logika se začíná vyvíjet jiným směrem, než filosofická logika
- moderní logika uplatněna jako základní nástroj tzv. logického novopozitivismu (Vídeňský kruh, Rudolf Carnap, ale i Ludwig Wittgenstein);

- Gödelovy objevy vedou k obratu pozornosti na výzkum algoritmů a rekurzívních funkcí Alonzo Churchž Alan Turing, Turingův stroj, Church-Turingova teze;
- Alfred Tarski definuje moderním způsobem vyplývání a korespondenční teorii pravdy, vybudoval teorii modelů;
- modální logika (s operátory „je nutné", „je možné"), C.I. Lewis, Ruth Barcan Marcus, sémantická reforma (60.léta): Saul Kripke (bohaté využití ve filosofii)
- vícehodnotové logiky postupně vedou k současné *fuzzy logice*
- v průběhu 60. let vzniká intenzionální logika - Richard Montague
- Richard Montague: teze, že není žádný rozdíl mezi umělými (tj. formálními) jazyky logiky a jazyky přirozenými (jakými jsou čeština, angličtina apod.)
- v současnosti jsou vyvíjeny hyperintenzionální logiky (Pavel Tichý)
- v současnosti dochází i k pokusu změnit paradigma logiky (logika na základě teorie her - Jaakko Hintikka, dynamická logika, nonmonotonní logiky)

Jeden z prvních výpočetních strojů (Analytický stroj) je založen na principu, který umožňuje operovat se symboly, přičemž nepoužívá pouze zákony aritmetiky a logiky. Již v té době bylo rozpoznáno, že pomocí logiky nelze popsat většinu způsobů běžného přemýšlení. Tedy že nestačí věci roztřídit do pouhých kategorií podle toho, zda mají nebo nemají určitou vlastnost. Věci v reálném světě mají vlastnosti ve větší či menší míře, což nelze redukovat na ano/ne .

Jednou z disciplín, které vývoj studia logiky ovlivnil je matematická logika. V roce 1854 přišel Angličan George S. Boole s takovým modelem matematické logiky, e kterém vystačil jen se třemi základními operátory (AND, OR a NOT) a s jejich pomocí dokázal z jednotlivých výroků sestavovat složitější formule stejným způsobem, jakým se v matematice (konkrétně v algebře) sestavují matematické vzorečky. Svou logiku mohl formálně vybudovat jako algebru, které se dodnes říká Booleova algebra. Boole jistě netušil, že se jeho algebra stane základním teoretickým aparátem pro modelování kombinačních obvodů číslicových počítačů. Netušil také, že technikům se budou nejlépe dařit takové konstrukční prvky, které budou mít jen dva možné stavy, a kterým bude odpovídat Booleova algebra, která má právě jen dva prvky (zatímco George Boole ji navrhl pro obecný počet prvků, nejméně však jako dvouprvkovou). Tím, kdo poprvé ukázal na souvislost dvouprvkové Booleovy algebry s číslicovými obvody, byl v roce 1937 Claude Shannon. Mezitím se musela zrodit ještě jedna velmi důležitá myšlenka, která si vynutila používání právě dvouprvkové Booleovy algebry.

Posuňme se ale dále k propoziční logice. Problém zde není jen ve vztahu mezi

propozicemi, ale také v propozici samotné. Pokud si vezmeme výrok „Každý pes má vlastníka", jedná se o vztah, kdy jeden je vlastněn druhým, ale ne naopak. Propoziční logika bere tento výrok jako ověřitelné tvrzení, ale nedokáže nic říci o jeho interních vztazích. Problém je rozlišován až v oblasti predikátového kalkulu, systémem formalizace prezentovaný Gottlobem Fregem (Hogan, 1998).

Jím navrhovaný systém měl umožňovat odvodit většinu pravidel aritmetiky a matematiky z jedné množiny axiómů. Frege se domníval, že se mu podařilo takový systém vytvořit, ale ukázalo se, že neměl pravdu. Bertrand Russell odhalil, že jeden z Fregeho axiómů může vést k teorému, že pokud je množina prvkem sebe sama, není prvkem sebe sama, ve slovní formulaci jako paradox vesnického holiče, který holí všechny, co neholí sami sebe - holí se i holič sám? Tento problém se obecně nazývá Russellův paradox (Hogan, 1998) a má úzkou souvislost s paradoxem dvojitého lháře (Jourdainův paradox): na papírové kartičce jsou napsány následující věty (každá je na jedné straně): „Nápis na druhé straně je pravdivý." „Nápis na druhé straně je nepravdivý". Nejen historicky zajímavá je i souvislost s Epimenidovým (Eubulidovým) paradoxem lháře (nejednodušší varianta: „Já lžu").

Příkladem může být kategorizace slov (Grellingův paradox). Některá přídavná jména popisují vlastnosti, které samy nemají. Říkáme o nich že nejsou autodeskriptivní (sebepopisné). Například slovo jednoslabičný, které není jednoslabičné. Vytvářejí tak druhou kategorii, slova heterodeskriptivní (nepopisující sebe sama). Tímto způsobem můžeme rozdělit všechna přídavná jména do zmíněných kategorií. Dostaneme se až k přídavným jménům autodeskriptivní a heterodeskriptivní. Pokud je chceme rozřadit, nastává paradox. Řekneme-li že heterodeskriptivní patří do skupiny heterodeskriptivních slov, dopouštíme se omylu, jelikož se tím pádem stává autodeskriptivní. Pokud ale prohlásíme že je autodeskriptivní, opět to není pravda (Hogan, 1998).

Následují další pokusy definovat základní axiómy tak, aby byly kompletní a bezesporné. Russell a Whitehead se vyhnuli sebereferenci (autodeskripci) tím, že ji zakázali. V návaznosti na pokusy o vyjasnění limit logických systémů přichází Kurt Gödel se svým důkazem. Jedná se o tvrzení, že neexistuje takový konsistentní formální systém, který by dokázal posoudit všechny své výroky. Tedy, že existuje vždy takový výrok, který je systémem axiómů a pravidel nedokazatelný (Hogan, 1998).

3.1.2 Gödel

Název této kapitoly odkazuje na období, které se vyznačovalo snahou ověření logick-

ého systému, jako univerzálně platného nástroje pro ověřitelnost a dokazatelnost tvrzení. Kurt Gödel je autorem výroku, který odstartoval polemiky kolem úplnosti formálního aparátu. Dostáváme se ke Gödelovu výroku, že pokud je systém (konkrétně systém aritmetiky definovaný Russellem a Whiteheadem) bezrozporný, nemůže dokázat vlastní bezrozpornost.

Příklad Gödelovy sentence:

Pokusíme se roztřídit všechny pravdivé sentence do dvou skupin - 1. pravdivé, nedokazatelné, 2. pravdivé, dokazatelné. Gödel sestrojil sentenci , která tvrdí, že patří do skupiny 2 a zní „nejsem v systému dokazatelná". Pokud je sentence nepravdivá, je v systému dokazatelná. Pak ale nemůže být dokazatelná, protože není pravdivá. Musí tedy být pravdivá, ale je nedokazatelná. (Smullyan, 2003).

Podívejme se, co říká o Gödelově větě Barrow: „Žádná z možných dedukcí, k nimž lze dospět na základě těchto axiómů užitím povolených vyvozovacích pravidel, nemůže obsahovat více informace, než kolik jí bylo obsaženo v axiómech (viz Monotonie). V tom tkví v podstatě příčina slavných omezení moci logické dedukce, jak to vyjadřuje Gödelova věta o neúplnosti." (Barrow, 1996,s. 112).

Peregrin na to reaguje následovně:„Obávám se, že charakterizovat Gödelův výsledek takto, je jako charakterizovat nehodovost na silnicích slovy „Příčinou tak mnoha havárií je to, že auta neumějí létat." Faktem totiž bezesporu je, že kdyby mohly důsledky axiómů obsahovat více informací než axiómy samy, nemusela by Gödelova věta platit; stejně tak jako kdyby auta uměla létat, nemuselo by docházet k tolika haváriím. Obě ta tvrzení jsou ale naprosto nezajímavá - chtít po autech, aby létala, či chtít po důsledcích axiómů, aby obsahovaly více informací než axiómy samy, nedává příliš rozumný smysl. To není žádná zvláštní vada aritmetiky a už vůbec ne doklad její rozpornosti" (Peregrin, 1997).

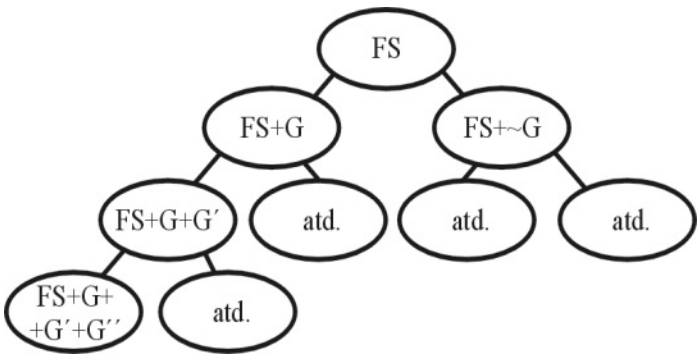
Bezrozpornost aritmetiky samozřejmě lze dokázat (není pravda, že se to nemáme šanci dozvědět), ale musíme to provést „my", nelze to uskutečnit v rámci aritmetiky samé, aritmeticky. Zde se blížíme interpretaci, kterou nabízí např. prof. Peregrin ve své knize Filosofie a jazyk, když zdůrazňuje právě onu potřebu vyskočit „nad systém" (Peregrin, 2003)., díky čemuž můžeme pravdivosti gödelovsky nerozhodnutelných / nedokazatelných výroků dokazovat úplně hravě.

Představme si, že existují systémy, v rámci nichž lze dokázat jejich vlastní bezrozpornost. Musíme ale takovému důkazu věřit? Lze dokázat, že ve vnitřně rozporném systému lze naopak dokázat jakýkoliv výrok - tedy i vlastní bezrozpornost. Smullyan přímo přirovnává k tomu, že je také pošetilé někomu věřit jen proto, že tvrdí, že vždy mluví

pravdu. A jde dokonce ještě dále. Ukazuje, že některé systémy nemohou dokázat vlastní bezrozpornost „právě proto", že jsou skutečně bezrozporné (Peregrin, 1997).

Pomalu se dostáváme do oblasti, kterou se zabýval Alfred Tarski a která posouvá problematiku do oblasti sémantiky (otázka „pochopení" související s možností vyskočit „nad systém").Ve své nejvýznamnější práci navázal Tarski na Kurta Gödela. Stejně jako Gödel totiž dokázal, že logické systémy jsou sémanticky neúplné. Tím se zhroutila snaha význačných matematiků první poloviny minulého století (B. Russell, A. Whitehead), kteří se snažili definovat matematiku jakožto zcela konzistentní systém založený na zákonech logiky. Zatímco Russell a Whitehead usilovali ve svém monumentálním díle Principia Mathematica vystavět matematiku na základech symbolické logiky tím, že z ní odstraní všechny logické paradoxy (což se jim samozřejmě nepodařilo), Hilbert šel ještě dále. Chtěl dokázat vnitřní konzistenci axiómů aritmetiky a všech dedukcí, které z nich mohou být vyvozeny.

Ve výsledku docházíme ke zjištění, že logické a matematické systémy, které jsou natolik bohaté, že obsahují Peanovu aritmetiku, jsou nejen formálně neúplné ve smyslu, že některé z jejich pravd jsou za použití prostředků systému nedokazatelné, ale že jsou také sémanticky neúplné v tom smyslu, že některé z jejich pojmů se nedají definovat pomocí jazyka a pojmů systému. Vždy je lze definovat pomocí většího systému, ale jen za cenu vytvoření dalších nedefinovatelných pojmů v rámci tohoto většího systému. Nakonec to tedy znamená, že neexistuje formální systém, v němž by mohla být rozhodnuta pravdivost všech matematických tvrzení nebo takový v němž by mohly být definovány všechny matematické pojmy.



Obr. 6 *Rozšíření formálního systému*

Kurt Gödel dokázal, že elementární matematika nutně obsahuje nedokazatelná tvrzení a že se částečně opírá o „axiómy víry" ve svou vlastní konzistenci. Vytvoříme-li tvrzení, jehož pravdivost či nepravdivost nelze určit, pak můžeme k množině axiómů definujících systém přidat libovolně buď toto tvrzení, anebo jeho negaci. Obě volby ovšem vytvoří větší logický systém, jenž nutně obsahuje nová nedokazatelná tvrzení.
Pokud Gödel svým důkazem potvrdil domněnku, že nejsou všechny výroky ve formálním systému dokazatelné, další důležitou otázkou byla, jestli existuje univerzální metoda na zjištění, zdali je daný teorém ověřitelný automatickým procesem ověřování (Hogan, 1998).
Následky Gödelova výroku se objevují i v oblasti tvorby inteligence. J.R. Lucas v roce 1960 upozornil na to, že Gödelovu větu o neúplnosti aritmetiky je možné použít k vyvrácení mechanistické teze, tedy názoru, že lidskou mysl lze simulovat strojem (Lucas, 1961). Lucasův článek obsahuje množství úvah a námětů, které stojí za uvedení.
Říkáme, že vědomá bytost něco ví, neříkáme tím jen, že to ví, ale též, že, ví, že to ví, a že ví, že ví, že to ví atd. Vědomá bytost může zacházet s Gödelovými otázkami tak, jak stroj nemůže, protože vědomá bytost může uvažovat sebe a své projevy, aniž by se lišila od zdroje těchto projevů (Lucas, 1961). Lucasem zdůrazněná potřeba vědomí odkazuje na nutnost uchopení a pochopení obsahu informací, které inteligentní systém zpracovává. Kognitivní teorie, která vidí základ pochopení funkce kognitivního aparátu ve zpracování informací, by měla začít klást větší důraz na „pracování s informací".

3.1.3 Formální a mentální logika

V klasické filosofii a psychologii je silnou tradicí považovat formální logiku nejen za normativní, ale i za deskriptivní teorii lidského usuzování. Naproti tomu například teorie mentálních modelů zdůrazňuje mentální logiku, která se od formální logiky liší tím, že je přirozenou vlastností člověka. Většina kognitivistů předpokládá její vrozený základ (Sedláková, 2004). Důkaz rozdílnosti mentální a formální logiky můžeme objevit již v samotném vymezení těchto oborů.
Formální logika se snaží pomocí co nejmenšího počtu operátorů a výrazového aparátu, vytvořit bezesporný popis světa a také posoudit výroky o tomto světě ohodnocením pravda/nepravda (v základní formě logiky). Při její tvorbě je aplikováno pravidlo Ockhamovy břitvy, snahy vyjádřit jevy či procesy světa v nejefektivnější (nejkratší) formě. Matematická logika se pokouší minimalizovat prostředky, které

používá k vyjádření nezbytných údajů (Mařík, 1993). V případě aplikace v oblasti matematických systémů nemusí dojít k redukci a je možné vyjádřit veškeré možnosti (snad až na bezrozpornost). Pokud ale použijeme logického aparátu k vyjádření znakového či symbolického systému, můžeme při dostatečné bohatosti základního aparátu dojít k neredukcionistickému popisu v syntaktické rovině, ale v sémantické rovině (která obsahuje dvojsmyslnosti, nepřesnosti apod.) bohužel logický aparát selže. Formální logika nefunguje v rovině významů. Příčiny můžeme vidět již v použití symbolického systému, který je nedostačující pro zachycení „obrazu reálného světa" s jeho hlavními konstituentami (dynamičnost, paralelnost, komplexnost).

Mentální logika vychází z opačného konce. Je na sémantice postavena. Nejednoznačnost či protichůdnost vývodů nevede ke zhroucení celého systému, pouze k vytvoření opravných mechanismů či omezením (přesné postupy jsou dodnes neznámé), které umožní jejich korekci, zavržení, omezení nebo změnu vyvozovacího mechanismu.

Prosazování myšlenky mentální logiky se stalo součástí studia mentálních modelů. Skupina vědců okolo Johnsona-Lairda předpokládá, že lidé mají vrozený logický systém nebo vrozené předpoklady pro přirozenou logiku, která se od formální logiky liší. Mentální logiky obsahují množinu inferenčních schémat, které jsou abstraktní a sledují vystižení účelovosti v přírodě, v lidském životě i v životě společnosti. Inferencí se rozumí proces myšlení, který vede od jedné množiny propozic k jiné. Propozice jsou většinou vyjadřovány verbálně, v případě praktických inferencí mohou premisy sestávat z percipovaných nebo představovaných stavů věcí či událostí a závěr pak může být zastoupen v akci (Sedláková, 2004).

Častou otázkou v oblasti mentální logiky bývá, zda je vrozená a jaké mechanismy ji konstituují. Zastánci vrozenosti (například Chomsky) tak oponují Piagetově teorii, že pravidla mentální logiky získáváme na základě učení. Problematika je blízká tématu *common sense* (zdravého rozumu), což je soubor pravidel, který není vázán na pravidla exaktně logická, ale jedná se o abstrahovanou formu základního lidského vědění. Je zatížen značnou subjektivitou, takže v mnoha aspektech se rozchází s vědeckým přístupem zkoumání, které klade důraz právě na objektivitu a exaktnost.

Snahou o napodobení *common sense* pomocí počítačů, metodou expertního systému (viz Expertní systémy) je projekt CYC. Snaží se vytvořit encyklopedii znalostí *common sense* , do které přispívají lidé z celého světa a zapisují ve formě pravidel různé poznatky reprezentující znalostní bázi „zdravého rozumu" (Crane, 2002). Vidíme zde klasický příklad použití formální logiky jako mentální. Mentální pravidla jsou zpracov-

ávána klasickou počítačovou architekturou založenou na komputaci a logice, čímž ztrácí své původní vlastnosti. Projekt CYC je monumentální velikostí báze pravidel, ale nepřesný použitou metodou.
Místo závěrečného shrnutí je možno říci, že metaforicky lze vidět vztah *common sense* a vědeckého přístupu jako vztah mentální a formální logiky, jako vztah subjektivního a objektivního, nejednoznačného a jednoznačného, sémantického a syntaktického atp. Bohužel nám znalost druhého nestačí k přesnému „popisu" prvého. Z uvedených argumentů vyplývá, že rozdíly mezi mentální a formální logikou jsou svou povahou fundamentální.

3.1.4 Monotonie

Živou reakci v kruzích logiků vyvolala ta část Minského studie, ve které podrobil nelítostné kritice monotónnost tradičních logických systémů. Za principiální vlastnost inteligence pokládá nemonotónnost lidského uvažování: „Naše mysl je výjimečná právě tím, že se dokážeme zříct včera odvozených pravd a nahradit je těmi, které jsme odvodili dnes" (Minsky,1968,s. 40). V monotónní logice však zůstává dokázána pravda již navždy pravdou a logické kalkuly dlouho nenabízely způsob, který by na tomto cokoli změnil. Až tato kritika uvnitř logiky přivedla ke studiu různých nemonotónních formálních systémů. Monotonie formálních systému se často uvádí na příkladu:
Vycházíme z předpokladu, že každý pták létá. Z toho samozřejmě můžeme usoudit, že i jistý konkrétní pták jménem Quido umí létat. Ovšem později se dozvíme, že Quido je tučňák a tučňáci nelétají. Náš systém by se měl v tomto okamžiku zhroutit, neboť je zřejmě sporný. Takový typ inkonsistence však lidskému uvažování nijak nevadí. Spokojíme se s tím, že speciální informace o tučňácích využijeme například k blokování odvození některých dalších údajů např. neodvodíme, že Quido má hnízdo vysoko ve větvích stromů (Mařík, 1993).

Základem monotónního logického vyvozování je deduktivní logika, která obsahuje několik zásadních omezení. Deduktivní logika je nekreativní, protože její závěry jsou vždy obsaženy už v premisách (jak poukázal při kritice scholastické logiky už Francis Bacon). Dedukce nepřidává nic nového, co by se dalo začlenit do zkušenosti. Pouze převádí do explicitní roviny implikace toho, co bylo obsaženo v předpokladech (Hogan, 1998). Deduktivní metoda usuzování je intuitivně sice velmi lákavá, avšak

setkává se s řadou výpočetních nesnází. V první řadě bývá pomalá - i pro nejjednodušší plány potřebuje obrovské množství inferencí; přitom existují výpočetní strategie, které dokážou řešit deduktivní úlohy mnohem elegantnějším způsobem. Za druhé, ryze deduktivní plánování je monotónní - je schopno vytvářet nové závěry, ale už ne zbavovat se těch předešlých. Matematická funkce je monotónní, jestliže její hodnoty průběžně bez oscilací stoupají nebo klesají (Thagard, 2001). Monotónnost predikátové logiky spočívá také v tom, že pokud přidáme nové axiómy do stávajícího systému, všechny původní teorémy zůstávají zachovány. Rozšíření logického systému nevede k restrukturaci jejich předchozích částí (Konar, 1999).V umělé inteligenci bylo několik pokusů, jak učinit logiku nemonotónní, jsou však náročné na počítačový čas a na vybavení. Pravdou také je, že čistě deduktivní plánovač se nedovede poučit ze zkušeností (Thagard, 2001).

Cestou k tvorbě nemonotónních systémů vede skrze přidání dalších mechanismů, které se podílejí na operacích s daty či reprezentacemi. Logický systém tak ale ztrácí svou stručnost a jednoduchost. Většina metod používá mechanismy, které buď vnesou do procesu logického vyvozování míru pravděpodobnosti, která činí výsledek logických operací pravdivým v určité míře (nepohybujeme se již oblasti pravda/lež), nebo jsou tyto mechanismy použity pro další zpracování jednoznačného výsledku. Vždy se však jedná o rozšíření základního logického aparátu. Původní požadavek Ockhamovy břitvy v oblasti logiky použité k simulaci lidského myšlení vede k redukci, která brání plnohodnotnému napodobení. Přesto (nebo snad právě proto) je používání deduktivních metod v oblasti simulace na klasických počítačích velice silným a efektivním způsobem, jelikož plně využívá architekturu a technické možnosti počítače. Nedostatky programů UI se spíše objevují v procesu indukce, tedy vytváření obecných závěrů z omezeného počtu příkladů.

3.1.5 Komputace

Kolem roku 1950 se začínají objevovat nové koncepce způsobu nazírání lidské bytosti. Člověk je (metaforicky) viděn jako stroj a nastává i určitý posun v terminologii používané k popisu kognitivních procesů. Lidé jsou přirovnáváni k výpočetnímu zařízení, které se rodí s určitým hardwarem a je programováno zkušenostmi, socializací a zpětnou vazbou svého vlastního chování. Cílem psychologie je zjistit způsob, jakým lidé zpracovávají informace. Behavioristický model S-R se ukazuje jako nedostačující a pozornost se přesouvá k interním procesům a stavům (Pfeifer&Scheier, 2001). Těm je

následně přisuzován statut výpočetních mechanismů, výsledky jejichž výpočtů tvoří podklady pro projevy lidského chování. Souvislost můžeme hledat v tehdejším rozmachu vývoje počítačů.

Jednou ze základních možností počítačů bylo použít je jako simulátory, na kterých lze napodobovat fungování neuronových sítí (kopií biologických systémů). Poté se ale názor na jejich využití změnil. Vidíme-li lidský mozek pouze jako biologický hardware manipulující s abstraktními reprezentacemi světa, kterému říkáme myšlení, proč používat počítač jako simulátor, když by mohl manipulovat s vlastními reprezentacemi (Hogan, 1998). Čímž se dostáváme k jádru komputace. Je ji možno považovat za teorii, která tvrdí, že veškeré jevy tohoto světa jsou převeditelné do formy rovnice, již je možno vypočítat a výsledek identifikovat s kauzálním následkem počítaných jevů. Komputace se nesoustřeďuje pouze na aritmetické operace s čísly, počitatelné jsou i jevy, které můžeme převést do symbolické roviny a následně je zpracovávat pomocí zákonů logiky (omezení logiky byla zmíněna dříve a jejich platnost přechází i do oblasti komputace).

Základní vyjádření komputační teorie je poněkud sporné, jelikož praví, že svět je počitatelný (tedy libovolná jeho část i jako celek). Posouzením zmíněného tvrzení bychom se dostali až do oblasti filosofie. Navíc se setkáváme s tvrzením, které si klade za nároky vytvořit teorii, jež by v sobě měla univerzálnost vysvětlení pro libovolný typ jevů ve vesmíru (teorii všeho). Netroufám si zatím s tímto tvrzením polemizovat, nabízím pouze krátkou kritiku komputační teorie z pera jiného autora. Je reflexí na komputaci jako přístupu k simulaci lidské mysli (protože pokud je celý vesmír počitatelný, musí být taková i mysl).

Jestliže je teorie myšlení počitatelná (algoritmizovatelná), neznamená to, že počítač myslí. Astronomie je počitatelná, ale vesmír není počítač (Crane, 2002).

Spíše než odpovědí na otázku, zda lze převést celý vesmír (s celou jeho historií) do podoby algoritmu, se jedná o metaforu, vyjadřující se ke klasické Turingově otázce, zda mohou počítače myslet. Crane se přiklání k názorům, které nevidí výpočet jako univerzální nástroj pro deskripci či reprezentaci okolního světa.

Margaret Boden si klade otázku, je-li možné komputační teorii nazývat paradigmatem? Odpověď, kterou sama nabízí, je spíše filosofickým zamyšlením než jednoznačným vyjádřením. Tvrdí, že paradigmatem je v podstatě dostatečně kvalitně formulovaný sociální úzus, což v nás v oblasti komputace příliš neposune kupředu (Boden, 1988).

Pokusme ale pominout pochybnosti o možnostech komputace a podívat se na základy a předpoklady, kterými se vyznačují komputační zařízení používaná jak pro oblast simulace, tak i obecněji pro jiné úlohy.
Komputace se dá vyjádřit třemi propojenými, leč samostatnými myšlenkami:

1. **Matematická funkce**
2. **Algoritmus**
3. **Systémová architektura**

Matematická funkce mapuje z množiny vstupních objektů, nazývaných doména, do množiny výstupních objektů, nazývaných rozsah (např. mapování všech možných vět z anglických slov do množiny všech možných anglických smysluplných vět).
Algoritmus pro funkci f je výpočetní procedura, která počítá f. Může jich být více pro totožnou událost. Můžeme ji vidět jako soubor primitivních operací, např. použití logaritmu místo sčítání a násobení.
Design počítacího stroje, který vykonává algoritmus, je nazýván systémová architektura. Ta obsahuje primitivní operace používané algoritmem a jiné nutné komunikátory mezi nimi, které jsou určeny podstatou stroje. Jsou výpočetními moduly používané algoritmem, ale nejsou částí algoritmu (Sternberg, 1999).
Podobné rozdělení nacházíme i u Marra. Ten rozděluje komputační teorii do 3 vrstev (Boden, 1988).

1. **komputační**
2. **algoritmická**
3. **hardware**

Komputační vrstva (odpovídající předchozí matematické funkci) není přesně ztotožnitelná přímo se slovem komputace – výpočet (proces). Marr jí vidí spíše jako otázku, **co** systém vykonává, než **jak** to vykonává. Blízkým termínem pro lepší pochopení může být Chomského kompetence, popřípadě analýza úlohy autorů Simona a Newella. Komputační vrstva poskytuje abstraktní formulaci zpracovávané úlohy, včetně možností a omezení, které vstupují do hry (Boden, 1988).
Algoritmická vrstva bere tyto informace v úvahu a v konkrétní rovině se snaží o tvorbu správných posloupností operátorů. Jak již bylo zmíněno výše, existuje několik možností, jak algoritmizovat zpracování určité úlohy. Jelikož se pohybujeme v sys-

témech, které jsou založeny na výpočtech, zajišťuje nám jejich aparát možnost isomorfismu. Jedná se o silnější formu ekvivalence mezi dvěma formálními systémy (Luger, 1994). Příkladem může být možnost naprogramovat klasický počítač PC, aby fungoval jako Turingův stroj a naopak. Tím, že zastřešujícím principem je komputační teorie, umožňuje být algoritmické vrstvě částečně nezávislá na použitém hardwaru. Přesněji architektura hardwaru musí být tak univerzální, aby na něm bylo možno provést libovolnou výpočetní úlohu.

Pro pochopení je důležité rozlišení mezi hardwarem a softwarem. Vztah mezi nimi je dán poněkud vágním, avšak filosoficky významným pojmem implementace, což je způsob, jak zajistit, aby daný softwarový program řídil reálný průběh příslušného výpočtového procesu v daném typu hardwaru. Podstata softwarových programů totiž tkví v jejich kauzální potenci *(causal efficiency)*, tj. schopnost řídit dotyčné procesy, která je invariantní k té či oné konkrétní implementaci. V tomto smyslu lze říci, že programy „přežívají hardwarovou smrt" (Havel, 2001).

V psychologii, která se snaží využívat výpočetní procedury k simulaci psychických jevů, vzniká problém černé skříňky, jelikož máme jen omezené možnosti ke zkoumání kognitivních procesů, takže nejsme schopni přesně rozhodnout u daného jevu, za kterou část je zodpovědná architektura a za kterou algoritmy (Sternberg, 1999). To je častým nedostatkem kognitivních architektur, které často vznikají jako myšlenkové experimenty podpořené statistickým výskytem jevu a jejich převod do předem definované architektury (například do von neumannovské) se setkává s mnoha problémy a omezeními.

3.2 Architektura

3.2.1 Charles Babbage

Jelikož se zde v této kapitole seznámíme s prvními pokusy o realizaci komputační architektury, je zde netradičně uveden její vývoj včetně některých podrobnějších údajů o autorovi.

Britský matematik a vynálezce Charles Babbage se již ve 20. letech 19. století pokusil zkonstruovat mechanický výpočetní stroj, jehož činnost byla založena na programovatelných instrukcích. Babbage se pokusil o obnovení myšlenky, kterou se zabýval už Leibnitz. Ten uvažoval o mechanickém uvažování jako o rozšířeném mechanickém počítání, ale nepodařilo se mu najít vhodný jazyk pro reprezentaci okolního světa. Babbagovi se to povedlo s tím, že použil Booleovskou algebru (Hogan, 1998). Chtěl

postavit takový přístroj, který by byl schopen počítat na aritmetické bázi. Přístroj měl navíc v sobě obsahovat prvky logické algebry, umožňující v mnohém napodobit lidské myšlení. Jednalo se Analytický stroj, teoretický koncept, který nebyl nikdy uskutečněn, ale který předznamenal způsob myšlení a pokusů o vytvoření myslícího stroje aktuální až do dnešních dní. Předchůdcem Analytického stroje a jediným realizovaným projektem Charlese Babbage byl Derivační stroj, který v sobě obsahoval tabulky pro výpočet první derivace a byl založen na mechanickém principu. Jeho funkce spočívala ve výpočtu derivace způsobem, který si vyžadoval sčítání jako jedinou funkci. Zkonstruovat jej ale nebylo jednoduché. Babbageovi došlo, že stroj nemůže naprogramovat běžným jazykem, jehož podoba je pro výpočty příliš rozvláčná a komplikovaná. Z tohoto důvodu vyvinul zvláštní jazykovou řeč, jakousi kombinaci čísel a typografických znaků, s jejíž pomocí hodlal stroj programovat.

Z hlediska hardwarové architektury základní model obsahoval aritmetickou jednotku (dnešní procesor) s jedním tisícem ozubených koleček a paměť dat pro tisíc padesátimístných čísel. Důležitou součástí výpočetního zařízení, které měl pohánět parní stroj, byla řídící sekce s programem činnosti zapsaným na řetězci papírových děrných karet (paměťové médium).

Děrné karty byly nápadem, k němuž Babbage inspiroval vynález francouzského obchodníka Josepha Jacquarda. Není bez zajímavosti, že právě „žakárový stroj" inspiroval později, v roce 1890, inženýra Hermana Holleritha k vynálezu děrných štítků přenášejících data. Hollerith těchto štítků využil ke konstrukci třídicího systému, využívaného při sčítání lidu. Jeho projekt sčítání, porovnávání a analýzy čísel na základě děrných štítků byl tak úspěšný, že Hollerith se svými společníky založil společnost, jež byla později přejmenována na International Business Machines – IBM.

V období dokončování tohoto stroje začal Babbage pracovat na návrzích jeho vylepšené varianty, kterým je výše zmíněný Analytický stroj *(Analytical Engine)*. Zařízení mělo kromě základních početních operací umět nejen řešit algebraické a numerické rovnice, ale současně i vystihnout výsledky a podle nich, což je nejpozoruhodnější, samostatně měnit průběh dalšího výpočtu.

K tomuto účelu navrhl Babbage systém tří programovacích karet, opatřených

děrováním (tyto předchůdkyně děrných karet byly paradoxně ještě dokonalejší než výše zmíněné Hollerithovy karty). Jednalo se o operační, číslicové a variační karty, přičemž operační karty dodávaly instrukce stroji (zastupovaly jakýsi operační kód), číslicové karty obstarávaly informace o hodnotě čísel (data) a variační (software) zprostředkovávaly druh výpočtu. Každý druh karet, rozlišený i odlišnými velikostmi, měl tedy svou specifickou funkci pro chod a funkce analytického stroje.
Právě schopnost postupovat a rozhodovat se podle stanoveného programu udělala z Babbageova návrhu předchůdce pozdějších typů počítačů. I když měl vynálezce ambiciózní projekt mechanického analytického stroje promyšlený do nejmenšího detailu, prototyp nebyl za jeho života bohužel nikdy zkonstruován. Docenění své myšlenky se sice nedočkal, ale to neznamená, že zapadla. Babbage je uznáván jako matematik, jehož principy ovlivnily architekturu počítačů, které jsou používány dodnes - von neumannovská architektura (Hogan, 1998). Sám Babbage o svém díle napsal: „Analytical Engine splňuje podmínky, které umožňují stroji vykonávat neohraničené výpočty" (Hogan, 1998,s.134). Nedá se však říci, že by si Babbage uvědomoval tak jasně univerzálnost navrhovaného stroje, jako to udělal o 100 let později jeho krajan A.Turing.

3.2.2 Von neumannovská architektura

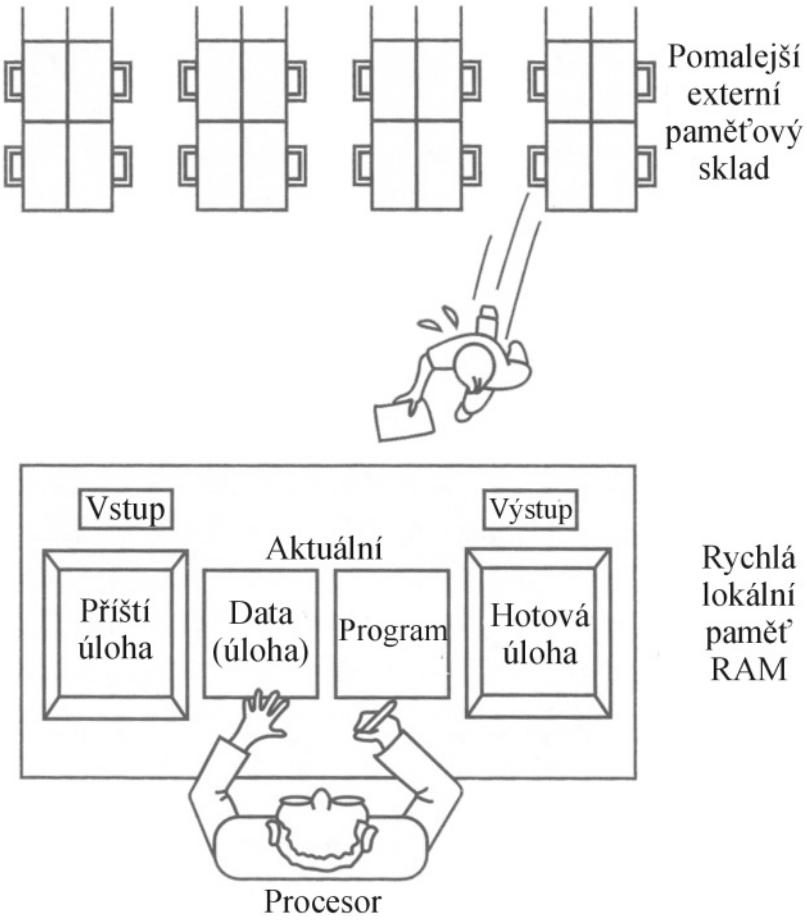
Právě myšlenky zmíněné v předchozí kapitole měly zásadní vliv na tvorbu architektury, která je považována za standard v oblasti komputace. Již v Babbageově přístupu můžeme identifikovat základní prvky současných počítačů (centrální jednotka/procesor, paměť, vstupní a výstupní zařízení pro data).Na konečné podobě se podílelo více badatelů. Norbert Wiener (zmiňovaný již v souvislosti s teorií informace) sepsal několik doporučení, určujících směr, kterým by se měla ubírat tvorba architektury budoucích strojů (Hogan, 1998). Požadavky byly následující:

1. Používat číselnou formu reprezentace, ne mechanická kolečka (v extrémní podobě by se jednalo o rozdíl mezi digitálním a analogovým, zde spíše míněno z hlediska efektivity).
2. Použít elektronky, pro jejich rychlost. Ne převodníky nebo relé.
2. Používat dvojkový kód místo decimálního.
4. Použít interní uložení programů, umístit odděleně od vstupních a výstupních dat.
5. K internímu skladu by měl být rychlý přístup.

Tyto požadavky předznamenávají vylepšení varianty stroje, který 75 let předtím navrhl Charles Babbage, která se dá považovat jako předchůdce von neumannovské architektury (Hogan, 1998).

Příspěvek samotného Johna von Neumanna do takto vylepšené architektury se zdá být zanedbatelný. V roce 1945 přišel s návrhem, že by výpočetní stroje měly obsahovat paměť rychle přístupnou pro procesor, ve které by byly uloženy aktuální program (soubor algoritmů) a také právě zpracovávaná data a jejich mezivýpočty. Narážel na pomalost tehdejších strojů, které ukládaly tyto informace vždy do externí paměti (z dnešního pohledu na pevný disk). Vznikla poslední část architektury počítače dnešních dní - operační paměť. Její uvedení do praxe neproběhlo okamžitě, ale muselo počkat na vynález paměti typu RAM.

Pro lepší představu celkové architektury je zde obrázek, který na metafoře úředníka zpracovávajícího např. výpočet daní, popisuje jednotlivé komponenty architektury (pro snadnější zapamatování jsou na obrázku přímo uvedeny termíny, které popisují současné počítače).



Obr. 7 *Von neumannovská architektura*

Rozdíl mezi tímto typem architektury a Turingovým strojem je následující. Von neumannovská architektura je již navrhována se zřetelem k její praktické aplikaci (vycházela z Ch. Babbage) na rozdíl od Turingova stroje, jež byl vytvořen jako myšlenkový experiment, který si nekladl za cíl praktickou aplikaci. Turingovým cílem bylo vytvořit co nejjednodušší (s nejmenším počtem prvků) univerzální systém. Praktické provedení Turingova stroje by bylo zbytečné. Jeho jednoduchost spočívá v počtu použitých komponent a jejich funkcí, nikoliv však v rychlosti a možnostech v oblasti zpracování dat, což jsou vlastnosti v praxi považované za hlavní přednosti výpočetních systémů.

Jen pro informaci na závěr uvádím krátkou poznámku o další práci Johna von Neumanna v historii tvorby výpočetních architektur. V oblasti umělých systémů se setkáme i s architekturou neuronových sítí, či s paralelními výpočetními systémy. Cílem von Neumanna byla právě taková architektura, která by dokázala zpracovávat informace paralelně. V jeho době však ještě technické možnosti nedovolovaly její tvorbu. Von Neumannových zkušeností bylo využito ke stavbě prvního moderního digitálního počítače EDVAC, který byl bohužel sériový (Caudill&Buttler, 2000).

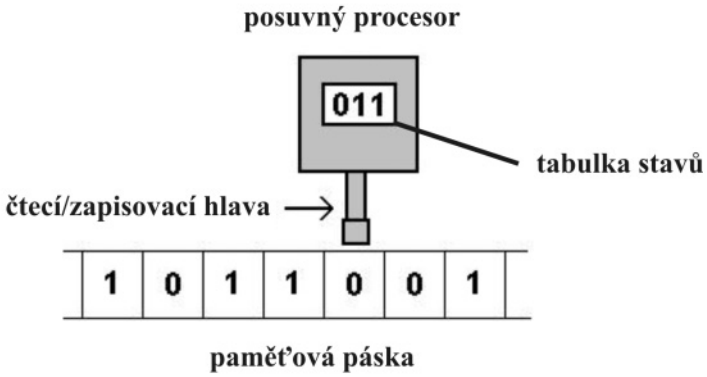
3.2.3 Turingův stroj

Nejznámější prací Alana Turinga je jeho článek s názvem „Computing Machinery and Intelligence" ve kterém vznesl otázku, jejíž řešením se zabývají vědci dodnes. Jedná se o to, zda je technicky možné, aby stroje dokázaly myslet. Tedy zda lze dosáhnout úrovně, kdy vytvoříme systém - stroj, který bude schopen samostatného myšlení, bez potřeby interpretace vstupů či výstupů. Cílem kognitivních věd má být kladná odpověď na položenou otázku. Turing ve svém článku naráží na problematičnost slov stroj a myšlení, které je velmi obtížné vymezit samostatně, natož ve společném kontextu. Zužuje tedy pojem stroj na digitální počítač. Pokud bychom dokázali najít řešení a sestrojili myslící stroj, nastává nám jiný problém. Jakým způsobem zjistit, jestli je daný stroj inteligentní? Což je problematika známá dodnes jako Turingův test (Hogan, 1998).

Podívejme se ale raději na princip fungování Turingem navrhovaného stroje, který má sloužit k vypočtení libovolné úlohy. Stroj (Turingův) není žádné hmatatelné zařízení, je to pokus o matematické zachycení intuitivního pojmu vypočitatelnosti či ještě obecněji vyřešitelnosti. Turing si uvědomil, že každý výpočet (či obecněji každé řešení) začíná nějakými vstupními daty, které si můžeme představit znak po znaku zapsané

na papírové pásce a končí nějakým výsledkem ve stejné podobě (Peregrin, 2002). Výpočet je přechod od jedné sekvence znaků na pásce k jiné, a tak Turing usoudil, že ať už přechod provádíme jakkoli, na té nejelementárnější úrovni se nemůže než skládat z několika základních operací (přečtení nějakého existujícího symbolu, posun pásky o jednu pozici tam či zpátky a zapsání nového symbolu či přepsání starého).

Obr. 8 *Schéma Turingova stroje*



Přestože je Turingův stroj pouze teoretickou konstrukcí, je možno jeho architekturu popsat pomocí následujících prvků. Jedná se o zařízení, které obsahuje tabulku s konečným počtem fyzických nespecifikovaných stavů a posuvnou hlavu schopnou číst, zapisovat a mazat symboly (nejčastěji se používá 1 a 0, ale je možno použít libovolnou konečnou abecedu symbolů). Hlava se pohybuje v diskrétních krocích po libovolně dlouhé pásce (může být nekonečná), která je rozdělena na políčka obsahující vždy jeden symbol. Na začátku každého kroku ovlivňují činnost stroje dva vstupy. Jeden z pásky (symbol) a druhý z tabulky stavů, která dle daného symbolu přiřadí hlavě operaci, kterou má provést. Výsledná operace se skládá z instrukce, co provést s přečteným symbolem (nechat, smazat, přepsat) a určením směru posunu hlavy vlevo nebo vpravo (Hogan, 1998).

Pojmem „vypočitatelný" se rozumí cokoli, co by dokázala vypočítat idealizovaná bytost, která by měla k dispozici neomezené prostředky a neomezeně času. Turing se však pokusil o nepříliš obvyklou věc: myšlenka počítače, tak, jak ho dnes známe, je myšlenkou přechodu od strojů, které vykonávají jeden určitý proces, ke stroji, který je univerzální v tom smyslu, že disponuje natolik flexibilním souborem natolik elementárních operací, že z nich lze skládat v podstatě jakékoli představitelné výpočty (Peregrin, 2002). Tím se dostáváme do ideálního stavu pro komputační teorii. Máme svět, který můžeme převést na soubor výpočtů a máme univerzální zařízení, které dokáže

všechny tyto výpočty uskutečnit. V čem je tedy problém?
Nejprve je nutné zmínit omezení, které výpočetní metoda prováděná Turingovým strojem, ale i jiným výpočetním zařízením obsahuje. Lépe řečeno, hovoříme o požadavcích, které je nutné splnit pro úspěšné použití komputační metody:

Metoda se skládá z konečné množiny jednoduchých a přesných instrukcí, které jsou popsány konečným počtem
symbolů.
Metoda bude vždy produkovat výsledek v konečném počtu kroků.

„Funkční sílu" Turingova stroje nejlépe vyjadřuje teze, která je kombinací myšlenek Turinga a Alonsa Churche.Vyskytuje se v mnoha publikacích v rozličných formách a s různou explanační silou (odvislou od přesvědčení autora). Uvádím zde nejčastěji zmiňované formy Church-Turingovy teze:

Všechny komputační modely jsou stejné nebo méně výkonné než Turingův stroj (Luger, 1994).
Složitost či efektivnost algoritmu je prokazatelná tím, jak ji lze provést Turingovým strojem (Crane, 2002).

Kromě toho, že můžeme používat Turingův stroj jako měřítko zdařilosti (efektivity a stručnosti) algoritmu (Craneova definice), můžeme tezi posuzovat podle míry, kterou je komputace účastná při tvorbě reprezentací a operacích s reprezentacemi prostředí, shrnutelného v psychologii pojmem kognice. Silnější varianta explanace teze praví, že pokud existuje problém, který není řešitelný žádným výpočetním způsobem, nemůže tento problém vyřešit ani lidská mysl (Pfeifer&Scheier, 2001).
Druhá myšlenka, vážící se k zmíněné tezi, tvrdí, že jestliže člověk dokáže řešit problémy, či vykazovat inteligentní chování, mohou být zkonstruovány stroje, které budou mít stejné schopnosti. Tvrzení tvoří jádro současných výzkumů v oblasti umělé inteligence (Pfeifer&Scheier, 2001). Bohužel myšlenka o schopnosti napodobit lidskou mysl výpočetním zařízením je závislá na odpovědi na výše zmíněnou otázku „kvantifikace a algoritmizovatelnosti prostředí" (probíraná také v kapitole Komputace).
Již zmíněná univerzálnost Turingova stroje je podpořena ještě jednou výhodou této geniálně navržené architektury. Tabulku stavů, která rozhoduje podle vstupu z pásky o tom, jaká bude příští operace hlavy, můžeme totiž převést do symbolické formy,

která je používána právě pro zápis na pásku (nejčastěji binární kód 1/0), a vytvoříme tak kopii Turingova stroje (v podobě dat). Tu pak můžeme implementovat do nového prázdného Turingova stroje pomocí hlavy, která načte tabulku stavů na pásce uloženou. Bohužel se v žádné publikaci neuvádí, jak takový převod vypadá v praxi. Jak říci prázdnému Turingovu stroji, že má načíst tabulku stavů (popřípadě jak je veliká)? Také ukončení načítání je problematické a je třeba vytvořit speciální symbol a také receptor v čtecí hlavě. Znamenalo by to rozšíření stávající architektury o nové prvky.

O možnostech rozšíření Turingova stroje a také jeho omezeních hovoří i některé současné práce. Klasická Church-Turingova teze nás ujišťuje, že každý algoritmus lze popsat pomocí standardního Turingova stroje. Nicméně pro výpočty současných osobních počítačů anebo velkých distribuovaných systémů se navržená architektura nehodí. Výpočty podobných systémů se liší od klasických ve třech směrech: nikdy nekončí, průběžně a nepředvídatelně interagují se svým okolím a výpočetní systém se dynamicky a nepředvídatelně vyvíjí. Wiedermann navrhuje rozšíření shora uvedené teze tak, aby pokrývala i právě zmíněné tzv. neuniformní interaktivní výpočty. Rozšířená teze tvrdí, že každý výpočet uvedeného druhu lze zachytit pomoci interaktivního Turingova stroje se speciálním typem orákula, které nezávisí na daném vstupu, ale pouze na jeho délce. Příklady, které rozšířená teze pokrývá, sahají od (formálních) modelů osobních počítačů přes Internet až po komunity inteligentních mobilních agentů podléhající neuniformní evoluci. Výsledné výpočetní systémy mají superturingovskou výpočetní sílu (Wiedermann, 2001). Používané metody rozšíření základního principu korespondují s těmi, které byly zmíněné v kapitole o formální a mentální logice. Je nutné přidat další mechanismy a vytvořit složitější architekturu, která by dokázala zpracovávat i komplexní a dynamické úlohy.

Podívejme se, jaké argumenty používají odpůrci komputační teorie, vyjádřené pomocí Turingova stroje. Velmi často je citována námitka Lady Lovelace, která tvrdila, že přístroje (počítače) vždy odpovídají stejným způsobem na stejné vstupy (Hogan, 1998). Lady Lovelace tedy tvrdí, že stroj nemá schopnosti cokoliv vymyslet. Může pouze informace zpracovávat. Je nutné připomenout, že Lady Lovelace žila ve stejném období jako Charles Babbage a její vyjádření se vztahuje spíše k obecné komputační teorii, než k samotnému Turingovu stroji (který ještě nebyl v její době navržen). Námitka směřuje k monotonii výpočetních systému a také k otázce, zdali můžeme automatický proces nazývat inteligentním.

Poněkud současnější námitka hovoří o rozdílech mezi diskrétním zpracováním oproti spojitosti v nervovém systému. Jakákoliv drobná chyba či odchylka v práci nervového

systému, způsobí sled události, který nelze předpokládat, či nasimulovat pomocí systému používajícího diskrétní stavy. V každém člověku je omezený počet pravidel, které mohou působit na jeho chování. Ale taková pravidla skutečně neexistují, takže člověk nemůže být stroj (Hofstadter, 1999). Částečnou odpovědí na Hofstadterův kategorický postoj je Wiedermannem navrhované rozšíření Turingova stroje o komponenty schopné zpracovávat komplexní dynamickou úlohu. Námitka ale útočí i na samotné základy komputační teorie s opodstatněním, že diskrétní forma informace či jejího zpracovávání není sto plně nahradit formu analogovou.
Což nás směruje k další problematické partii, která se týká například oblasti *fuzzy* množin. Bohužel v této práci není dostatek místa, pro podrobnější pohled na danou oblast. Pouze pro doplňující informaci na závěr uvádím několik postřehů cizícch autorů, týkajících se rozdílů mezi analogovým a digitálním. Což je tématika, jejíž odpovědi jsou předpokladem pro otázky ohledně „kvantifikovatelnosti (reprezentace) světa".

Analogové měření je průměrování průběžné kvantity (Luger, 1994).
Analogový x Digitální – reálné číslo x celé číslo – kupování mléka x kupování vajec
Diskrétní forma je rozpojitá, kvalifikovatelná, operacionalizovatelná, neobsahuje kontinuum a tím problematiku nekonečně malých veličin. Kontinuum lze rozdělit na diskrétní informace, ale ztratíme přitom kvalitu informace a také některé její hodnoty. Pokud ale chceme diskrétní informace propojit v kontinuum, musíme extrapolovat (Sedláková, 2004).

3.2.4 Finite state automaty

Speciálním případem Turingova stroje jsou konečné automaty *(finite state automata).* Přesněji se jedná o automaty s konečným počtem stavů. Zastupují jednoduchou formu výpočetního zařízení, jejichž výchozím principem je právě Turingův stroj. Oproti němu se jedná u těchto automatů o konkrétní systémy, které se používají v praxi, ale bylo nutné provést některá zjednodušení. Požadavek na nekonečnou pásku je v praxi neuskutečnitelný. A proto používají konečné automaty buď omezeně dlouhou pásku nebo žádnou (v tom případě jsou vstupy reprezentovány jinou formou). Interní paměť je tvořena tabulkou stavů (jako u Turingova stroje), která je podle názvu automatů konečná a reprezentuje sled úkonů (algoritmů), které může stroj plnit (z pohledu počítače se jedná o software). Interní paměť tvoří tedy pouze software a neumožňuje

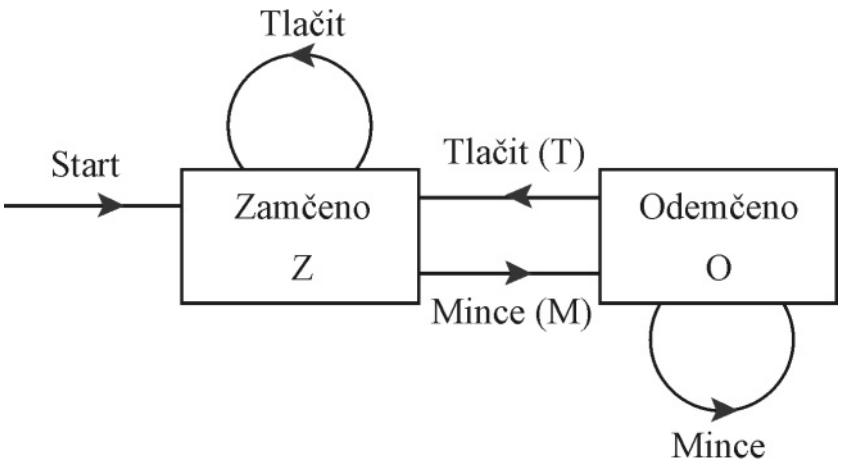
uložení informací o reprezentacích či mezivýpočtech vstupních dat. Základní definice konečných automatů se dá shrnout do těchto bodů:

1. Počet stavů automatu je diskrétní a přesně rozlišitelný.
2. Počet stavů je konečný.
3. Vstupy a výstupy probíhají v libovolném z těchto stavů.
4. Neustále probíhá přechod mezi jednotlivými stavy
5. Systémy nemají žádnou externí paměť. Veškerá interní paměť jsou pouze stavy a jejich posloupnosti. To, jak stavy budou přecházet, je částečně dáno jejich obsahy, ale také informacemi které do tohoto systému vstupují

(Hogan, 1998).

Dobrou demonstrací jednoduchých konečných automatů je turniket u vchodu do metra, který se dá projít po vhození mince. Obrázek znázorňuje formalizaci zařízení, obsahující funkční diagram a také tabulku stavů.

a) Diagram



b) Tabulka stavů

Počáteční stav	Cílový stav	
	Vstup T	Vstup M
Z	Z	O
O	(Otočit) Z	Z

Obr. 9 *Finite state automata – příklad turniketu v metru*

Rozšířením možností konečných automatů můžeme přejít od deterministických k

nedeterministickým KA. U deterministických je v tabulce stavů přiřazen pouze jediný přechod na následný stav při určitém vstupu. U nedeterministických automatů obsahuje tabulka stavů více než jednu možnost přechodu k následnému stavu.

3.3 Aplikace

3.3.1 Symbolické systémy

Symbolický funkcionalismus, proud zvaný též jako „stará-dobrá-umělá-inteligence" (GOFAI), je založen na dvou základních hypotézách – funkcionalistické hypotéze a hypotéze fyzického symbolického systému . Funkcionální hypotéza tvrdí, že:

Inteligentní chování daného systému je dosaženo interakcí mezi jednotlivými komponenty, které disponují odlišnou funkcionalitou, což je dosaženo tím, že v rámci systému hrají odlišnou roli (Pěchouček, n.d.).

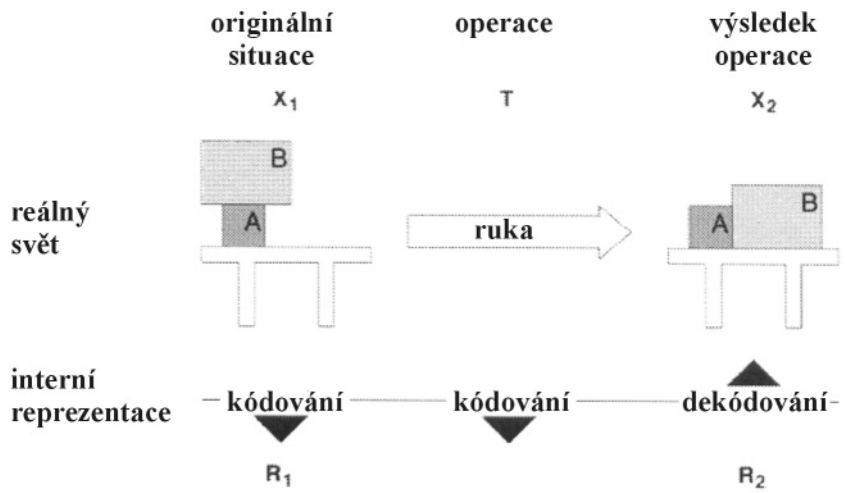
Skrze tuto poněkud tautologickou definici se dostáváme dál ve vymezování a revizi prostředků používaných při snahách o simulaci inteligentního systému. V předchozích kapitolách jsme rozebírali jak zastřešující téma komputační teorie, tak architektury, které dokáží vytvořit podmínky pro komputační možnosti. V následujících kapitolách se posuneme o kousek dál. Z oblasti „hardwarové" se přesouváme do oblasti „softwaru". Zmíněny budou některé základní programy, které jsou implementovány do klasických von neumannovských architektur. Autoři se snažili o jejich nejefektivnější využití při tvorbě „inteligentních algoritmů". První kapitola nastiňuje oblast zpracování symbolů matematické i nematematické povahy, a obtížnosti jejich zachycení ve vhodné podobě k následnému zpracování. Jedná se o fyzický symbolický systém zmíněný na začátku kapitoly.

3.3.1.1 Fyzický Symbolický systém

Hypotéza fyzického symbolického systému tvrdí, že:
„Fyzický symbolický systém je dostatečným a nezbytným prostředkem pro prezentaci inteligentního chování". Autory hypotézy a následné teorie jsou Allen Newell a Herbert Simon. Z výchozích principů, které autoři zastávali, se jedná spíše o empirický než teoretický přístup ke zkoumání lidské inteligence. Inteligence je zde nazírána jako manipulace se symboly, přičemž systém fyzických symbolů je nutnou podmínkou pro

její realizaci. Slovo fyzický je myšleno ve smyslu nutnosti uložení symbolů do určitého fyzického média. To, jakým způsobem je uložen je nepodstatné.

Mechanismus práce se symboly je následující. Výpočetní proces pracuje s reprezentacemi, které mají formu symbolických struktur. Symbolická reprezentace se řídí podle Simona a Newella zákony reprezentace. Příkladem je způsob umístění objektu v prostoru a zachycení jejích vztahů. Pokud máme stůl, na kterém leží objekt A a na něm leží objekt B, můžeme to zachytit následující reprezentací:



Obr. 10 *Zachycení změny pomocí symbolické reprezentace*

(objekt A)(Objekt B)

(Stůl S)

(na B A) (na A S)

Jestliže provedeme operaci přemístění objektu B z objektu A na stůl, vzniká nám nová reprezentace

(objekt A)(objekt B)

(stůl S)

(na A S) (na B S)

Jde tedy o vytvoření interních reprezentací vždy korespondující s externím světem. Tím že je použit k interní reprezentaci jazyk (symbolický systém), vznikají určitá omezení plynoucí z jeho použití. Jedná se o zachycení prostorových vztahů, které by v případě složitější situace vedlo k nutnosti použít stále více operátorů. Nevýhodou při převodu je právě vyjádření prostorových vztahů jazykovým kódem místo obrazového. Se vzrůstající komplexitou (větší počet prvků,vztahů a jejich proměn v čase) vznikne situace, kdy pro záznam měnící se situace potřebujeme takové množství popisujících

prvků, že se daný symbolický systém stane neefektivním a bude vyžadovat značné zatížení výpočetního aparátu. Přes všechny námitky, se tento způsob reprezentace stal standardem v oblasti umělé inteligence. Největší možnosti jeho využití nabízejí produkční systémy, které pracují se symboly uloženými ve formě znalosti a jejichž problematiku zpracovali stejní autoři (viz následující kapitoly).

3.3.2 Simon-Newell

V této kapitole načrtneme hlavní způsoby používání výpočetních automatických programů. Mezi průkopníky praktické aplikace výsledků kognitivní psychologie a pokusů o simulaci rozhodně patří zakladatelé tohoto odvětví Herbert Simon a Alan Newell. Blíže se seznámíme s výchozími principy, které tito autoři zastávali a také s jejich programy tvořící základy oblasti umělé inteligence a strojového myšlení. Důležitou otázkou se zde stává tvorba univerzálního algoritmu (či programu) schopného zpracovávat úlohy z různých oblastí (libovolný typ úlohy). Většina zde zmíněných programů implicitně předpokládá svou aplikaci v klasické von neumannovské architektuře, o které víme, že je založena na principech komputace.

Jedním z možných přístupů při vytváření inteligentních systému je metoda *„top-down"*, což znamená, že postup zpracování informace a tvorby chování je řízen z vrchní abstraktní úrovně a poté je spodní konkrétní vrstvou vykonáván. Newell tuto koncepci rozvrhl do tří rovin. Vrchní úroveň se nazývá „znalostní" a obsahuje přehled znalostí a cílů, které může systém dosahovat. Prostřední logická úroveň vytváří posloupnost operací vedoucích k řešení a zpracovává je do algoritmické formy. Spodní úroveň implementační poté přiřazuje jednotlivé operace konkrétním programům na jejich vykonání (Pfeifer&Scheier, 2001). Setkáváme se s podobným rozčleněním, jaké zastupují v obecné rovině jednotlivé vrstvy komputační teorie.

Simon tvrdí, že lidská mysl je v rovině informačního zpracování velmi jednoduchým zařízením, a její zmiňovaná komplexita je způsobena komplexitou prostředí. Pokud chceme odhalit základní výpočetní procesy, které používá mysl, musíme studovat zdánlivě jednoduché situace, které nám tyto procesy odhalují. Zde podle něj leží základní jádro výzkumu kognitivních psychologů. Na druhou stranu lze o daném přístupu říci, že badatelé v této oblasti vytvořili obrovské množství paradigmat s předpokladem , že se jim podaří izolovat čisté případy kognitivních akcí, a také uskutečnili spoustu laboratorních studií, o kterých předpokládali, že budou fungovat i mimo laboratoř (Sternberg, 1999). Za těmito studiemi lze vidět potřebu, nalézt či odvodit výchozí

„atomy" psychologie, jejichž kombinace mohou vést k tvorbě (a také dokonalému pochopení) jevů komplexnějších.
Simon také kriticky připomíná že od Woodswortha k Andersonovi (autoři knih týkajících se kognitivní psychologie) se událo jen málo změn v kapitolách věnovaných myšlení a uvažování. Jednou z příčin může být malé propojení mezi problematikami řešení problému, uvažování a indukce (Pick, 1992).
V této oblasti Simon své výchozí principy abstrahoval spíše z pozorování prostředí než racionalistickým přístupem. Procesy myšlení či rozhodování jsou řízeny principy zisku, minimálních nákladů, efektivity apod. Matematické a logické formule hovoří o přesných hodnotách a o nutných podmínkách, kterých je třeba k dosažení závěru, zatímco člověk je daleko více řízen principem „dostatečnosti". Stačí nalézt dostatečné množství faktů či argumentů k rozhodnutí.
V rozhodovacích postupech identifikoval hierarchickou strukturu. Jednotlivé akce vedoucí k cíli, jsou členěny na podúkoly, samostatně vykonávané specializovanými centry na nižší úrovni. Příkladem pro tuto analogii mu byla hierarchická organizace a plánování ve velkých podnicích. Přestože je tento postup plně formalizovatelný, dospěl Simon k závěru, že tradiční matematika není dostačující pro modelování rozhodovacích procesů a začal hledat jiný způsob reprezentace (Newell and Simon, 1972, s.124).

3.3.2.1 Logic Theorist

Autoři prvního z programů umělé inteligence (Simon a Newell) se zaměřili na ověřování teorémů v elementární symbolické logice. Ověřování logických teorémů je jednoduchý a posloupný proces, který je podobný odvozování geometrických teorémů. Vycházíme z množiny premis nebo axiomů neredukovatelných na nižší úroveň. S axiomy můžeme poté operovat podle pravidel inference abychom získali tvrzení v podobě teorémů, které díky přesně daným pravidlům, jsou vždy pravdivé v axiomaticky vázaném systému. Newell a Simon používali v procesu ověřování teorému „výrokového kalkulu", složeného z propozičních výroků spojovaných pomocí operátorů „nebo" a „implikuje" do výrazu, které jsou také propozicemi a je možné jim přisoudit pravdivostní hodnotu. Na takových základech byl postaven jejich první program Logic Theorist. Primárním cílem programu nebylo ověřovat teorémy, které již logikové potvrdili. Autoři měli cíl mnohem více psychologický. Chtěli se dozvědět, které druhy pravidel lidé používají, pokud hovoří o intuici a dalších principech, jež

nejsou přímo přístupné zkoumání. V případě potvrzení využitelnosti pravidel s použitím rozsáhlého systému axiomů, by tak bylo možné vytvářet programy, které dokážou výkonnostně předstihnout lidské schopnosti. Ačkoli bylo hlavním cílem při tvorbě LT odhalení zákonitostí myšlení, ve svém důsledku přinesl výsledky spíše v oblasti generování či prohledávání rozsáhlých stavových prostorů, čehož bylo později využito v oblastech jako jsou šachové programy, řešení problému apod.
Program při prohledávání stavových prostorů používal také heuristická pravidla, umožňující ohodnotit jednotlivé mezistupně cílového stavu, čímž jsou schopny redukovat počet variant ve stavovém prostoru, právě podle blízkosti k cíli. Simon s Newellem použili zpočátku program LT k ověřování teorémů z knihy Russella a Whiteheada Principia Matematica. Z 52 teorémů dokázal Logic Theorist ověřit 38. Nutno podotknout, že tento první program z oblasti UI dokázal ověřit jeden z teorému v knize Principia Matematica elegantněji než její autoři (Hogan, 1998).

3.3.2.2 General Problem Solver

Výzkum a tvorba programů napodobujících lidskou inteligenci narážel v dobách, o kterých jsme hovořili v minulé kapitole a o které budeme hovořit i nyní na jeden zásadní nedostatek. Technologie nebyla ještě na úrovni, umožňující zabývat se **obecnou inteligencí.** V teoretické oblasti se objevují dostatečné podklady pro tvorbu obecných výpočetních systému, ale jejich konstrukce (období relé a elektronek) byla extrémně nákladná a výpočetní rychlost stále zanedbatelná. Vývoj a výzkum v oblasti aplikací se tedy vydal směrem oddělených projektů, soustřeďujíc se na jednotlivé aspekty, konstituující celkovou inteligenci potažmo mysl . Výzkumné projekty si přestaly činit nárok na řešení problematiky umělé inteligence „jedním tahem".
Přesto byl dalším počinem dvojice program General Problem Solver, spuštěný v roce 1957. GPS vycházel z předpokladu, že zpracování informace je spíše doménově obecné než doménově (oborově) specifické. (Sternberg, 1996) V různých obměnách a variacích na něm pracovali Simon a Newell po dobu 10 let. Kniha Human Problem Solving obsahuje 920 stránek podrobného záznamu celého projektu (Newell and Simon, 1972).
Jak již název napovídá, byl GPS určen k tomu, aby dokázal řešit obecné problémy. To je základní rozdíl oproti LT, který sloužil pouze k ověřování logických teorémů. Jejich teoretické předpoklady vycházely z lidské schopnosti, řešit libovolný problém. V praxi se jednalo o využití heuristických principů a jejich zakódování do „jádra", které

by dokázalo zpracovávat úlohu nezávisle na jejím zadání. Rozdíl oproti LT je, že axiomy (tedy základní stavební kameny vyvozování) nahradíme znalostní bází, obsahující sérii základních postupů (receptů) při řešení problému. Pokud je systému dána znalost o určitém aspektu světa, schopnost obecného usuzování mu umožní vyřešit problém. Požadované informace byly systému dodávány formou „diferenčních tabulek", specifikující rozdíl mezi danou situací a cílovým stavem. Pro zpracování těchto tabulek se používá *means-end analýza* (analýza prostředků a cílů). Na podobném principu je založeno například zpětnovazebné inženýrství Norberta Wienera.V základě se dá způsob práce GPS shrnout do následujících kroků:

1.Zjisti rozdíl mezi současnou pozicí a cílovým stavem.
2.Najdi operátor který typicky redukuje tento rozdíl.
3.Urči , jestli může být operátor aplikován na danou situaci.
-pokud ano, použij jej
-pokud ne, urči situaci, za které může být operátor použit.
(tvorba nového podúkolu)
4.Vrať se na 1.

Díky použití *means-end analýzy,* která je v současnosti standardní technikou UI, získal GPS základní schopnost formulovat plány. Je mu to umožněno právě díky tvorbě podúkolů *(subgoal)*, umožňující řetězené použití operátorů, které samy nedokážou dosáhnout cílového stavu. Modifikací podmínek pomocí jiných operátorů je dosaženo výsledků, jenž nejsou ovlivněny pouze algoritmem programu, ale i vstupními daty. Ty ovlivní způsob práce a tvorbu posloupností (sekvence) operátorů (Hogan, 1998). Program splňuje podmínku obecnosti *(general purpose)*, jelikož jeho algoritmus je definován v takové rovině obecnosti a znalostní báze takovou formou, že je schopen fungovat nejen v oblasti logiky ale i při řešení problémů, hraní her apod.
Systém pracuje s jedním typem heuristiky, tvořící základní funkční jednotku, používanou iterativně až do vyřešení problému. Oproti souboru axiomu u LT lze vidět znalostní bázi jako krok kupředu, ale jelikož je tato tabulka doplňována a spravována tvůrcem či uživatelem, má systém jen omezené informace o řešení úlohy (není dostatečně propojen se světem, ze kterého úloha pochází) a tak dokáže řešit pouze úlohy, na které dostačuje diferenční tabulka.
Koncem 60-tých let začalo být jasné, že GPS není schopen flexibilně zachytit variabilitu lidského chování. Nejslabším článkem systému jsou zmíněné diferenční tabulky.

Způsob uložení znalosti do těchto tabulek je zcela mechanický a nemá možnost se modifikovat, či přizpůsobit měnící se či jinak definované situaci. Vzniká problém ve způsobu uložení znalostní báze. GPS také na rozdíl od lidského uvažování neprobíhá vždy jasným směrem. Člověk není otrokem svých cílů, a během řešení problémů může použít i několik kroků „stranou", čehož není algoritmus GPS schopen. Obecnost algoritmu a jeho volnost se v očích kritiky jeví jako příliš jednoduchá a hraničící s redukcionismem.

Pokus o vylepšenou verzi GPS proběhl nahrazením diferenčních tabulek pomocí produkčních pravidel (znalost je uložena v logické operaci IF-THEN – zachycení vztahu, souvislosti). Opět se ale objevuje otázka, zdali jsou produkční pravidla dostatečná pro reprezentaci znalostí takovým způsobem, aby systém fungoval flexibilně. Přestože byly vytvořeny obří projekty, zaměřené na tvorbu a shromaždování velkého množství produkčních pravidel (například Large Production System Project, Instructable Production System či výše zmíněná encyklopedie *common sense* CYC), způsob uložení ani vylepšená verze algoritmů se nebyly schopny přiblížit lidské schopnosti řešení problému, který je součástí obecné inteligence (Hogan, 1998).

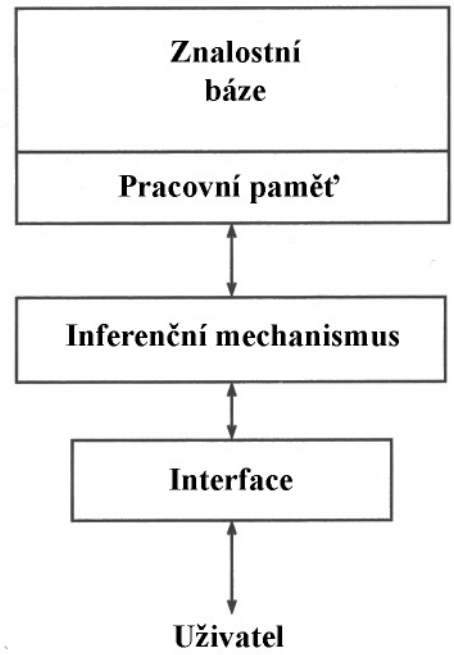
Posledním příkladem je SOAR, systém založený na architektuře produkčních systému. Jeho tvůrci jsou Newell, Laird a Rosenbloom. Opět je použit princip zpracování symbolů, přičemž využití různých typů symbolů mu umožňuje zpracovávat širokou škálu úloh (krok směrem k univerzálnosti). Z hlediska použitých algoritmů v sobě integruje dva základní principy používané v oblasti UI. První z nich je heuristický princip využívající metodu *means-ends analýzy,* založené na redukci rozdílu mezi současným a cílovým stavem. Druhý princip je více specifičtější. Jedná se výkonnou heuristiku založenou na principu zpětné vazby. SOAR je velmi podobný předchozímu systému GPS (kterému byl později přisouzen statut expertního systému založeného na znalostech). Mnoho kritiků proto hodnotilo SOAR pouze jako expertní systém. dodávají, že hovořit o něm jako o systému simulujícím obecnou inteligenci, je zavádějící (Sternberg, 1999).

3.3.3 Expertní systémy

Na konci předchozí kapitoly je zmíněno, že poslední z řady programů simulujících obecnou inteligenci SOAR byl spíše považován za expertní systém než obecný. Je až kuriózní, že právě pokusy o tvorbu univerzálního systému pro řešení problémů ve svém důsledku vedly ke konstituci opačné oblasti. Expertní systémy (také nazývané

produkční systémy) jsou v současné době využívány k naprosto specifickým činnostem v rozličných oblastech lidské činnosti. Ať se jedná o diagnostiku v lékařství (MYCIN), hledání ložisek drahých kovů (PROSPECTOR) či organickou chemii (DENDRAL), ve všech oblastech nacházejí expertní systémy úrodnou půdu. Předpokladem úspěšné aplikace jsou možnosti jejich znalostní báze a rychlost s jakou k ní může přistupovat, tvořící společně výkonné jádro pro práci v předdefinované oblasti. Expertní systémy selhávaly díky tomu, že nedokázaly dobře identifikovat obecnou úlohu. Stačilo vytvořit pevný rámec pole působnosti jejich činnosti, a schopnosti automatického zpracování informací rázem předčí možnosti člověka v daných oblastech.
Existuje rozdíl mezi expertními systémy a konečnými automaty či programy? Ano, zpracování informace není kontrolováno programem nebo procedurou jako u klasických procedurálních programovacích jazycích, ale je prováděno pomocí produkční pravidel (párů IF-THEN), které mohou být použity kdykoliv jsou jejich podmínky uspokojeny. Takže zpracování úlohy se mění během zpracování v důsledku dynamicky se měnícího obsahu pracovní paměti, která ovlivňuje produkci (Harnad, 1990). Můžeme nalézt i podobnosti plynoucí z použité architektury. Stejně jako u konečných automatů, můžeme i zde identifikovat deterministické a nedeterministické typy expertních systémů. Jestliže je v jednom okamžiku možno použít více možností aplikace pravidel, nazýváme systémy nedeterministické. V případě, že posloupnost pravidel nenabízí možnost alternativy, tedy že máme k dispozici vždy pouze jediné pravidlo v daném stavu, hovoříme o deterministickém systému. Testujeme-li determinismus v expertním systému, snažíme se aplikovat znalostní bázi na konkrétní výchozí a cílový stav. Pokud bude algoritmus deterministický, nemůžeme tento závěr aplikovat na celou znalostní bázi, ale pouze na tento konkrétní příklad (Konar, 1999).
Příkladem hybridního expertního systému, založeného na použití dvou architektur je systém 3CAPS. Obsahuje soubor produkčních pravidel (procedurálních znalostí), umožňující manipulace se symboly, které se nacházejí v pracovní paměti. Jeho odlišnost od klasického produkčního systému spočívá ve využití mechanismů paralelních počítačů a dá se shrnout do 3 bodů. Zaprvé, každá reprezentace má svou aktivační úroveň, která vyjadřuje dostupnost reprezentace pracovní paměti. Aktivační úroveň musí dosáhnout určitého prahu (stejně jako v neuronových sítích váha propojení), aby bylo produkční pravidlo opravdu přístupné v pracovní paměti a mohlo být použito. Za druhé je zpracovávání postupné. Pokud je produkční pravidlo používáno častěji s určitým elementem, zvyšuje se aktivační úroveň daného elementu ale také

výstupu, který vniká propojením použitého pravidla a elementu. Za třetí je to využití paralelního zpracování, tedy že může být používáno více pravidel v jednom okamžiku, pokud jsou splněny podmínky pro jejich použití. Architektura je ukázkou propojení komputačního (symbolického) a konekcionistického přístupu. Symbolický systém pracuje v horní vrstvě zpracovávání (abstraktní úroveň), zatímco konekcionistické sítě zajišťují funkci v základní rovině (Sternberg, 1999).



Obr. 11 *Komponenty expertního systému*

Mezi základní obecné vlastnosti produkčních systémů patří citlivost a stabilita. Systémy s dobrou citlivostí dokáží reagovat odlišnými inferencemi i když jsou rozdíly ve vstupních datech malé (kategorizace). Stabilita pak hovoří o schopnosti provést inferenci pro libovolná relevantní vstupní data (komplexita). Dobré výsledky zajišťuje nerozpornost inferenčního mechanismu společně s dostatečnou velikostí znalostní báze (Konar, 1999).

Expertní systémy (produkční systémy) mohou být přirovnány k technikám z oblasti **řešením problému (problem solving)**, založené na hledání cíle pomocí stavového pole. Pojmem stavové pole je myšlena grafická forma reprezentací stavů systému tak, že jsou možné stavy kauzálně rozmístěny v prostoru, což zvyšuje přehlednost vyjádření. Objevují se i ve formě matematické formalizace , která již nenabízí výhody přehlednosti. Nejpodobnější produkčním systémům je algoritmus hledání pomocí

uspořádaného prohledávání (best-first search). Rozdíl je pouze ve vybírání následujícího stavu. Produkční systémy používají **strategii rozlišení konfliktu**, **algoritmus uspořádaného prohledávání** vybírá stavy, který mají nejmenší číselné vyjádření **hodnotící funkce**. Ta se uvádí jako součet hodnoty optimální cesty od počátečního do daného stavu a **hodnoty optimální cesty** z daného do cílového stavu (Konar, 1999, s. 184, zvýraznil M.V.).

3.3.3.1 Problem Solving (Řešení problémů)

V předchozím odstavci jsou některé pojmy vytištěny tučně. Jedná se o odbornou terminologii užívanou v oblasti řešení problémů. Důležité je však rozlišit, z pozice kterého vědního oboru o řešení problémů hovoříme.

Počátky dané problematiky bychom našli v experimentální psychologii, která se snažila v rámci studia myšlení pokrýt i oblast řešení problémů. Psychologie vymezuje rámec dle typů problémů se kterým se setkáváme (problémy s mezerou, příliš složité problémy) dle způsobu zadání apod. Obecné způsoby řešení jsou následně přisuzovány mechanismům, které jsou uznávané tím kterým psychologickým směrem (restrukturace, vhledem). Příkladem může být Sternbergovo dělení faktorů, které způsobují složitost řešené úlohy (Sternberg, 2000):

1. počet kroků (procesů)
2. počet komponent
3. zátěž paměti a pozornosti
4. zátěž adaptability (exekutivy a metakognice).

Úroveň obecnosti při studiu řešení problému je v psychologii značná. Většinou se snaží vytvořit univerzální model, pomocí kterého lze řešit problém ať je zadán libovolně. Je pravděpodobné, že při takovém způsobu přemýšlení můžeme lehce sklouznout k redukcionismu (stejně jako při míře zhuštění v tomto odstavci, která předpokládá základní psychologické znalosti).

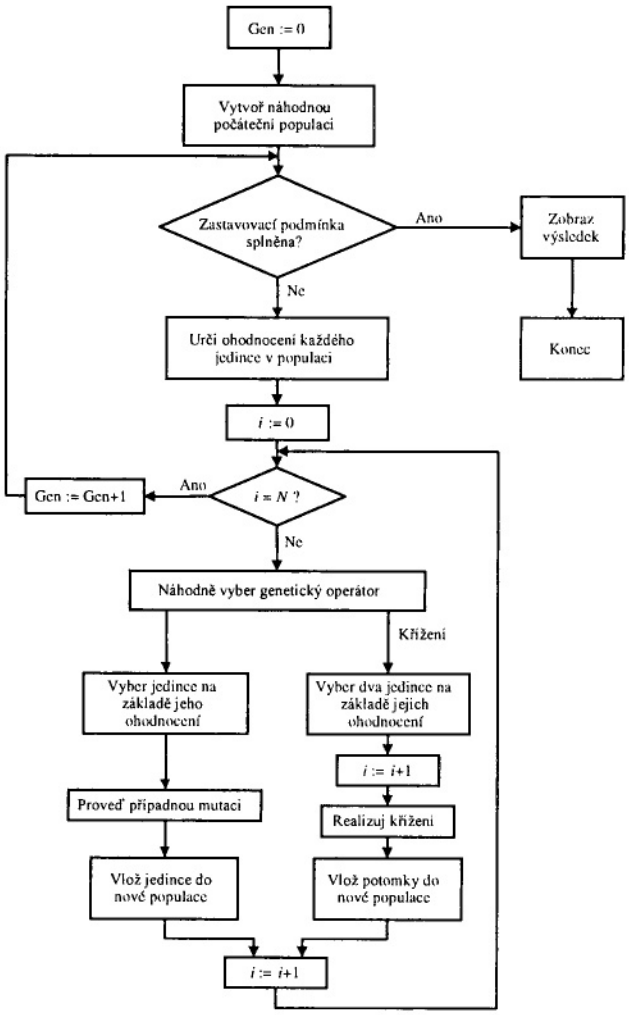
Řešení problémů je ale také samostatnou pasáží v oblasti UI. Oproti psychologii se ocitáme v jiné pozici. Pokud se psychologie snažila vypozorovat z chování a introspekce principy, kterým člověk řeší problémy, UI se snaží aplikovat tyto postupy v oblasti napodobování. Jestliže psychologie vytvořila o způsobu řešení problémů několik teorií, UI se snaží vyjít z jednoho principu (prohledávání stavového pole) a jeho neustálým vylepšováním a přidáváním mechanismů (což jsou ona tučně vytištěná slova v předchozí kapitole) z něj udělat obecně použitelný princip.

Již v pasáži o expertních systémech se objevuje pochybnost, zdali je zvolený směr správný. Jeho volba je daleko více spojena s možností efektivní aplikace v konkrétních oblastech (spojené s ekonomickou stránkou) a také dokonalé využití stávající architekturou. Metoda prohledávání stavového pole se z počáteční fáze „brute-force", tedy mechanického prohledávání všech možností stavového pole, stává přidáváním doplňujících mechanismů stále sofistikovanější, přičemž hranice jejich možnosti jsou totožné s limitami použité architektury a symbolického systému.
Možným způsobem při vylepšování metod prohledávání stavového prostoru, je přidání takového mechanismu, který je známý z oblasti biologie, a jehož „rozhodovací síla" a výkonnost byla ověřena v průběhu miliónů let . Hovoříme o využití principů genetiky a eugeniky v oblasti UI, hovoříme o genetických algoritmech.

3.3.3.2 Genetické algoritmy

Genetický algoritmus (GA) je stochastický algoritmus sloužící k napodobování procesů biologické evoluce. Vychází ze základních Darwinovských principů s využitím pravidla „přežití nejlepšího" (cílový stav je ztotožnitelný s pojmem „nejlepší"-*fitness*). GA jsou nejčastěji používány v oblasti inteligentního prohledávání, strojového učení a problému optimalizace. Informace o stavech systému jsou zachycovány jako binární řetězce, které pak tvoří jednotlivé „chromozomy". Operace, které jsou na nich následně aplikovány, se dají přirovnat k biologickým termínům křížení a mutace. Právě křížení a mutace mohou zajistit úpravu počátečního stavu do podoby, která v sobě obsahuje nové prvky, tzn. že je restrukturována. Jádrem je pak mechanismus, který každou novou generaci (tedy pozměněnou verzí problému) porovnává s cílovým optimálním stavem. Míra shody s cílovým stavem je právě míra jeho *fitness*. Nejčastěji používáme náhodnou změnu v každém kroku kontrolovanou s ideálem, až dosáhneme ideálního cílového stavu. Oproti klasickému náhodnému prohledávání stavového prostoru tedy nepostupujeme systematicky (Konar, 1999). Pro lepší představu uvádím diagram, který zachycuje posloupnost kroků GA.
Pokud ale neznáme cílový stav, nelze odvodit míru fitness a základní schopnost algoritmu je narušena. Poté existuje možnost použít (stejně jako v přírodě) zpětné vazby prostředí, které umožní organismu rozpoznat, zda nová varieta jeho chování zajišťuje efektivitu a přiblížení se k *fitness* (Konar, 1999). V případě použití GA přímo v reálném prostředí (hypoteticky) vznikne problém ověřování nové variety v externím světě.

Musí dojít k expresi daného chování (v úrovni aktuální populace), což vede k nevratným procesům. Pokud bylo nové chování špatné (nevedlo přiblížení se k *fitness*), může tento krok vést k zániku organismu. Jelikož systémy, ve kterých jsou aplikovány genetické algoritmy, nemají schopnost se rozmnožovat, znamená to v praxi poškození či zánik systému po prvním použití nevhodného algoritmu. Tím se nám ukazuje výhoda interní reprezentace prostředí v úlohách, kdy neznáme cílový stav. Možnost provádět jednotlivé postupy stavovým prostorem virtuálně zamezí potřebě testovat každé řešení dané úlohy přímo v prostředí.



Obr. 12 *Diagram funkce genetického algoritmu*

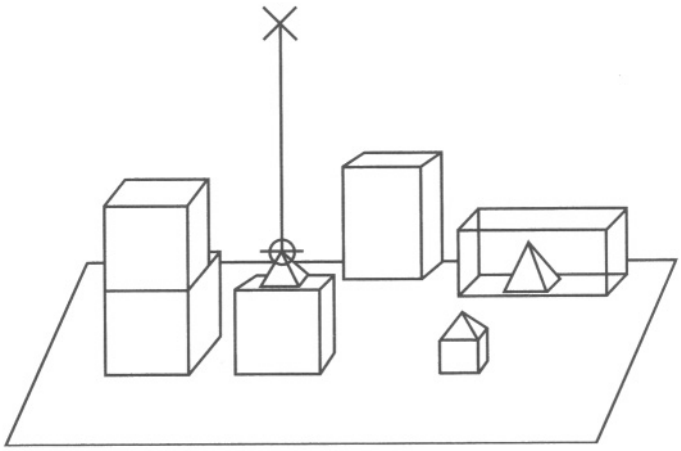
Výhodou genetických algoritmů je, že v sobě obsahují mechanismus zapomínání jako Hebbovské neuronové sítě (viz Hebbovské učení), což zajistí systému optimální využití paměti. V průběhu procesů dochází k odstraňování nevýhodných variant z paměti a systém uchovává jen relevantní a jemu blízká řešení, jejichž kombinací se snaží dojít k nejlepšímu řešení (Konar, 1999).

3.3.4 SHRDLU

To co charakterizuje období 70-tých let a co nabízí pokrok v oblasti obecné inteligence je koncepce *mikrosvěta (minisvěta)*. Vychází z akceptace teze, že komplexnost prostředí a pohybu jedince v něm není v oblasti napodobování nepřekonatelnou překážkou. *Mikrosvět* je zde brán jako doména, která může být zkoumána izolovaně. Koncepce implikuje, že přestože každá oblast diskurzu se zdá být otevřená pro zbytek lidských aktivit, její nekonečné větvení je jen zdánlivé a začne brzy konvergovat k uzavřené množině faktů a vztahů. Je možné provádět mnoho aktivit ve světě a je mnoho způsobů odezvy prostředí, ale celkový výčet je konečný, tvoří kategorie a je možné jej považovat za uzavřený systém. Lépe řečeno je možné koncipovat „svět", který obsahuje nejen simulaci jedince, ale i simulaci okolního prostředí a obě části dohromady tvoří uzavřený systém *mikrosvěta*.

Mezi klasické zástupce tohoto přístupu patří program SHRDLU. Jeho autor, Terry Winograd, popisuje svou práci termíny, které lze nalézt ve fyzice. „Zajímáme se o tvorbu formalismu nebo „reprezentaci", kterou se snažíme popsat …znalost. Hledáme „atomy" a „částice" ze kterých je postavena a „síly", které na ni působí" (Winograd, 1976, s.118). Pravdou je, že fyzikální teorie mohou být založeny na studiu relativně jednoduchých a izolovaných systémů a poté z těchto principů tvořit stále komplexnější modely a integrovat do nich další domény či fenomény. To je možné díky tomu, že je jich možno dosáhnout aplikací zákonů a vztahů na soubor základních elementů, které Papert a Minsky nazývali „strukturální primitiva".

Důležité je zmínit, jak vlastně program SHRDLU funguje. Jedná se o vytvoření světa bloků a geometrických primitiv, které jsou pouze virtuálně simulované strojem a prezentované například grafickou formou na monitoru. Tím je koncipován *mikrosvět*. V něm existuje inteligentní systém (součást programu), schopný primitivy manipulovat a mající „znalost" o jejich poloze. Uživatel se může pomocí komunikace v angličtině bavit se systémem o poloze předmětu, počtu objektů apod. Systém „chápe" svůj mikrosvět (je jeho součástí) a tak s ním dokáže inteligentně interagovat.



Obr. 13 *Grafické znázornění mikrosvěta SHRDLU*

Jestliže se nám podaří vytvořit inteligentní systém v této rovině, můžeme přenést výsledky a poznatky do reálného světa?

Winogradovy předpoklady, vycházejí z přístupů zkoumání neživé přírody. Pomocí SHRDLU chtěl ukázat, že jeho program bude schopen „rozšiřovat" koncept vlastnění i přesto, že základ programů spočívá v *mikro-teorii* vlastnění, která je založena právě na koncepci *mikrosvěta.*

Pokud ale na systém nebudeme nazírat jako na uzavřený systém, který obsahuje jak reprezentace světa, tak jeho prezentace, začnou se objevovat nesrovnalosti. Simon ve své analýze dospívá k závěru, že SHRDLU nerozumí a nechápe pojem vlastnění, jelikož nepracuje s významy. Dále pokračuje: „SHRDLU systém pracuje s problémy ve světě samostatných bloků s fixní reprezentací. Když dostane instrukci „vezmi velký červený blok", provede pouze asociaci termínu „vezmi" s příslušnou procedurou, poté identifikuje pomocí testů, co je asociováno pod „velký", „červený" a „blok", z čehož odvodí argumenty pro danou proceduru a provede řešení metodou řešení problémů (Haugeland, 1997,s.151). Winograd se vyhýbá problematice transdukce a následné reprezentace tím, že systém může kdykoliv použít prezentaci prostředí, jež vytváří, jako reprezentaci prostředí, kterou jako inteligentní systém zpracovává. Odpadají nám veškeré problémy a otázky týkající se vnímaní, jelikož jej nepotřebujeme. V reálném světě ale bez vnímání nedokážeme postulovat inteligentní systém.

Winogradovi odpůrci tvrdí, že množina propojených faktů může tvořit universum, doménu, skupinu atp., ale nemůže konstituovat svět. Svět je organizovaná masa objektů, důvodů, schopností a dovedností ve smyslu, kterým jim lidské aktivity přiřazují význam. Mohli bychom pokračovat, že i v dětském světe jsou mezi ostatními věcmi bloky, ale neexistuje zde něco jako blokový svět. A proto není možné souhlasit s Winogra-

dem, že pokud jeho program pracuje s „malým zlomkem světa", pracuje s mikrosvětem. Protože jím prezentované *mikrosvěty* nejsou světy. Není možné je kombinovat či rozšiřovat na svět našeho každodenního života. Neschopností odpovědět na otázku, co to svět vlastně je, znamenalo pro UI pět let stagnace (Haugeland, 1997,s. 153). Kritika se snaží nalézt slabinu přímo v základech modelování, které kromě zkoumaného jevu, vytvoří i jeho umělé prostředí.

3.3.5 Hry jako model i nástroj

Hry jsou oblastí, kterou průkopníci v oblasti UI rádi používali při svých simulacích. Výhodou her je přesně specifikovaný stavový prostor, základní pravidla, jejichž platnost je univerzální a omezený počet diskrétních stavů systému. Nehovoříme zde obecně o všech hrách (jako například fotbal, který nemá omezený a definovaný počet stavů) ale o hrách, založených na pravidlech logiky a kombinatoriky. Většina těchto her se nazývá **tvrdé systémy**, což vyplývá právě z přesného vymezení prostředí i pravidel. Nejčastější formou takových her jsou šachy, dáma, go apod. Tyto hry také patří do kategorie **her se sumou nula**. Což znamená, že proti sobě stojí dvě strany (hráči) a každý tah jedno hráče (diskrétní krok) znamená změnu v poměru sil. Hráč se svými tahy snaží získat profit odpovídající (rovnající se) ztrátě protihráče. Součet zisku a ztrát je v každém tahu roven nule (equilibriu). V libovolné fázi hry je systém tvořený protihráči v rovnovážném stavu. Což ale není hlavní výhodou, proč si tyto hry oblíbili průkopnící Umělé inteligence. Devízou je hlavně výše zmíněný tvrdý systém. Nemůže se objevit stav, který by nebyl odvoditelný ze základního nastavení a sady pravidel. Prostor je také přesně vymezen a diskrétní (plocha je dělena na pole). Poslední devizou je diskrétní rozdělení času na kroky, ve kterých probíhají změny. Takto definovaný systém se jeví jako naprosto ideální pro simulaci klasickou výpočetní architekturou. Snad právě proto je oblast šachových a jím podobných programů první oblastí simulace, kde počítač „předehnal" člověka ve schopnostech „být inteligentní". Důvody jsou právě výše zmíněné vlastnosti „tvrdých" her.
Simulace her však nabízí i jinou možnost uplatnění na poli Umělé inteligence.
Alternativním postupem je použití převráceného principu současných počítačových her. Nejprve však musíme zmínit neinvertovaný princip. Počítač má za úkol vytvořit trojrozměrný svět, ve kterém se hráč pohybuje. Tento svět nemusí být reálný, ale většinou obsahuje fyzikální zákony reálného světa i jeho design je kopií základních geometrických objektů. V určitém typu hry (RPG), je úlohou hráče zorientovat se v daném

prostředí a nacházet smysl, kterým toto prostředí funguje, nalézt své postavení v něm, a možné akce, vedoucí k lepšímu pochopení prostředí, popřípadě k restrukturaci. Požadavkem na uživatele je adaptace na prostředí. Výhodou hráče je, že se jedná o inteligentní bytost, která pouze aplikuje a modifikuje své zkušenosti z reálného světa do světa modelu. Zkusme se ale na hru podívat trochu jinak. Můžeme posunout počítač do role hráče.Trojrozměrné prostředí, které počítač konstruuje pro hráče se dá použít jako interní reprezentace okolního hypotetického světa pro počítač. Získáváme tak reprezentaci v obrazové (ikonické) podobě, a přitom její zpracování, uchování a manipulace je ve výpočetních schopnostech stroje, na kterém simulace běží. V této rovině se ještě stále ocitáme ve komplexnější variantě Winogradova SHRDLU.

Můžeme se ale posunout o krok dále, a snažit se vytvořit interní reprezentace reálného nehypotetického prostředí. Z povahy informací (obrazový kód) budeme mít o něco jednodušší pozici než u klasického symbolického přístupu. Za součinnosti kamer a senzorů je možné získat materiál dostatečný na to, aby počítač dokázal vytvořit trojrozměrný model prostředí ve kterém se pohybuje (určitě s jistou mírou degradace komplexity) a v reálném čase umístit sebe do reprezentace prostředí. Nyní máme k dispozici dvě reprezentace prostředí a také situovaný systém. Vylepšení je ve schopnosti systémů paralelně vytvářet hypotetické mentální reprezentace (vnitřní systém) ve formě, kterou dokáže plně „pochopit" a zpracování informace z prostředí ze senzorů a kamer) sloužící k upřesnění metrik a poloh předmětů v reálném prostředí, pro potřeby práce vnitřního systému. Laicky řečeno, tento systém vychází vstříc informacím z venku tím, že je nejen zpracovává, ale ještě si z vlastních „představ" vytvoří přibližný model toho co je a bude vnímáno.

Čímž se dostáváme za hranice směru Umělé inteligence, který s se nazývá *Artificial Life (A-Life).* Zabývá se simulací agenta i prostředí na jedné platformě. Nejčastěji bývá využíván k simulaci dynamických jevů (kooperace agentů ve virtuálním světě, evoluce organismu apod.). Bohužel si zachovává stejnou nevýhodu jako i zmíněný Winogradův SHRDLU. Kromě zmíněného redukcionistického modelu světa, je to nulová reference vytvořeného prostředí k reálnému světu. Některá omezení týkající se umělého života *(A-Life)* byla zmíněna v kritice Winogradova programu SHRDLU.

Vraťme se tedy k původní myšlence. Díky adekvátnímu modelu prostředí a poloze systému v něm splníme požadavky, uplatňované při konstrukci autonomních agentů, situovanost a orientovanost. Otázkou zůstává, jestli agent bude využívat při svém pobytu v prostředí, poznatků ze svých kamer a senzorů, nebo bude brát jako „reálné" prostředí právě svůj interní model tohoto prostředí. U člověka je daná otázka námětem

pro dlouholeté filosofické spory. Počátky se dají vidět v platónské ideji jeskyně (otázka „Vnímáme svět nebo jeho obraz?"). Ať je odpověď na tuto otázku jakákoliv, přístupem tvorby modelu okolí agenta v obrazové (ikonické) formě, získáváme mentální reprezentaci(e) okolí, které v sobě obsahují metrické a topologické uspořádání prostoru, kterého není jazyková (symbolická) reprezentace schopná. To nás posouvá trochu dále v oblasti napodobování inteligence (přesněji v jednom ze způsobů, jak se dá simulovat vnímání).

Další kroky budou obtížnější. Pokud se vrátíme k naší metafoře počítače jako hráče, hrajícího hru zvanou svět, dostáváme se situace , ve které je na tom lépe lidský hráč. Jde o orientace a prozkoumávání prostředí, identifikace předmětů, , tvorba a způsoby adaptace. Orientaci jsme vyřešili modelováním prostředí korespondující s reálným prostředím pomocí součinnosti dvou nezávislých reprezentačních systémů. Prozkoumávání prostředí souvisí z části s motivy, které počítač má (některé základní musí být zakódovány natvrdo, stejně jako u člověka rozmnožování, příjem potravy, orientačně pátrací reflex a jiné). Předpokládáme, že budou následně emergovat v cíle, jež jsou kombinacemi základních cílů a které jsou také determinovány zpětnou vazbou z prostředí, v našem případě změnami ve vnitřním 3D modelu prostředí.

V takto fungujícím systému se značně zjednodušuje identifikace předmětů (proces rozpoznávání a kategorizace). Pokud počítač svou reprezentaci neustále generuje v reálném čase, generuje všechny předměty v jeho okolí a ty jsou přesně identifikovatelné. Což platí pro prostory a předměty, které mají optimální velikost pro jeho senzorický systém (zde narážíme na omezení rozlišovacích schopností vstupu z kamer a senzorů). Jestliže se objeví objekty, které svou komplexností převyšují jeho rozlišovací schopnosti, nastává problém z jejich převodem na úroveň modelu.

Další oblastí, která nám nabízí zlepšení oproti symbolickému přístupu, je problematika ukotvení významu. Pokud budeme vycházet z Harnadových prací, splňuje zmíněný způsob reprezentace podmínku pro ukotvení významu (interní reprezentace musí obsahovat ikonickou formu). Požadavek na ukotvení významů je ale splněn pouze v případě, že budeme postulovat (v případě komputační architektury) speciální počítačový modul (program) vědomí, o kterém si povíme níže, a který by dokázal přijímat výsledky modelování vnějšího prostředí ve formě kom-

plexního obrazu. Tím, že používáme k simulaci klasický počítač se všemi nevýhodami jeho architektury, můžeme dojít k tvrzení, že model prostředí není uložen v ikonické formě, ale pouze ve formě čísel (symbolů), které určují polohu a velikost jednotlivých objektů v prostoru. Způsob uložení v paměti je následně symbolický a ikonická podoba vzniká až samotným převodem dat do podoby modelu. Druhým požadavkem Harnada na ukotvení je, kromě ikonické, také kategoriální reprezentace. Způsobů, které vedou k přiřazování do kategorií je několik. V našem případě se pro splnění Harnadovy podmínky nejvíce hodí přístup, který je založen na tvorbě prototypu. Každý objekt modelu by musel obsahovat ještě svou referenci v podobě prototypu, definující limitami svých vlastností a rozměrů skupinu, do které daný objekt modelu patří. Při dostatečných výpočetních možnostech stroje je podmínka splnitelná.

Poslední problematickou oblastí při metafoře počítačového hráče je otázka vědomí. Při jejím objasňování se dostáváme do oblasti, ve kterém jsou lidské vysvětlovací principy velmi obtížně aplikovatelné. Při popisování způsobu simulace založené na konstruování vnitřní trojrozměrné reprezentace je v pozadí obsažen předpoklad homunkula.

Homunkulus

V oblasti filosofie mysli patří k jednomu ze stěžejních témat Ryleho „Ghost in machine". Jestliže se snažíme vytvořit umělý inteligentní systém, setkáváme se otázkami týkající se jeho aktivní, kontrolní komponenty. Možností je postulovat „živého" agenta v hlavě (homunkula), který ovládá kognitivní aparát a supluje tak vědomí. Z čehož vyplývá, že inteligentní záměrné a motivované chování potřebuje systém, který je živý a má přístup ke všem funkcím a částem systému a může je reorganizovat s určitým stupněm volnosti (Luger, 1994). Problematika homunkula se objevuje i ve spojitosti s ukotvením symbolů. Dá se říci, že systém používající pro svou činnost neukotvené symboly, vždy bude potřebovat homunkula pro své vysvětlení.

Tedy, že po konstrukci modelu prostředí a všech aspektů pobytu v něm, předpokládáme něco, co se na celou věc dívá a rozumí jí, aktivní komponentu, které je přístupný dostatek informací o interních procesech i reprezentaci. V našem případě lze

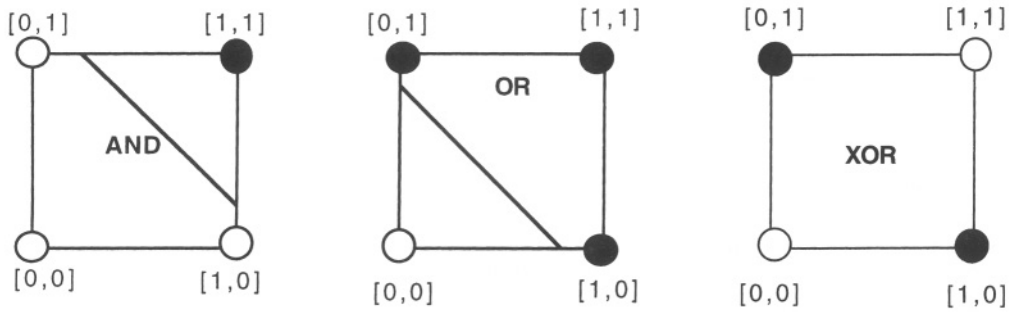
tento požadavek obejít pouze postulací programu vědomí, samostatně spuštěného uvnitř počítače (či na jiném stroji, který je propojen se strojem simulujícím prostředí). Jeho oblasti působení jsou příliš komplexní na výčet a příliš spekulativní na přesnou deskripci, takže zůstává pouze v úrovni myšleného. Jeho základní vlastnosti lze shrnout jako schopnost, mít přístup k libovolným procesům a obsahům, které vedou právě ke konstrukci tohoto platónsky virtuálního světa uvnitř stroje.

4 KONEKCIONISMUS

4.1 Neuronové sítě

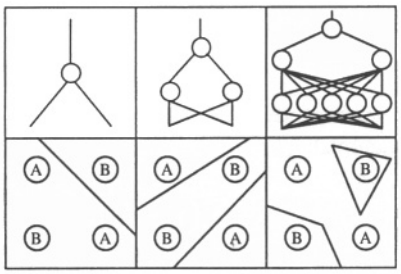
V 80-tých letech se začíná objevovat nová technika modelování, či přesněji nová architektura. Pro svou fundamentální odlišnost od předchozích způsobů se tento přístup stává paradigmatem v oblasti kognitivních věd. Přístup se nazývá konekcionismus a jeho aplikovanou oblastí jsou neuronové sítě, architektura principielně odlišná od předchozích přístupů. Základ tvoří samostatné velmi jednoduché jednotky, jejichž propojení konstituuje sítě. Největší rozdíl oproti klasické von neumannovské architektury spočívá v paralelním způsobu práce. Mnoho vědců z oblasti UI a kognitivních věd daný přístup přivítalo, protože umožňuje alternativně řešit problémy, jež se u předchozího přístupu setkávaly s obtížemi. Výchozí myšlenkou je postulace základní jednotky, funkčně podobné lidskému neuronu.

Dvacet let předtím (v roce 1958) publikoval Rosenblatt práci o perceptronu. Již u něj bylo použito architektury založené na modelu neuronu a váhových propojení (Rosenblatt, 1958). Základní myšlenka perceptronu ale nebyla uvedena do praxe. Způsobily to připomínky Minského a Paperta , kteří matematicky dokázali (Minsky, 1969), že základní forma perceptronu není schopná simulovat některé logické operátory (přesněji XOR, což souvisí s problematikou lineární separovatelnosti).



Obr. 14 *Funkce AND, OR, XOR a možnosti jejich lineární separovatelnosti*

Publikování práce znamenalo pokles zájmu o konekcionismus, jako o perspektivní obor. Na konci 80-tých let ale přichází období stagnace klasického symbolickém přístupu v oblasti modelování, hledají se alternativní metody a konekcionismus prožívá svou renesanci (Pfeifer&Scheier, 2001). Důležitým je i fakt, že Minský opravil svá tvrzení o omezenosti neuronových sítí jako architektury vhodné pro simulaci (funkce XOR je řešitelná neuronovou sítí za použití více vrstev).



Obr. 15 *Možnost linearní separovatelnosti pomocí a) jednoho neuronu b) dvouvrstvé sítě c) vícevrstvé sítě*

Bylo by zbytečné uvádět zde popis funkce neuronu (jak živého, tak umělého), jelikož bývá v každé základní učebnice psychologie. Méně častá ale již bývá základní charakteristika neuronových sítí (Caudill&Buttler, 2000).

1. Jsou tvořeny počtem jednoduchých procesních jednotek, komunikujících přes množinu propojení, které mají různou váhu a sílu.
2. Paměť je reprezentována jako vzorec hodnot vah, mající propojení mezí jed notlivými prvky. Informace je zpracovávána jako šíření se měnících se vzorců aktivity mezi prvky.
3. Sítě jsou spíše učeny a trénovány než programovány.
4. Místo oddělené paměti, procesoru a externímu programu, který řídí operace systému jako u digitálního počítače, operace neuronových sítí jsou implicitně kontrolovány třemi vlastnostmi: kombinační funkcí neuronu, způsobem propojení a učícím pravidlem
5. Neuronové sítě jsou přirozené asociační paměti
6. Neuronové sítě jsou schopny generalizace; mohou se naučit charakteristiky becné kategorie
7. Jsou odolné proti chybám. Díky paralelní distribuované formě uložení pamětí „degradují s grácií"
8. Neuronové sítě mají schopnost sebeorganizace. Dokáží reagovat na vstupy z prostředí změnou své funkční dynamiky
9. Neuronové sítě jsou schopné emergence nových vlastností či chování.

To co výzkumníky nejvíce lákalo na neuronových sítích byly jejich dvě základní vlastnosti. Za prvé, že jsou sítě schopné se učit a za druhé, že mají schopnost emergence. Emergence v tom smyslu, že systém vykazuje výstupy, které nebyly předpro-

gramovány. Chování je výsledkem interakce jednotlivých komponent systému (Pfeifer&Scheier, 2001).Hlavním důvodem pro vznik emergentního chování je redundance v propojení. Klasická sériová architektura obsahuje pouze propojení, která jsou nutná mezi jednotlivými jednotkami. I proces zpracování informace v takovém systému je pevně zakotven designérem systému . Naopak neuronové sítě jsou typ architektury, která je hned v počátcích připravená na více způsobů implementace úlohy. Pokud je jeden neuron propojen se všemi neurony vedlejší vrstvy a pro učení jsou používány jen některá propojení pro funkci systému, v klasické von neumannovské architektuře by to vedlo k odstranění těch spojení, která nejsou používaná. Vznikl by systém, který není použitelný obecně, ale pouze specificky (ne z hlediska zpracování úlohy, ale architektury) . Rozdíl ve způsobu propojení komponent je, že pokud se objeví nový stimul, který má neuronová síť zpracovat, mohou být používány spojení, které byly v předchozím případě nevyužita. Což nás vede k závěru, že redundance je nutnou podmínkou emergence (více v samostatné pasáži o emergenci).

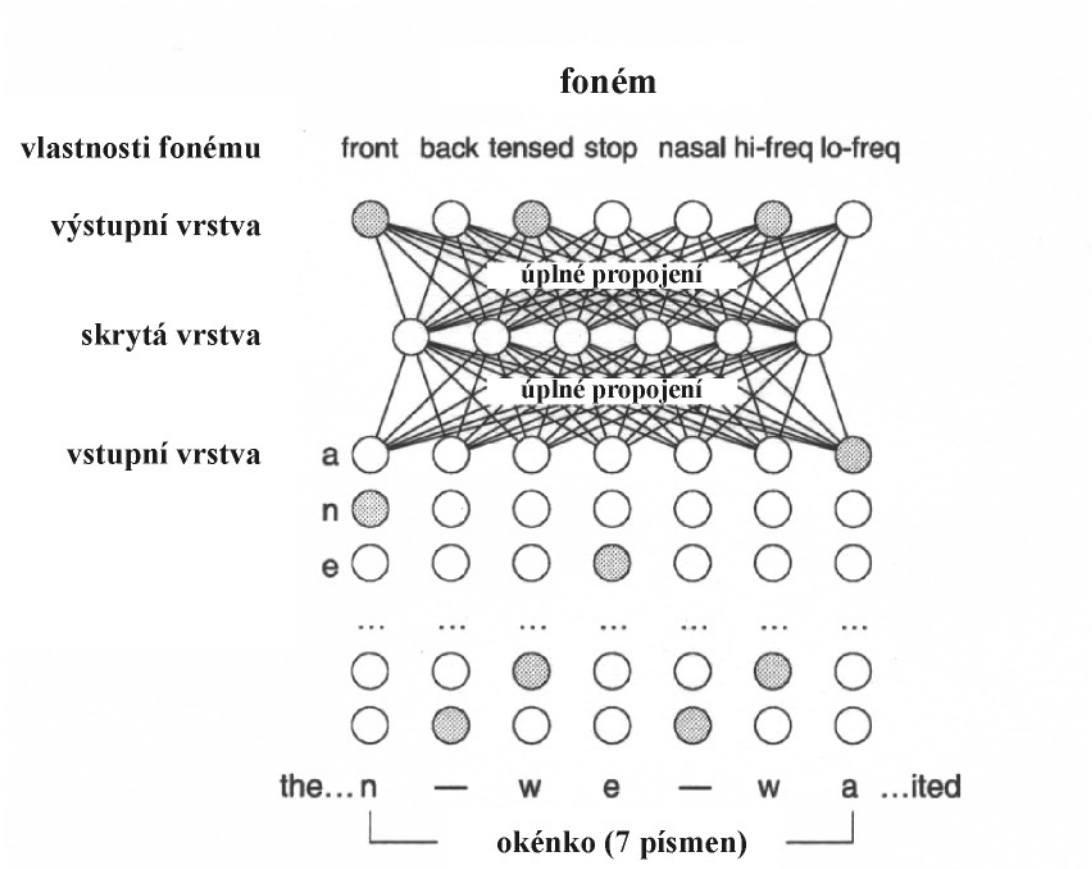
4.1.1 Způsoby učení

Jak bylo zmíněno výše, jsou neuronové sítě velmi dobře použitelnou architekturou pro modelování či simulaci různých typů učení. Automatické neuronové sítě (ANN) podporují oba typy učení (kontrolované i nekontrolované). Kontrolované *(supervised)* typy učení se používají při aplikaci v oblastech kontroly, automatizace, robotiky a počítačového vidění. Nekontrolované *(unsupervised)* učení se používá při plánování, osvojování si zkušeností (akvizice) a při převodu analogového do digitálního kódu. Je samozřejmé, že tyto oblasti jsou zaměnitelné, protože se nejedná o komplementární kategorie, ale o zdokonalování principu kontrolovaného učení použitím poznatků a strategií z oblasti nekontrolovaného. Klíčový rozdíl je právě v přítomnosti interpretátora dat (člověka) u prvního typu a jeho absenci u druhé formy (Konar, 1999). V české terminologii se pro *supervised* a *unsupervised learning* také objevuje varianta učení učitelem a bez učitele.

Při kontrolovaném (supervised) učení musíme znát správný výstup, abychom jej použili jako korekci. Cílem je najít odpovídající mapování pro skrytou vrstvu. Rozhodujícím činitelem je rozdíl mezi skutečným a požadovaným výstupem. Hodnoty rozdílu slouží k upravení vah ve skryté vrstvě. Nejčastěji se používá metody *back-propagation* (Pfeifer&Scheier, 2001).

Dobrý příkladem pro kontrolované učení je projekt NETTalk. Jedná se o neuronovou

síť spojenou s řečovým syntetizérem, která je schopná převádět psaný anglický text do mluvené podoby. Práce sítě končí přiřazením fonému, jehož expresi již provádí samostatný hlasový syntetizér. Síť se skládá ze tří vrstev. Vstupní vrstva je navržena velmi robustně. V sedmi slotech na písmena je zpracováváno najednou sedm následných znaků textu, který má být převeden do řečové podoby. Výhodou tvorby tohoto textového „okénka" je možnost paralelního zpracování bloku textu, jehož prezentace a zpracování v jeden okamžik umožní zachycení vazeb mezi jednotlivými variantami výskytu znaků. Ty jsou důležitým vodítkem pro diskriminaci a následné přiřazení správných fonému v závislosti na sedmiznakovém „minikontextu". Každý slot (pro jeden znak) má 29 neuronů, reprezentujících celou abecedu plus mezeru a znaménka. Vstupní vrstva obsahuje 29x7 neuronů. Ve skryté vrstvě pak dochází díky robustnosti vstupní vrstvy ke kvalitnímu rozlišení dané posloupnosti znaků a poté ke kategorizaci vedoucí k přiřazení fonémů ve výstupní vrstvě. Učení je zde prováděno *backpropagation* algoritmem.

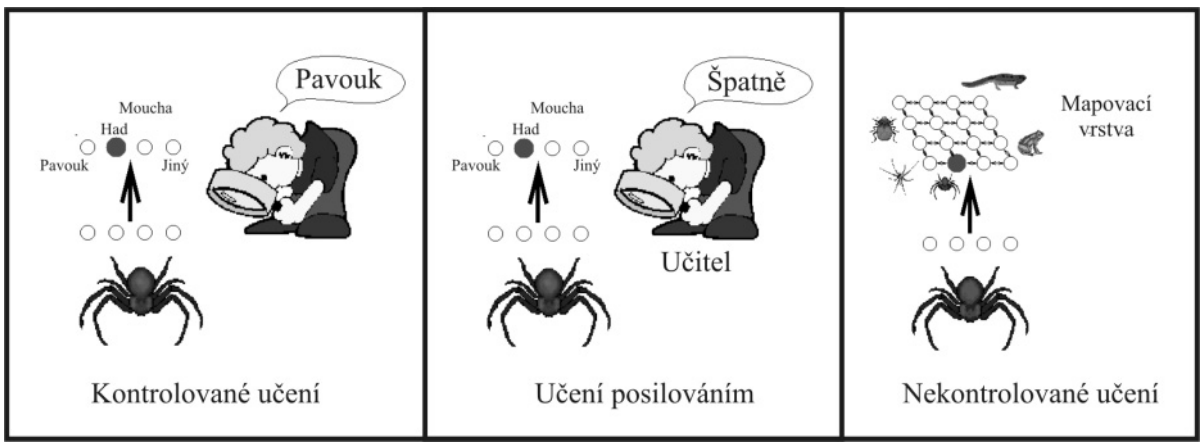


Obr. 16 *Struktura a funkce sítě NETTalk při zpracování anglického textu*

Zde jsou některé postřehy autorů zachycující průběh aktivace. Při učení síť postupovala podobně jako probíhá řečový vývoj u dítěte. Nejprve se naučila rozlišovat samohlásky od souhlásek a její řeč připomínala dětské žvatlání. Přes fázi neumělé a nepřesné výslovnosti malých dětí se vypracovala až na 95 procentní úspěšnost (Pfeifer&Scheier, 2001). Výhodou sítě je, že v sobě obsahuje možnost tvorby kontextu v rámci jednoho slova (díky sedmi slotům na písmena) a také paralelní zpracování, protože každý ze slotů má k dispozici svou sadu znaků. Nevýhodou zůstává kontrolované učení. Síť není autonomní a kontext nesouvisí se sémantickým významem zpracovávaných slov, ale slouží pouze pro správnou výslovnost.

Dalším typem kontrolovaného učení je **učení posilováním.** Vychází z klasické psychologické teorie o trestu a odměně.V praxi to znamená, že systému, který se učí, je dána pouze zpětná vazba o tom, zda provedl úlohu správně (ano/ne). Díky tomu, že se jedná vždy pouze o užití jednoho typu (odměna nebo trest), je přítomen pouze jeden druh signálu. Tím, že je výsledek označen například jako negativní (trest), provede systém buď korekci svého algoritmu či vah sítě. Záleží na něm, jakým způsobem (v této fázi učitel nezasahuje). Učení posilováním leží spíše na pomezí mezi kontrolovaným a nekontrolovaným učením. Používá se jak v oblasti neuronových sítí, tak v oblasti strojového učení (Pfeifer&Scheier, 2001).

V přírodě se nesetkáváme jen s přístupem, kdy je naše efektivnost či adaptabilita určována pouze okolím (společností), ale i hodnocením a změnou nastavení, u kterého je kritériem naše vlastní historie. Pokud se funkce sítě upravuje (adaptuje) na základě vnitřních hodnocení svého chování a nejen pouze externím kritériem, hovoříme o samoorganizaci (Mařík, 1993). Neuronové sítě s takovými možnostmi jsou schopné **nekontrolovaného učení** (učení bez učitele).



Obr. 17 *Typy učení neuronových sítí*

4.1.2 Paměť neuronových sítí

Rozhodnout, zda neuron obsahuje nějakou paměť, záleží na způsobu nazírání. Pokud je hodnota potenciálu uvnitř neuronu brána jako paměťová informace, neuron paměť obsahuje. Také je možno vidět v této hodnotě stav systému. Nejde už o reprezentaci (paměť), ale spíše prezentaci a neuron žádnou paměť neobsahuje. Bereme-li, že je neuron používán jako médium (například uzel sémantické sítě), do kterého ukládáme obraz prostředí, je mu přiřazen status paměti (což předpokládá velmi nestandardní typ neuronu). Nejčastěji hovoříme o celé síti neuronů, ve které je paměť obsažena v distribuované formě.

Během posledního desetiletí vzrostl zájem o konekcionistické modely paměti. Vychází to již z možností základních prvků. Propojíme-li totiž dva neurony obousměrně (O=O), můžou implementovat určitý typ paměti (Pfeifer&Scheier, 2001). V celkovém kontextu neuronových sítích je jednotka paměti reprezentována jako vzorec vzruchů mezi velkým množstvím neurodů (neuronů). Množina těchto hodnot aktivity (vah) je zobrazitelná jako vektor v multidimenzionálním prostoru, přičemž počet dimenzí odpovídá počtu hodnot. Interakce mezi jednotkami způsobují ukládání paměti.

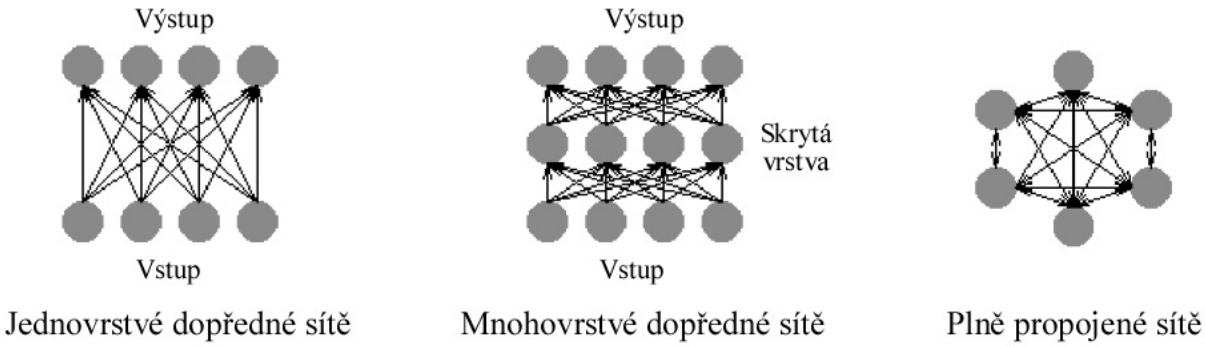
Konekcionistické modely paměti se dají rozčlenit do tří základních skupin. První je tvořena vícevrstvými dopřednými sítěmi pro rekognici a kategorizaci. Druhou tvoří autoasociativní sítě pro rekognici a rozpoznáváni vzorů a třetí rekurentní heteroasociativní sítě paměťových sekvencí (Sternberg, 1999). To nás posouvá ke kapitole, která uvádí obecné třídění neuronových sítí s přihlédnutím k jejich paměťovým možnostem.

Rozdíly neuronových sítí a biologických mozků při simulaci	
	1.Komplexnost
	2.Lidské schopnosti nejsou arbitrární, ale pevně kódované.
	3.Není znám přesný způsob excitace či inhibice.
	4.Model učení z chyb není použitelný obecně.
	(Kosslyn, 1992)

Tab. 2 *Rozdíly neuronových sítí a biologických mozků při simulaci*

4.2 Typy neuronových sítí

Neuronové sítě bývají nejčastěji rozděleny podle způsobu svého propojení. Jestliže propojovací matice (viz níže) obsahuje hodnoty 0 v diagonále a nad ní, jedna se o dopředné *(feedforward)* sítě, protože obsahuje pouze dopředné propojení, tedy propojení v jednom směru neobsahující zpětnovazebné smyčky. Sítě obsahující několik vrstev propojených pouze dopředně se nazývají vícevrstvé dopředné sítě. Je-li každý neuron v jedné vrstvě sítě propojen s každým neuronem následující vrstvy sítě pouze jednosměrně, nazýváme tuto síť plně propojenou. Sítě, kde jsou všechny neurony propojeny s ostatními obousměrně se nazývají Hopfieldovy sítě (Pfeifer&Scheier, 2001).



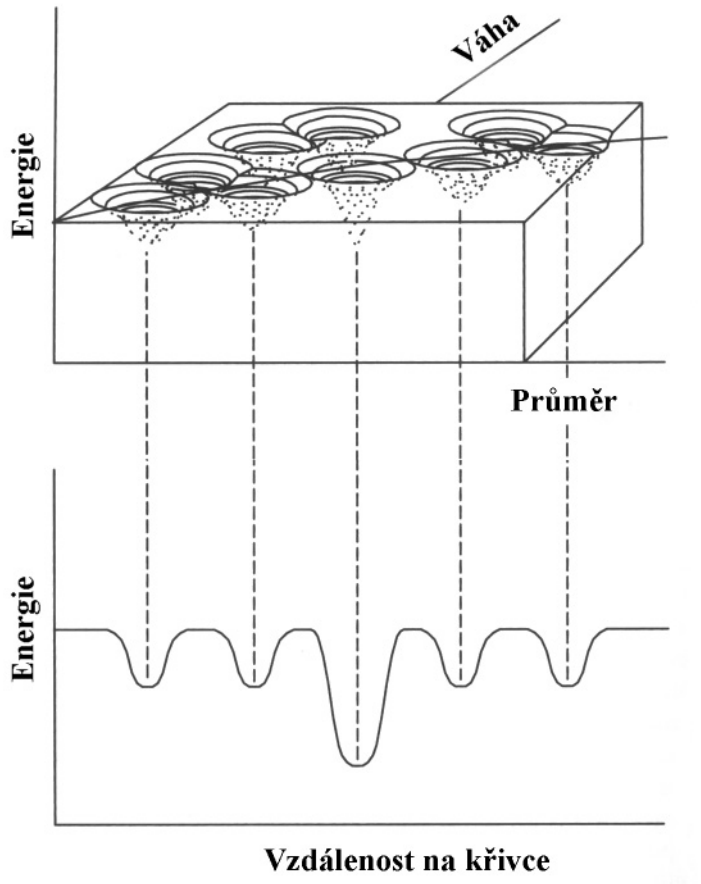
Obr. 18 *Některé základní typy neuronových sítí*

4.2.1 Hopfieldovy sítě

Hopfieldova neuronová síť, navržená začátkem osmdesátých let, je typickým příkladem autoasociativní sítě. Během vývoje J. Hopfield rozpracoval koncept energetické funkce, která má zásadní význam pro správnou funkci sítě a z ní jsou odvozena pravidla pro učení a vybavování. V současné době existuje několik modifikací sítě. Může být použita jako asociativní paměť, klasifikátor (kategorizace) nebo k řešení optimalizačních problémů.

Energetická funkce nám umožňuje lépe pochopit chování Hopfieldovy sítě a speciálně pak princip učení a vybavování. Pro názornost si ji můžeme představit jako trojrozměrnou scénu, kde funkce představuje krajinu s údolími a kopci. Údolí pak představují již naučené vzory. Hopfieldově síti při učení předkládáme pouze trénovací vzor. Při vybavování neznámého vzoru, který chceme pomocí Hopfieldovy sítě iden-

tifikovat, jej budeme reprezentovat kuličkou, pohybující se po pomyslné krajině a snažící se dostat na co nejníže položené místo. Může se však dostat i do lokálního minima, které také představuje řešení, avšak nikoliv řešení optimální (kulička zůstane v údolí, které neodpovídá její velikosti). Z tohoto důvodu bude energetická funkce (rozlišovací schopnost pro hledání minima) sítě poněkud složitější. Potřebujeme totiž, aby obsahovala vše, co popisuje její chování. Budeme chtít, aby její hodnota byla velká pro velké chyby (kopce) a malá pro malé chyby (údolí). Problematika se také nazývá *annealing* a bývá řešena Boltzmannovým strojem.

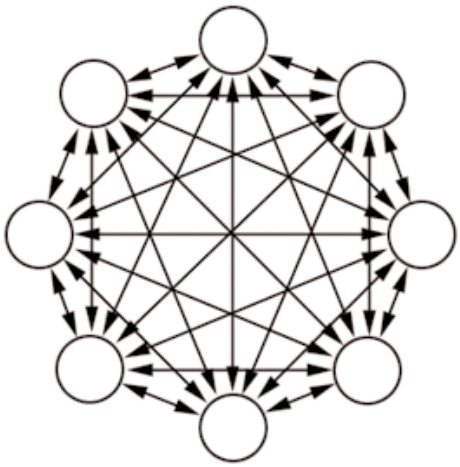


Obr. 19 *Energetická funkce*

Asociativnost hopfieldovské paměti je dána tím, že vybavovaný vzor zadáváme jistou jeho částí, do značné míry libovolnou. Vzor je vybavován podle části svého „obsahu", tj. prostorového umístění, nikoliv odkazem na nějakou jeho „adresu" (jako u paměti RAM). Jedná se o *content adressable memory* (CAM), což je u neuronových sítí používaných jako paměťový systém běžné.

Pro rozsáhlejší (Hopfieldovy) sítě je výhodnější reprezentovat je pomocí matice. Jedná se pouze o způsob vyjádření výhodný pro člověka, jelikož je přehlednější pro čtení. Fyzicky mohou být neurony umístěny libovolně v prostoru, pouze jejich propojení si

zachovávají při převodu do ortogonální soustavy horizontální a vertikální seřazení. Tento způsob vyjádření se nazývá **propojovací matice** (Pfeifer&Scheier, 2001).



Obr. 20 *Schéma Hopfieldovy sítě*

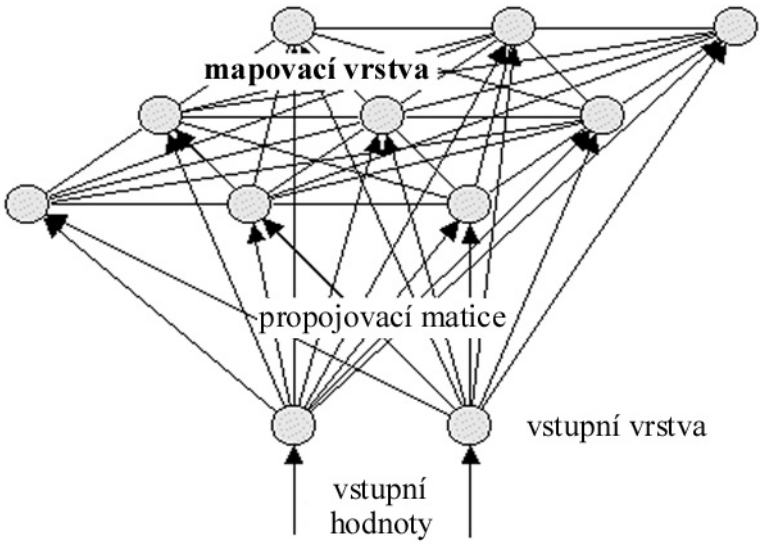
Je-li Hopfieldova síť použita jako asociativní paměť, má dvě hlavní omezení. Prvním je, že počet vzorů, které můžeme síť naučit, je poměrně nízký. Jestliže naučíme síť příliš mnoho vzorů, může síť konvergovat k nějakému zvláštnímu obrazci, na který nebyla naučena. Síť je potom přeučená.
Nevýhodou Hopfieldovy sítě jsou také veliké nároky na paměť. To může způsobit chybu při identifikaci předloženého vzoru. Další důležitou vlastností Hopfieldovy sítě, obtížně pojmenovatelné jako výhoda či nevýhoda, je automatické rozpoznávání inverzních vzorů k již naučeným vzorům. Hopfieldovu síť tedy nemusíme inverzní vzory učit.

4.2.2 Kohonenovy sítě

Jestliže tvoří Hopfieldovy sítě způsob paměťového systému blízký lidskému, rozšiřuje Kohonenův model schopnosti neuronových sítí ještě blíže směrem k lidské inteligenci. Jím navržená síť obsahuje schopnosti, které bychom v psychologii nazvali: exaktnější kategorizace, zapomínání, selektivita a díky mapovací vrstvě, informace o funkci systému s možnosti jejich využití při kontrole či zpětné vazbě.
Často potřebujeme pro správnou kategorizaci získat několik jeho vlastností. Zjištěné vlastnosti pro určitý vzor nám vytvářejí tzv. vektor vlastností. Dimenze neboli délka vektoru je určena počtem měřených vlastností N a odpovídá dimenzi prostoru, ve

kterém provádíme kategorizaci. To znamená využití všech předností neuronových sítí. Pokud síť obsahuje mnoho vstupních neuronů, přičemž každý je citlivý na určitou vlastnost (která ale musí být převedena do číselné formy nejprve člověkem), vzniká v rozhodovací vrstvě mnoharozměrný vektor vah, jehož hodnota slouží jako kriterium kategorizace. Silným nástrojem kategorizace jsou Kohonenovy sítě.

Základem Kohonenových sítí je plné propojení vstupní vrstvy s mapovací vrstvou. Tato vrstva má laterální propojení se všemi neurony své vrstvy. Pro nebližší neurony jsou propojení posilující), pro vzdálené inhibiční při zachování topologie. V praxi to znamená, že jsou blízké body ve vstupním prostoru mapovány na blízké body ve výstupním prostoru. Již zmíněnou vlastností Kohonenových sítí je schopnost Hebbovského učení. Díky tomu se sítě dají považovat jako neurobiologicky plausibilní (Pfeifer&Scheier, 2001).



Obr. 21 *Způsob propojení mapovací vrstvy u Kohonenových sítích*

4.2.2.1 Hebbovské učení

Základní pravidlo kanadského neurobiologa Donalda Hebba říká, že pokud se nachází neuron A v dostatečné blízkosti, aby excitoval neuron B a opakovaně pálí, tak v obou neuronech nastanou metabolické procesy, které zvýší pálení. A podporuje pálení B, je-li dostatečně blízko.

Možnosti biologických sítí se dají převést i do jejich umělých variant. Základní princip Hebbovského učení vychází ze zvyšování hodnot vah neuronů, které jsou spolu propojeny a během učení jsou aktivovány společně. V neuronových sítích však jeho

použití může vést k růstu vah kooperujících neuronů nad limity. Proto je nutné omezení na maximální délku společného jednotkového vektoru (růst jednoho na úkor druhého, přičemž součet vah nemůže překročit fixní hodnotu). Užívá se k tomu metody „Použij nebo ztrať" a je nutná přítomnost excitační i inhibiční váhy od –1 do +1 (Caudill&Buttler, 2000,s. 147). Hebbovské učení je možné rozšířit o mechanismus zapomínání, kdy se při absenci aktivace či koaktivace neuronů snižuje hodnota jejich vah až k „úplnému zapomnění".
Koukolík uvádí složitější variantu Hebbovského učení, kdy neurony mají schopnost se koaktivovat podle vzorce pálení. Nestačí pouhá aktivita a blízkost v prostoru, ale nutná je i znalost vzorce. Využití pulzních neuronů v oblasti Hebbovského učení rozšiřuje možnosti koaktivace i mezi vzdálenými neurony. Kriteriem totiž není pouze míra aktivity, ale i její způsob. V umělých neuronových sítích se však zřídka používá modelu neuronu založeného na pulzním principu, který se vyskytuje u živých neuronů (Caudill&Buttler, 2000). Ojedinělým modelem využívající této vlastnosti je způsob uložení paměti *Memory surface* (viz níže).

4.3 Redundance

Pojem redundance se v textu této práce již několikrát objevil. Jeho význam je celkem jasný a většinou vyplývá z kontextu použití. Podívejme se ale, kde nalezneme prapůvod potřeby redundance. Pojem je obsažen již v základní práci informačního paradigmatu, v knize Shannona a Weavera, týkající se informační teorie. Definují redundanci jako „zlomek struktury zprávy, která není podmíněna volbou odesílatele, ale spíše uznávanými statistickými pravidly ovládajícími volbu symbolů braných v potaz" (Shannon&Weaver, 1948). Názorným příkladem může být přenos zprávy v češtině, která je omezena možnými kombinacemi písmen tvořícími slova, která jsou výrazy českého jazyka. Důležité je také vzít v potaz, že některá slova jsou častější než ostatní. Tyto omezení nám jako příjemci umožňují pochopení přenášených českých slov, i pokud by některá písmena ve slově chyběla. V takovém případě redundance obsažená v jazyce vyplývá z omezení kombinace písmen (Ashby, 1956). Což platí pouze v případě, že vezmeme jednotlivá slova (omezená daným smyslem v jazyce) jako výchozí situaci. Tehdy je redundance brána jako možnost vynechání určitého písmene ve slově při zachování smyslu pro zkušeného uživatele jazyka (nezkušený uživatel tuto schopnost ztrácí). Takto se spíše ale definuje informační šum (o který šlo Shannonovi v jeho práci) než redundance (Pfeifer&Scheier, 2001). Také vznikají prob-

lémy při tvorbě nových slov, které uživatel jazyka nezná, a také u slov, která jsou tvořena jedním písmenem (spojka a). Takto definovaná jazyková redundance je spíše přesah za omezení slov, které mají význam. Jazyk je redundantní právě ve své schopnosti být arbitrárním systémem bez významu, jehož nové kombinace se mohou stávat označeními pro jevy v prostředí.

4.4 Robustnost

Stejně jako redundance je i tento pojem považován za nutnou podmínkou při tvorbě inteligentního systému. Vyskytuje se u biologických systémů v podobě jedné, evolucí nejsilněji podporované vlastnosti. U umělých systémů se snaha o tvorbu robustních architektur setkává z mnoha technickými obtížemi.

Pro toto slovo můžeme najít mnoho synonym jako pružnost, houževnatost apod. Jeho hlavním cílem je obrana proti zhroucení i pokud je část systému poškozená, nebo pokud jsou podmínky pro uskutečnění úlohy nedostačující. Příkladem (poněkud zjednodušujícím) je rýma u člověka, zabraňující dýchání nosem. Robustnost systému znamená možnost dýchání pusou. Slovo je v podstatě velmi blízké pojmu redundance. V oblasti inteligence se spíše mluví o robustnosti mozkové neuronové sítě, mající takové množství propojení a funkčních jednotek (neuronů), že při i výpadku většího počtu neuronů či propojení nedojde k vážnějšímu porušení funkčnosti systému. Pokud jsou některé oblasti trvale zničeny, může dojít k novým propojením, která „mrtvou oblast" obchází. Kvůli těmto vlastnostem se o mozku říká, že „degraduje s grácií".

Radikální přístup v oblasti simulace inteligence vycházel z konstatování, že lidská mysl je příliš komplexní na to, aby mohl být složen z mnoha elementárních, složitě propojených jednotek *(top-down přístupy)*. V poslední době se ale ukazuje, že spíše použití přístupů vycházejících zespod *(bottom-up)* a výstavba systému od základních jednotek, přes základní reflexy až po komplexnější netriviální funkce může přinést lepší výsledky. Zmíněný přístup vytváří teoretický podklad pro tvorbu robustních systémů.

5 KLASICKÝ PŘÍSTUP VERSUS NEURONOVÉ SÍTĚ

Z obecného hlediska představují neuronové sítě univerzální výpočetní prostředek se stejnou výpočetní silou jako klasické počítače např. von neumannovské architektury (tj. pomocí neuronových sítí lze principiálně spočítat vše, co umí např. osobní počítač a naopak). Z hlediska formalizace je funkce sítě popsána velkým počtem váhových parametrů, což má nepříjemné konsekvence. Někteří autoři tento fakt nazírají značně antropocentricky (ve smyslu, co nepůsobí „esteticky" na člověka, nemůže být užitečné).

U klasických počítačů se funkce a algoritmy vkládají do stroje metodami, které kopírují sekvenční uvažování člověka. Navíc vytvořená rozhraní pro zadávání programů umožňují jejich tvorbu v takových jazycích, které jsou pouze redukovanou verzi lidského jazyka. Uživatel tedy není příliš zatěžován překladem svých požadavků do „jazyka" systému. U neuronových sítí se s takovými výhodami nesetkáme. Jejich paralelní povaha a způsob reprezentace „programů" je pro lidského uživatele nepřehledný a neintuitivní. Veškeré požadavky na návrh či změnu systému vyžaduje získání speciálních znalostí a zkušeností. Zmíněný nedostatek však není překážkou, která by znamenala zásadní omezení při tvorbě neuronových síti. Jedná se pouze o zdůraznění ikompaktibility systémů.

Hlavní výhodou a zároveň odlišností neuronových sítí od klasické von neumannovské architektury je jejich schopnost učit se. Požadovanou funkci sítě *neprogramujeme* tak, že bychom popsali přesný postup výpočtů její funkční hodnoty, ale síť sama abstrahuje a zobecňuje charakter funkce v adaptivním režimu procesu učení ze vzorových příkladů. V tomto smyslu neuronová síť připomíná inteligenci člověka, který získává mnohé své znalosti a dovednosti ze zkušenosti, kterou ani není ve většině případů schopen formulovat analyticky pomocí přesných pravidel či algoritmů. Neuronové sítě se nesnaží o popis ale o imitaci chování (Caudill&Buttler, 2000). Navíc také způsob reprezentace paměti neuronovými sítěmi umožňuje dynamickou modifikaci uložených informací v závislosti na prostředí a také možnost pracovat s těmito obsahy (například procesem generalizace). Klasická architektura nabízí pouze statické uložení informace v paměťovém skladu.

Schopnost učit se a zobecňovat je typickou vlastností lidské inteligence. Velkým problémem pro hodnocení generalizační schopnosti neuronové sítě je, že není jasné, jakým

způsobem definovat, co je to správná generalizace. Díky tomu, že neumíme definovat (formalizovat) ani měřit generalizační schopnosti neuronových sítí, chybí základní kritérium, které by rozhodlo, jaké modely neuronových sítí jsou v konkrétním dobré či lepší než jiné apod. Generalizační schopnosti navržených modelů neuronových sítí se většinou ilustrují na jednotlivých příkladech, které (možná díky vhodnému výběru) vykazují dobré vlastnosti, ale tyto vlastnosti nelze nijak formálně ověřit (dokázat). Tento stav je příčinou krize základního výzkumu neuronových sítí (Šíma, 1996).

Na závěr hodnocení těchto architektur se samozřejmě nabízí možnost vzájemné koexistence a kooperace obou architektur v jednom systému. V oblasti simulace se setkáváme s těmito tendencemi v oblasti tvorby agentů, multiagentních systému či paralelních počítačů. Konkrétních aplikací však není ještě mnoho a jejich výsledky se v dostupné literatuře příliš nevyskytují. V rovině předpokladu je dobrým zhodnocením vzájemné kooperace Peregrinova stať, hledající inspiraci v evolučním principu.

Myslím, že nejefektivnější systémy jsou, tak jako ostatně i náš mozek a naše mysl, heterogenní. Přirozený výběr je, zdá se, mistr bastlení. Vezměme šachové programy. V učebnicích se dočteme, že jsou založeny na kombinaci funkce, která ohodnocuje pozice na šachovnici a výkonného modulu, který na několik tahů dopředu propočítává hodnoty všech dosažitelných postavení a volí cestu k tomu, které má tu nejvyšší. Skutečně funkční šachový program ale tohle jistě musí zkombinovat s komponentami zcela odlišného druhu: třeba s knihovnou zahájení či s modulem pro koncovky, který pracuje na úplně jiném principu atd. Myslím tedy, že nejperspektivnější jsou skutečně systémy, které ne příliš lahodí oku teoretika - systémy zbastlené z různých principů a technik (Peregrin, 2004).

6 PARALELISMUS

Hlavním cílem této kapitoly bude odlišení paralelního přístupu (také PDP, paralelismus, paralelní zpracování) od architektury neuronových sítí, se kterou bývá často zaměňován. Z velké části je to způsobeno skutečností, že oba přístupy zpracovávají informace paralelně. Rozdíl nacházíme v jednotlivých stavebních prvcích. U neuronových sítí je základní prvek schopen velmi malé výpočetní kapacity (pouze sčítat váhy a při překročení prahu „pálit") a neobsahuje paměť ve smyslu interního pracovního skladu pro výpočty základní jednotky.

U paralelních počítačů je základní jednotka tvořena procesorem, jehož výpočetní kapacita mu umožňuje provádět rozsáhlé výpočetní operace se svými vstupy. Obsahuje i paměť, která může sloužit jako skladiště pro mezivýpočty či ukládání výsledku pro pozdější odeslání. Z hlediska lokalizace paměti existují i varianty, kde jsou jednotky tvořeny pouze procesory a veškerá paměť je uložena v centrální paměťové jednotce mimo procesory (což odpovídá předchůdcům von neumannovské architektury). V takovém případě je přístup značně neefektivní pro tvorbu systémů s nutností rychlé odezvy na vstupy, protože je třeba neustále přenášet výstupy procesoru do centrální paměti i pokud se jedná o mezivýpočet.

Podstatný rozdíl je v základním pohledu na architekturu. Paralelní počítače provádějí výpočty, zatímco neuronové sítě spíše odpovídají na podněty. Každý paralelně propojený procesor má své taktovací hodiny, které udávají rychlost vykonání jednoho výpočetního kroku a v případě komunikace mezi sebou je informace také vysílána s ohledem na taktovací frekvenci. Neuronové sítě pracují asynchronně. Pokud přicházejí vstupy, jsou váženy, pokud na vstup nic nepřichází, neuron nepracuje.

V případě hybridní architektury paralelních počítačů (je tvořena kombinací neuronových sítí a klasických počítačů) musel doznat změn i jazyk, skrze který lze tvořit algoritmy, řídící průběh paralelních výpočtů. Příkladem může být procesy popisující jazyk (PDL), který byl prezentován Steelsem v roce 1992. Vychází z paralelistické idey, že procesy v systému probíhají kontinuálně a paralelně vedle sebe. Neexistuje kontrola chování na základě centralizovaného výběru. Místo toho každý proces může ovlivňovat jistý počet hodnot, které spolu s ostatními vedou k určitému chování pomocí sdílení. Jedná se o paralelistický přístup, který je založen právě na propojování modulů a jejich kooperaci či soutěžení modulů. Hrozí zde možnost zastavení činnosti, pokud není dobře definováno vzájemné ovlivňování se modulů. Pokud dojde ke konfliktu protichůdných modulů (či jejich skupin), systém se zablokuje a není

schopen pokračovat v činnosti (Pfeifer&Scheier, 2001). Obecně lze říci, že oproti programovacím jazykům klasických počítačů bylo nutné rozšířit jejich bázi o nové příkazy, které jsou schopné ošetřit komunikaci mezi jednotlivými moduly. Objevují se tedy jazyky, které dokáží vytvářet algoritmy pro ovládání synchronních přenosů, sdílení dat, kooperaci výpočtů, hierarchie součinnosti a jiné, zajišťující přechod klasických počítačů do oblasti paralelního zpracovávání. Naprosto identická je situace při tvorbě komunit z jednotlivých agentů v oblasti agentového přístupu. Ačkoli se vyskytuje velmi málo literatury, týkající se popisovaného přístupu, patří svou povahou a možnostmi, mezi velmi nadějnou kombinaci na poli simulace inteligence. Stejně jako multiagentní přístup, má i paralelismus předpoklady stát se klíčovým paradigmatem kognitivních věd.

7 PŘÍSTUP ZALOŽENÝ NA AGENTECH

V našem bádání o způsobech tvorby inteligentního systému se dostáváme k přístupu založeném na agentech. Oproti předchozím kapitolám (kromě paralelismu), zabývajících se možnostmi v oblasti architektury a možnostmi vhodně reprezentovat vstupní informaci, přecházíme do aplikovanější oblasti. Těžiště zájmu je přesunuto z oblasti tvorby univerzálních architektur a reprezentačních systémů do syntetického zpracování stávajících poznatků v jeden celek. Znamená to využití stávajících architektur (komputační, neuronové sítě, paralelismus) a reprezentačních systému k nalezení takového propojení, které využije jejich potenci.
Prvořadým cílem přístupu je však jiné propojení. Jedná se o interakci systému s jeho prostředím. Předchozí kapitoly vypovídaly o teoretických přístupech, které se snaží nalézt nejefektivnější formu a podmínky pro výstavbu systému schopného reprezentovat a zpracovávat informace. Vždy jsme však vycházeli z předpokladu, že v počátku (na vstupu) pracujeme s takovým kódem, který je vlastní simulovanému systému. Zcela jsme pomíjeli převod informace z prostředí do interní formy (transdukce). Proto je nutné přejít do komplexnější roviny přemýšlení (ve smyslu větších celků) a nazřít problematiku inteligence systému v kontextu jeho pobytu a pohybu v prostředí.
Rozšiřujeme tak původní schéma o vnímání a konání, čímž konstituujeme agenta (racionálního agenta, inteligentního agenta, autonomního agenta). Pokusů o jeho definici je stejné množství jako pokusů o definování inteligence. Vyplývá z faktu, že pojem agent můžeme ztotožnit s pojmem inteligentní systém. Uveďme si některé pokusy o vystižení pojmu :

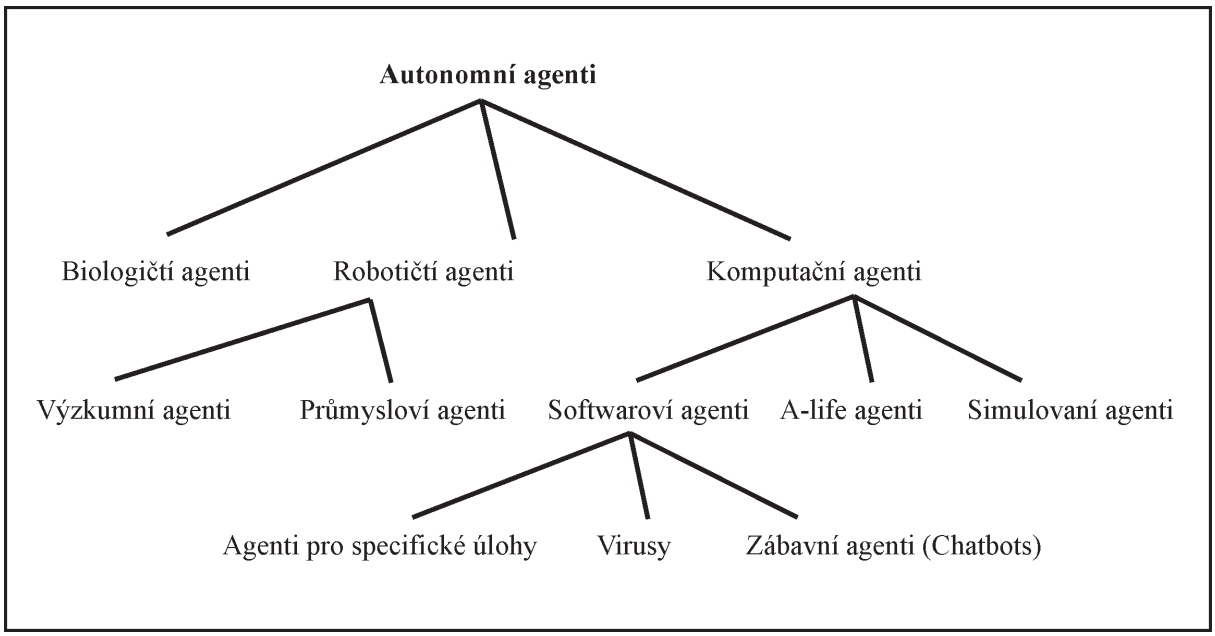
- Agent je cokoliv, co může být nahlíženo jako vnímající prostředí skrze své senzory a konající v prostředí skrze efektory (Russel&Norwig, 1995).
- Inteligentní agent vykonává nepřetržitě tři funkce: vnímá dynamické podmínky prostředí; koná tak, aby ovlivnil prostředí; a uvažuje aby interpretoval vnímané, řešil problémy, vyvozoval závěry a vytvářel jednání (Hayes-Roth, 1995).
- Autonomní agent je systém schopný autonomních a cílevědomých činů v reálném světě (Brustoloni, 1991).

Zmiňované definice jsou značně obecné a nepřesné, ale můžeme na nich vidět, že základním požadavkem agenta je právě jeho umístění v prostředí - situovanost a

schopnost vnímat a konat. Shrnutím může být definice:

Autonomní agent je systém situovaný v prostředí jako součást prostředí, který jej vnímá a koná v něm, ve snaze získat takové vlastní uspořádání, které mu umožní ovlivnit své budoucí vnímání (Franklin, 1996).

Tato definice se snaží postihnout působení agenta včetně možnosti interních reprezentací a anticipace, což jsou vlastnosti a schopnosti člověka. V celkovém vyznění ale touha po přesné definici agenta odpovídá stejně marnému úsilí o neredukcionistickou definici člověka (který je také inteligentní systém) v celé šíři jeho možností.
Díky tomu, že se agentový přístup přesunul do aplikované oblasti simulace a modelování, vzniklo nepřeberné množství konkrétních projektů, které svou specifičností umožnily vznik mnoha specializovaných oblastí využití agentů (podobně jako expertní systémy). Jejich taxonomie přesahuje možnosti této publikace, uvádím zde proto pouze základní dělení, které si uchovává potřebnou rozlišovací schopnost. Třídění v sobě obsahuje různé formy kombinací použitých základních substrátů, architektur a reprezentačních systému.



Obr. 22 *Taxonomie autonomních agentů*

7.1 Vtělená kognitivní věda

Brooksem navržená architektura potlačování *(subsumption architecture)* byla prvním přístupem který směřoval k ustavení nového paradigmatu ve zkoumání inteligence, označovaného jako **robotika založená na chování**, dnes nazývána **vtělená kognitivní věda.**

Brooksova terminologie není zcela jednotná a používá některá slova v jiném významu než je obvyklé. Například slovo chování používá ve významu interakce systému s prostředím, ale také jako úlohou požadované chování, pokud mluví o vnitřních modulech (vrstvách). Obecně vidí chování jako proces, který není založen na klasickém zpracování informací a je v opozici proti expertním systémům, tedy těm, které používají *high-level ontologie* (abstraktní procesy). Rozhodovací procesy vedoucí k chování je možno klasifikovat jako decentralizované a dynamický rekonfigurovatelné. Každá z robotických senzorických zpětných vazeb (interní procesů) funguje kontinuálně. To znamená, že energie do jeho motorů (chování) záleží na propojení s prostředím. Pokud máme hovořit o rozhodovacích procesech, které aktivují určitý reflex, nejedná se o procesy uvnitř robota. Obecně se dá říci, že prostředí pro které je robot (agent) vytvořen, určuje výběr mezi mechanismy *low-level* reflexů (Hendriks-Jansen, 1996). Argumentace se na konci stává spíše obhajobou samotné subsumpční architektury (viz níže).

Vtělená kognitivní věda často vytýká tradičním přístupům zanedbávání interakce systému s prostředím. Hlavní námitka je vznesena proti způsobu vkládání informace na informační vstup. Klasický přístup používá předzpracovanou informaci (transdukované člověkem) směrovanou na vstupy systému. Komplexní informace o prostředí je transformována do podoby, které systém rozumí. V reálném prostředí ale taková možnost není. Systém sám musí obsahovat mechanismy, které umožní transdukci informace na podobu, se kterou je schopen pracovat (Pfeifer&Scheier, 2001).

Problematika vtělenosti inteligentního systému obsahuje několik sporných míst. Podle tvrzení R. Brookse inteligence vyžaduje tělo. Souvisí to v podstatě s kořenovým problémem dualismu či monismu, čili otázky, zda se svět skládá pouze z jediné entity, či vyžaduje interakci energie a hmoty. Případná odpověď dělí vědce na dva tábory, které jsou následně fragmentovány na přesněji formulovaná paradigmata daných oborů. V aplikované oblasti se termínu **vtělenost** – *embodiment* - používá přímo se zřetelem na tvorbu autonomních agentů (inteligentních robotů). Pojem vtělenost je brán jako potřeba robota se pohybovat v prostředí, které je mimo něj a mít možnost toto okolí

vnímat skrze svou senzorovou výbavu. Velmi blízkým termínem a podobným termínem je situovanost. Agent (robot) je situován, jestliže dokáže získat dostatečnou informaci o své poloze v prostředí a odpovědět na ni patřičným chování. Vtělenost i **situovanost** jsou předpoklady vedoucí k **autonomii** agenta (robota), tedy že bude schopen samostatně interagovat s prostředím (Pfeifer&Scheier, 2001). Oba požadavky jsou ve své podstatě důsledkem nevyřešenosti základního problému, kterým je ukotvení symbolů *(symbol grounding).* Schopnost orientace v prostředí a vydělení sebe coby jeho součást, vyžaduje jako základní požadavek schopnost prostředí „rozumět". Tedy překódovat informace externího světa do interního kódu tak, aby jejich zpracovávání obsahovalo sémantickou rovinu. Poté už agent nepotřebuje člověka jako *interpreter* (překladač) mezi externím světem a svými interními stavy.

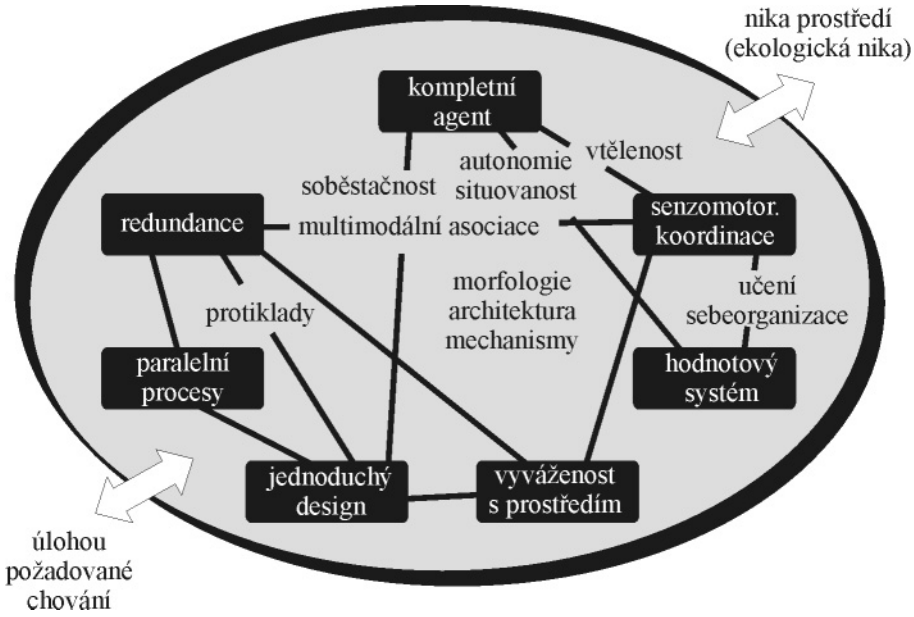
7.2 Principy tvorby autonomních agentů

Konstruktéři při tvorbě systémů často vycházejí z intuitivních předpokladů jevů, které jsou svou povahou implicitní. Při tvorbě systémů (agentů) jsme nuceni znalosti a postupy co nejvíce převádět do explicitní formy, kterou lze jasně popsat.Existuje totiž množství teoretických přístupů, které v sobě obsahují principy, o kterých je známo, že fungují, ale není zřejmé jak. Možnosti pochopení systému se tak stávají omezené a výhodou je pouze, pokud se designér věnuje problematice delší dobu, jeho schopnost pracovat s implicitními předpoklady daleko kvalitněji, díky prohlubující se znalosti, vedoucí k identifikaci a explikaci potřebných mechanismů. Ve vtělené kognitivní vědě se jedná o využívání znalostí z oblasti biologie a jejich užívání při návrhu agenta. Biologické mechanismy nejsou přesně známy, ale jsme schopni je ve zjednodušené formě napodobovat. Vysvětlit si to lze buď jako nemožnost některé biologické procesy plnohodnotně napodobit pomocí prvků neživé přírody (viz Biologie), nebo také nedostatečnými výsledky poznání v těchto oblastech.

Jeden ze základních principů při navrhování autonomních agentů postupuje následně. Předpokládáme, že chování je vždy propojeno s prostředím, ve kterém se agent pohybuje a neexistuje abstraktní chování mimo prostředí. Pokud dokážeme identifikovat druhy chování, které se v daném prostředí vyskytuje, snažíme se nalézt mechanismy, které toto chování tvoří a následně vztahy mezi jednotlivými typy chování. Tímto postupem se však většinou dostaneme do pozice, kdy se snažíme vytvořit agenta, který dokáže plnit pouze specifické úlohy (Pfeifer&Scheier, 2001). Množina možných chování začíná v komplexnějším prostředí narůstat a stane se výpočetně náročnou.

Pokud začneme hledat vztahy mezi různými typy chování, můžeme nalézt úspory plynoucí z kooperace mechanismů zajišťujících jednotlivé chování, ale dostáváme se i do situace, kdy složitost architektury agenta začíná přesahovat kapacitní možnosti designéra.

Pokud se podíváme na konstrukci autonomního agenta spíše z hlediska použitých principů, vlastností a mechanismů než omezení, můžeme ji vyjádřit následujícím obrázkem. Nutno zdůraznit, že jde o jednu z možných variant (lépe řečeno z předpokladů) a proto jí nelze generalizovat jako obecný princip tvorby agentů.



Obr. 23 *Podmínky tvorby automního agenta*

Při tvorbě agentů bývá hlavním cílem aby dosahoval emergentního chování. Častěji se emergentní princip používá v oblastech jako simulace umělého života, dynamických systémů a neuronových sítí. Přístupy umožňují díky principu sebeorganizace, možnost vzniku vlastností, které předtím v systému nebyly přítomné.

7.3 Emergence

Emergentní princip hovoří o vynořování se nových vlastností či výsledků, které nejsou v systému obsaženy v explicitní formě. Přesnější analýzu problematiku a některé předpoklady pro vznik emergence uvádí následující odstavec.

V začátku si rozložíme kompletní systém (prostředí+inteligentní systém) na prostředí

a inteligentní systém jako samostatné části. V první pasáži budeme brát v potaz pouze inteligentní systém. Předpokladem pro vznik „nového" je takový design architektur, které nabízejí více variant propojení jednotlivých operátorů. Pokud by systém obsahoval pouze propojení nutná pro speciální účel navržený designérem, hovoříme o emergenci pouze v případě, že se poškodí jeden s operátorů (systém zůstane funkční i bez něj), nebo operátor funguje vadně a produkuje špatné výstupy (což není cílem tvůrců inteligentních systémů). To je ale nepravá emergence, jelikož nedošlo k vytvoření alternativní stavu , ale k přechodu jednoho stavu do téhož modifikované stavu (determinismus). Pokud systém obsahuje redundantní propojení (tedy některá propojení jsou v určitých fázích nevyužitá a jakoby zbytečná), mohou vzniknout dvě varianty emergence. První možností je **sekvenční emergence**. Jednotlivé operátory zpracovávají informace z prostředí v určité posloupnosti, ale možnosti propojení jim umožní i jiný průběh zpracování a tak je možný vznik „nového" jinou sekvencí práce operátorů (nondeterminismus). Typickou oblastí aplikace je sériový způsob zpracování.

Systém pracující paralelně, může tvořit nové chování ještě jiným způsobem. Jestliže pracují operátory nezávisle na sobě, mohou se vzájemně ovlivňovat v činnosti (vyměňují si informace o svých stavech) a výsledné chování pak obsahuje nové varianty kooperace operátorů, jelikož operátory měly více informací a ty změnily průběh „výpočtu" **(kooperativní emergence).** Důležitým kritériem pro vznik emergence je tedy patřičná architektura systému.

Nyní již budeme hovořit o inteligentním systému v prostředí a o klíčové úloze prostředí pro možnost emergence u inteligentního systému. Konstantním prostředí vytváří stále stejné vstupy a systém tak nemá možnost tvorby alternativy, protože není k čemu. V dynamickém prostředí, je emergence možná. Právě změna v prostředí může být kauzální příčinou vzniku nového chování. Bez vlivu prostředí není možné restrukturovat pořadí či způsob kooperace operátorů.

Současné přístupy v oblasti neuronových sítí hovoří o emergenci v poněkud pozměněném smyslu. Síť se nejprve trénuje, tím že jí prezentujeme podněty, které jsou pomocí sítě kategorizovány. Pří prezentaci vzoru síť funguje určitým způsobem a kategorie je funkcí (způsobem propojení). V takovém případě emergentní chování znamená, že síť je schopna kategorizovat i podnět, který nebyl v cvičné sadě (je nový). Síť sice produkuje nové chování tím, že dokáže zařadit neznámý podnět, ale to je pouze rozmezí platnosti kategorizační funkce a ne emergence v pravém slova smyslu. Podobně je tomu u subsumpční architektury. Zde pracuje několik paralelních

operátorů, které mají přímé napojení na vstup a výstup, ale jejich činnost není kooperativní. Pravidlo naopak určuje, že vyšší úrovně jsou potlačovány, pokud se systém ocitne v situaci, která je pro něj problematická (nová či kritická), a činnost přebírají nižší vrstvy, podobné reflexům u člověka. Komunikace mezi vrstvami je minimální, jedná se o typ, který není ani čistě sekvenční ani paralelní. Emergence vzniká činností jednotlivých vrstev s omezenými možnostmi jejich součinnosti při produkci chování. Zde je nejlépe vidět vliv prostředí na vznik emergentního chování. Vzniknuvší agent (v případě vtělené subsumpční architektury) neustále akomoduje na okolní prostředí a nemá možnost řídit své chování. K tomu je potřeba interní reprezentace prostředí (paměti) a také „schopnosti", která je u živých bytostí nazývána vědomím. To je poslední důležitou podmínkou pro vznik emergentního chování. Pokud je součástí činnosti vědomí, schopnost kontroly a zpětné vazby na prostředí a interní procesy, může docházet u vědomého systému k emergentnímu chování i bez přítomnosti prostředí. Pokud je systém schopen práce se svými interními reprezentacemi, může provádět nové způsoby propojení operátorů, bez toho aby jej prostředí ovlivňovalo či limitovalo.

Klíčovou se pro vznik emergentních vlastností uvnitř inteligentního systému jeví architektura obsahující redundantní propojení, schopná vytvářet interní reprezentace a vědomě s nimi manipulovat. Systém musí být navíc situován v dynamickém prostředí, se kterým interaguje.

7.4 Výhody a nevýhody agentů

Tato kapitola již nemůže mít argumentační sílu, jako podobné odstavce z počátku práce. Bylo zmíněno, že agentový přístup je syntézou. Výhody a nevýhody zde zmíněné se vztahují spíše ke konkrétním základním architekturám, které jsou při konstrukci agentů použity. Přestože některé z nich už zazněly v předchozích kapitolách, jsou zde uvedeny v pozměněném kontextu.

Pokud aplikujeme architekturu neuronové sítě do autonomního agenta, získáme tak několik výhod. Naučí se například generalizovat podněty. V konečné fázi vede proces generalizace k „anticipaci" objektu v prostředí (Pfeifer&Scheier, 2001). Není to ovšem vědomá anticipace, neboť agentům dosud nemůžeme přisoudit tento status pro jejich malou komplexnost. Schopnost anticipace mu přisuzuje pozorovatel (člověk), protože si takto anticipaci vysvětluje sám na sobě.

Agent vybavený neuronovou sítí může kromě učení také zapomínat, pokud použijeme

model Hebbovského učení. Při jeho aplikaci se může stát (při opakovaném vystavení určitého podnětu), že váhy neuronu dosahují velkých hodnot. Tomu je zabráněno mechanismem zapomínání, který snižuje při každém nové vstupu předchozí hodnotu váhy neuronu. V okamžiku, kdy není spojení dlouhou dobu využíváno (není prezentován patřičný podnět), váha se sníží na nulu a neuron tento podnět zapomene. Takto vybavená neuronová síť se může učit nekonečně dlouho. Pokud je prostředí neměnné, váhová matice se dostane do stavu equilibria. Jestliže se prostředí mění v optimálním poměru k velikosti hodnoty zapomínacího mechanismu, je funkce sítě optimální, protože si pamatuje pouze podněty, které aktuálně potřebuje ke své činnosti (Pfeifer&Scheier, 2001).

Oblastí, která je při tvorbě agentů často přehlížená, je motivace. Je to dáno její blízkostí s emocemi, které se tradiční kognitivní věda vyhýbá, zcela jí pomíjí nebo se jí snaží nahradit pomocí mechanismů založených na algoritmech. Častými pokusy je také nahrazení tohoto pojmu pravděpodobnostními hodnotami vzdálenosti současného stavu agenta od cílového, či vytvořením systému operujících s konstrukty typu domněnka *(belief)* a potřeba *(desire)*. Složitější systémy se snaží vysvětlit motivaci pomocí emergentního principu, tedy propojením mnoha modulů dohromady vytváří chování, ale také interní stavy, ve kterých se projevuje dopředu nedefinována soutěživost či kooperace mezi moduly (Pfeifer&Scheier, 2001).

Teorie agentů stojí z hlediska využívání architektur mezi teoriemi, které využívají neuronu jako základní stavební jednotky, a teoriemi, které používají centrální procesor jako jádro, přes které procházejí všechny informace z okolí a jsou zde zpracovávány. Podobají se zčásti fodorovým modulům či paralelním počítačům. Architektura je postavena na kooperaci několika samostatných jednotek, z nichž každý má speciální funkci, kterou může plnit. Architektura autonomních agentů může být založena na klasickém přístupu či neuronové síti, s vědomím, že přebírá nevýhody jedné či druhé.

7.5 Reaktivní agenti

Rodney Brooks hovoří o tom, že tradiční přístup k UI je chybný. Většina myšlenek, jež se týkaly oblastí myšlení, logiky a řešení problémů, jsou odvozeny z naší vlastní introspekce, tedy jak my vidíme sebe samy. Tvrdí, že inteligentního chování lze dosáhnout použitím velkého množství „volně spojených procesů" *(loosely coupled processes),* které pracují převážně asynchronně a paralelně. Argumentoval, že není třeba příliš

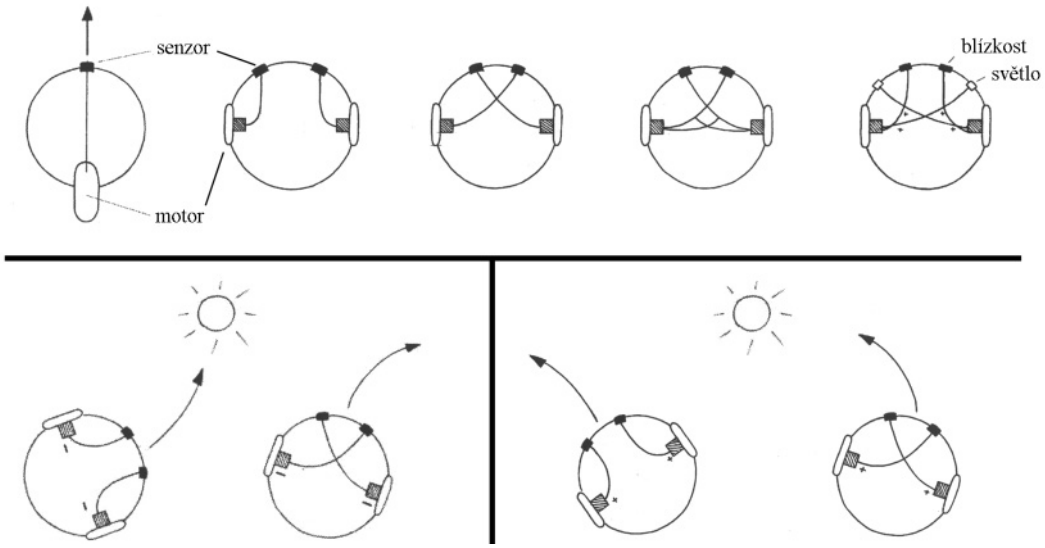
interních procesů zpracování a že sensorické vstupy mohou být mapovány relativně přímo na efektory (Pfeifer&Scheier, 2001). Brooksovým mottem je, že „Svět sám je svou nejlepší reprezentací". Interní reprezentace je třeba konstruovat pouze v případech, pokud je to nezbytně nutné. Tento typ architektury patří do skupiny **reaktivních agentů** (Mařík, 2001).

Takový typ architektury vede k úzké vazbě systému na prostředí . Inteligence je emergentní vlastností interakce organismu s prostředím. Jím navržený přístup je velmi odlišný od klasického přístupu zpracování informace.

Známým příkladem jednoduchého reaktivního agenta je mravenec Herberta Simona, kterého sledoval, jak běží po pláži (Simon mravence). Křivka jeho trajektorie byla nesmírně komplikovaná. Z hlediska lidského pozorovatele se může zdát, že mravenec provádí spoustu obtížných rozhodovacích procesů. Přitom se jedná o typ interakce, kdy o mravencově trajektorii rozhoduje z velké části prostředí (na podobném principu jsou založeny Braitenbergova vozítka). Mravenec se stává **externě řízeným systémem** (Mařík, 2001).

Základy reaktivního přístupu položil Braitenberg v experimentech se svými vozítky, které jsou nejvyšší formou redukce tvorbou systémů, jejichž vstupy jsou propojeny přímo s výstupy (v pokročilejší verzi s určitou formou redundance). Braitebergovy vozítka slouží jako dobrý příklad architektury, která neobsahuje interní reprezentace, ani žádné výpočetní mechanismy, ale kde jsou vstupy rovnou propojeny s výstup.

U neuronových sítí se to dá přirovnat k typu, který neobsahuje střední skrytou vrstvu. U klasické architektury jde o systém, kde je procesor pouze místem, přes které prochází propojení, přičemž nedochází k žádné formě zpracování.



Obr. 24 *Braitenbergovy vozítka - příklady (Braitenberg, 1984)*

Hlavním přínosem reaktivních agentů je zdůraznění způsobu propojení při konstrukci. Jejich architekturám nemůžeme přisoudit pojmy jako paměť nebo znalost, protože funguje vždy v přímé interakci s daným prostředím. (Pfeifer&Scheier, 2001).

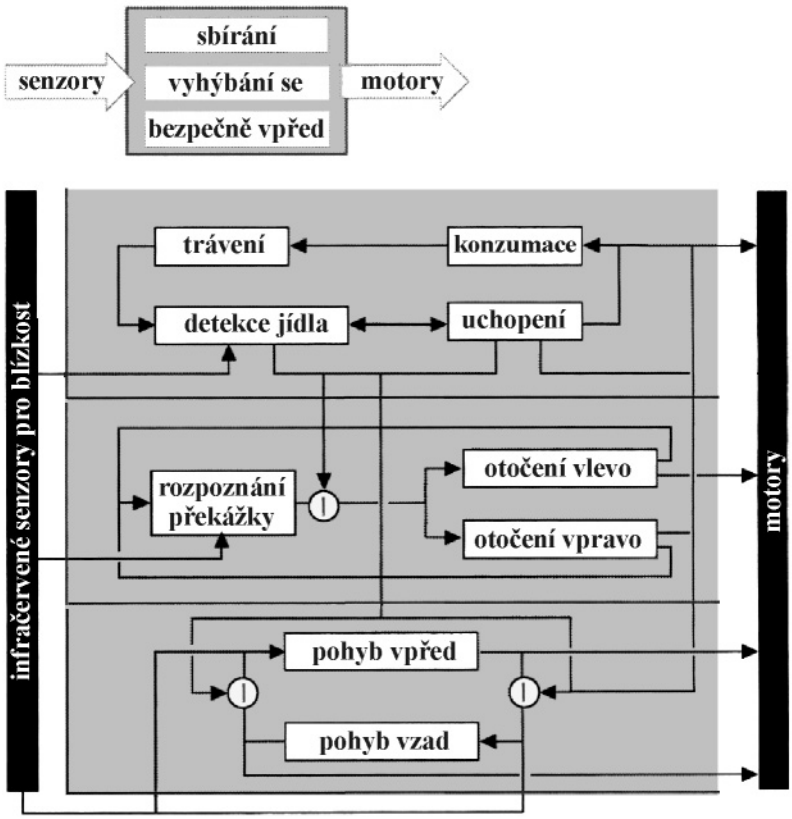
7.5.1 Subsumpce

Subsumpční architektura využívá metody dekompozice jednotlivých kontrolních architektur na množinu prvků plnících úkoly (Pfeifer&Scheier, 2001). Klasická funkční dekompozice provádí zpracování jednotlivých modalit samostatně. Ty jsou převáděny na centrální reprezentaci. Poté přicházejí na řadu interní procesy, které se snaží vytvořit model světa odpovídající vnějšímu prostředí. Následně dochází rozhodování a plánování akcí a jejich provedení. Taková dekompozice je založena na klasickém informačním přístupu využívajícího cyklus vnímání-myšlení-konání. Jelikož tento přístup umožňuje modelování a plánování (rozšíření položky myšlení), je cyklus rozšířen na vnímání-modelování-plánování-konání.

Oproti tomu Brooksův přístup obsahuje jednotlivé moduly, uspořádané ve vrstvách, které sice mají hierarchickou strukturu, ve smyslu od jednodušších po složitější operace, ale rozdíl je v tom, že není nutná kauzální posloupnost jednotlivých vrstev od jednoduché po složitou. Vrstvy pracují paralelně a nezávisle na sobě. Každá má stejné vstupy a má možnost ovládat efektory (výstupy). Můžeme hovořit o paralelním zpracování informací.

Mezi ideové předchůdce daného přístupu patří hierarchické systémy, jako například TOTE, které se také skládají z vzestupně uspořádaných množin (vrstev) modulů. Každý modul obsahuje funkci, která se podílí na celkové funkcionalitě. V architektuře TOTE ale nevzniká prostor pro emergentní chování, jelikož vše je specifikováno předem (hierarchie v pravém slova smyslu) (Pfeifer&Scheier, 2001).

V subsumpční architektuře je vrstva komponována jako soubor modulů. Každý modul je tvořen **rozšířeným strojem s konečným počtem stavů** (speciální *finite state automata*). To znamená výpočetní mechanismus založený na stavech, které se mění v závislosti na předchozím stavu a hodnotě vstupu. Rozšířenost hovoří o přidání některých dalších prvků oproti nejjednoduššímu stroji s konečným počtem stavů, kterým je Turingův stroj (bez nekonečně dlouhé pásky). Přídavné prvky mohou být různé, například časovač, který dokáže po určité časové periodě změnit vnitřní stav stroje nezávisle na hodnotě vstupu. (Pfeifer&Scheier, 2001).



Obr. 25 *Příklad subsumpční architektury*

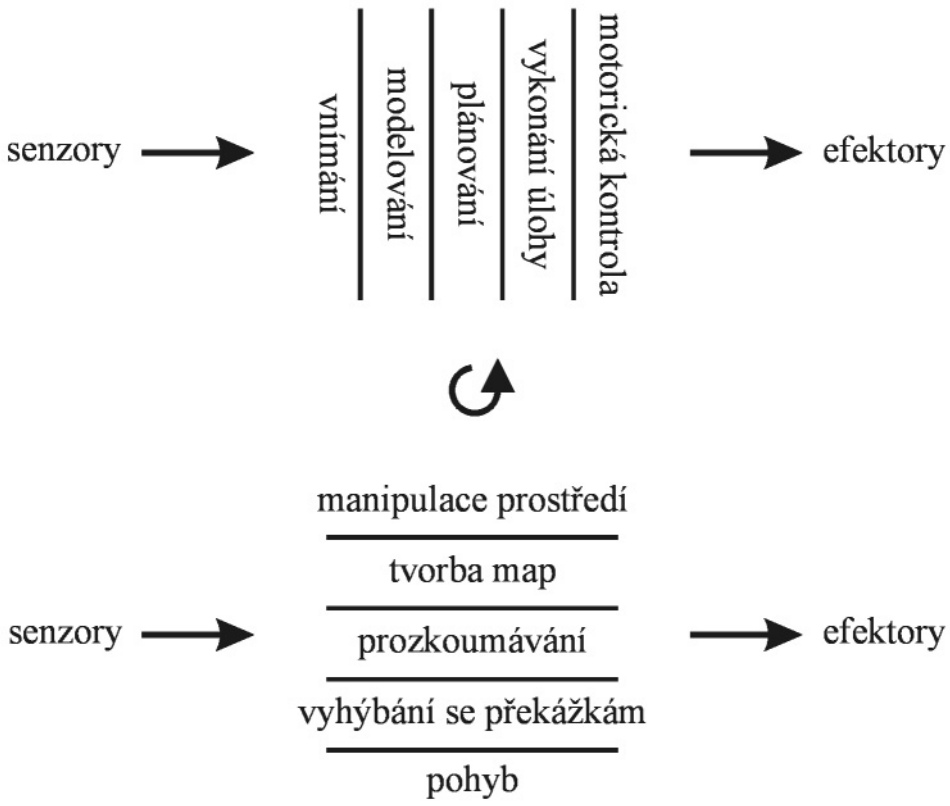
Jednotlivé vrstvy se mohou navzájem ovlivňovat a předávat si informace či výsledky svých činností (velmi omezeně a specificky), ale pozornost je soustředěna na jejich samostatnou činnost. Slovo hierarchie ve vztahu k vrstvám zde není používáno v klasickém významu. Má vztah k potlačování – subsumpci. Pokud nastane situace, při které by pracovaly dvě vrstvy antagonisticky, přichází na řadu inhibiční hierarchie. Nižší vrstvy jsou programovány jako reflexy, a proto získávají přednost ve vykonání akce proti vyšším vrstvám. Základním propojení vrstev mezi sebou tedy slouží k inhibici.

Existuje však více variant subsumpčního chování. Základní typ má napevno zakódováno, jaká je funkční hierarchie používání vrstev. Prostředí na tom (kromě přítomností aktivátorů) nemůže nic změnit. Existují ale komputačně náročnější verze těchto systémů, rozhodující o nadřazené aktivitě jedné vrstvy pomocí metod hlasování či rozprostřené aktivace. V takovém případě hraje prostředí roli aktivátoru, který nenabývá hodnot jedna/nula, ale jeho jednotlivé vlastnosti (subaktivátory) podle míry přítomnosti v dané situaci rozhodují o subsumpci, tedy o nadřazení vrstvy nad ostatními a následném chování. Paralelu můžeme hledat v Lorenzově „psycho-hydraulickém" modelu, který je založen na předpokladu, že každé chování, které vyvíjí konstantní úsilí, aby bylo aktivováno, se časem zvyšuje, pokud je dané chování používáno (Lorenz, 1981, s.47).

Druhý typ subsumpce vládne rozšířenější rozlišovací schopností pro volbu správného chování. Z lidského pohledu je schopnost emergence efektivního chování pravděpodobnější.

Výhodou architektury je její vnitřní rozšiřitelnost. Přidáním nové vrstvy nepotřebujeme rekonfigurovat celý systém, protože vrstvy jsou z velké části autonomní, což je dobré pro designéra, který tak může přidávat další prvky dle požadavku prostředí (podobně jako axiomy do formálního systému). Nevýhodou je nepropojenost jednotlivých komponent. Nenabízí se možnost emergence nového chování v sekvenci (viz Emergence). Jednotlivé procesy pracují samostatně a tak nemohou organizovaným skládáním vytvořit kvalitativně odlišné celky – sekvence (Konar, 1999).

Jak bylo zmíněno, Brooksův přístup nerozlišuje mezi periferním a centrálním jako jiné architektury. Reaktivní agent může být nahlížen jako systém s decentralizovanou aktivní reprezentaci svých schopností (Mataric, 1997). Pojem reprezentace je zde použit jinak než v klasické kognitivní psychologii. Jedná se o procedurální reakci systému na daný vstup. Těžko říci, zda se jedná o reprezentaci jako takovou. Systém neobsahuje paměť (formou interní reprezentace) a je vždy pouze reagujícím na současný stav prostředí. Použití slova reprezentace v tomto kontextu je zavádějící. Lze jej použít pouze metaforicky, tedy že reprezentace je tvořena součinností modulů v daný okamžik.



Obr. 26 *Tradiční architektura versus architektura vrstev*

Vraťme se však k porovnání s klasickými modely kognitivního cyklu. Systém je dělen horizontálně a neodděluje vnímání od konání jednotlivými mezistupni zpracování informace. Propojením mezi vnímáním a konáním je zajištěno v každé vrstvě a každou vrstvou podle toho, k čemu je specializovaná (Hogan, 1998). Tento základní posun ve schématu se může zdát překvapivě jednoduchým a efektivním, ale jedná se jen o pominutí některých faktů. Pokud bychom základní schéma dle obrázku otočili o 90 stupňů a propojili každou vrstvu s vstupy i výstupy, ztratí se nám někde procesy, které informaci předzpracovávají. Každá vrstva, která by chtěla pracovat se vstupní informací na netriviální úrovní, by v sobě musela obsahovat moduly, které by ji umožnily převod informace na takovou formu, se kterou dokáže pracovat (interní reprezentaci) . To znamená, že každá vrstva by obsahovala složité mechanismy práce s interními reprezentacemi, čímž by se tato architektura stala robustní a redundantní, ale již na úkor poměru velikost/efektivita. Přístup horizontální vrstev totiž vertikální přístup (čití/vnímání/myšlení/plánování/konání) neřeší, pouze jej zamlčuje. Většinou pracuje jen s jednoduchými reflexními vrstvami (například vrstva, která zajišťuje styk s předmětem), u které není třeba interní reprezentace či složitých procesů zpracování. Se stoupajícími nároky na inteligenci systému je ale nutné přidávat vrstvy, které již nedokáži svou činnost provozovat pouze na základě reaktivity, ale potřebují pro svou činnost lepší vybavení. Tvrdit, že komplexnější vlastnosti systému budou emergovat nekoordinovanou součinností velkého počtu vrstev, je nereálné.

Pokusem o posunutí v oblasti subsumpční architektury je imunologický přístup. Samotný název ukazuje, jakým směrem se badatelé vydali. Jedná se o vytvoření metafory k lidskému imunitnímu systému. Teoretické možnosti přesahující rámec této aplikace spočívají v předpokladu , že jednotlivé moduly jsou aktivní, stejně jako protilátky reagující na přítomnost antigenu v organismu. Moduly jsou opět základními funkčními jednotkami, ale nejsou pasivními ve smyslu očekávání zpracovatelné informace. Neustále mapují senzorické vstupy i ostatní moduly na přítomnost informace, kterou patří do jejich kompetence. V případě stimulace z prostředí vznikne situace, kdy je posloupností zpracování informace vytvořena unikátní kooperace modulů, jelikož aktivní moduly podle své funkce samy identifikují podněty objevující se na senzorech (filosofie mysli nabízí metaforu senzorického projekčního plátna, na které se dívá místo homunkula aktivní společenství modulů). Problematické je říci, jak lze tuto aktivitu modulů definovat. Představíme-li si zjednodušený model agenta v prostředí jako kauzální posloupnost prostředí-senzory-moduly-chování-prostředí,

pak můžeme říci, že kauzální posloupnost u subsumpční architektury je následující. Příčinou je prostředí, které vysílá informace o sobě na senzory, které je vysílají všem vrstvám a modulům v nich. Modul, který dokáže informaci nejlépe zpracovat potlačí činnost ostatních modulů (pokud je to nutné) a provede výpočet, který vede k výběru chování. Neobsahuje interní reprezentaci (paměť) a tak provádí algoritmus přiřazení chování vždy stejně.

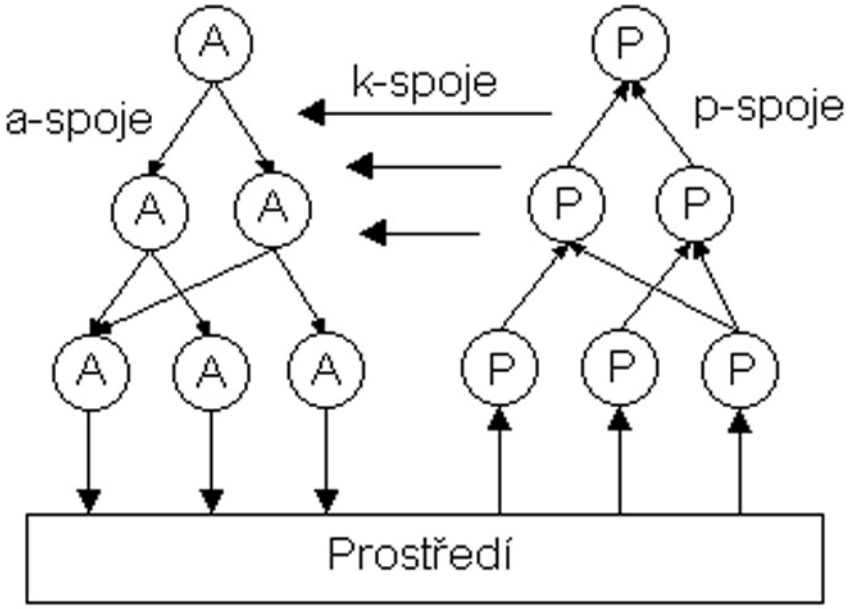
V případě imunologického přístupu jsou vrstvy mapující vstup aktivní. To znamená, že systém je schopen ovlivňovat to, co bude vnímáno (v psychologii bychom řekli anticipuje). Jelikož kauzální proces prostředí-senzor-modul je ireversibilní, můžeme pomocí anticipace vytvořit chování nezávisle na prostředí. Tento proces je možný pouze v případě, že je modul vybaven schopností paměti, která mu umožní uchovat znalost o prostředí či o svém stavu i v nepřítomnosti podnětu. Nelze ale říci, že by aktivita vyplývala pouze z přítomnosti paměti. Přístup se poněkud vymyká z kategorie čistě reaktivních agentů.

Ve své podstatě je Brooksova subsumpční architektura rozšířením klasické von neumannovské architektury o paralelismus (Pfeifer&Scheier, 2001). Jednoduché počítače pracují vedle sebe, všechny mají stejné vstupy a mohou pracovat nezávisle na sobě nebo využívat informace od ostatních. Architektura ovšem nenabízí možnost emergence kooperativního chování ze dvou důvodů. Propojení mezi vrstvami není redundantní, takže v případě potřeby komunikace je možno používat jen spojení, která byla dopředu naplánována designérem jako nezbytná pro dané prostředí. Tím, že nová spojení může provést pouze designér, architektura nenabízí výhody redundantního propojení neuronových sítí.

7.6 Multiagentní přístup

Vedle centralizovaných systémů (s bázemi pravidel a faktů a inferenčním mechanismem) existuje od počátku 80. let i myšlenka rozdělení systému na moduly, které mezi sebou kooperují podle přesně vytyčených pravidel. Jednoduché agenty se seskupují do **agentur**, které mohou produkovat složitější jednání než jednotliví agenti.

Tato myšlenka vychází ze **societní teorie mysli** Marwina Minského. Ten se domnívá, že se mysl skládá z jednoduchých agentů, kteří se sdružují do agentur (jedná se přenesení myšlenky tvorby agenta sdružováním jednotlivých modulů). Jsou to p-agenty a a-agenty (percepce a akce). Mezi nimi jsou spoje několika typů: **p-spoje** (zpracovávají percepce), **a-spoje** (vykonávají akce) a **k-spoje** (přenášejí znalosti).



Obr. 27 *Architektura societní teorie mysli*

Z prostředí se vnímají vzruchy. Ty se vedou pomocí p-spojů a jsou zpracovávány p-agenty. V okamžiku, kdy se informace dostane do takového p-agenta, který ji umí zpracovat, je přenesena k-spojením do příslušného a-agenta a dále vedena směrem "dolů" k a-agentu, který mění některou vlastnost prostředí.

Novinkou jsou právě k-spoje, znamenající způsob nakládání a ukládání znalostí. Jsou založeny na teorii *knowledge lines* (Minsky, 1986). Paměť při tomto typu ukládána zaznamenává jen část procesů. Jeho následným opakováním lze získat přesnější obraz prostředí, či procesů v okolním prostředí.

Totální mentální stav – (TVAR,BARVA,VELIKOST) – všichni patřiční agenti pro jednotlivé vlastnosti jsou aktivováni.

Částečný mentální stav – (BARVA,TVAR) – jen někteří agenti jsou aktivováni. Tento případ umožňuje ekonomický způsob zpracování informací. Může přejít v totální, pokud je to organismem vyžadováno (Minsky, 1986).

Tím je umožněno spojení paměťových obsahů v nový celek, jehož vlastnosti jsou emergentní a také značně odlišné od vlastnosti původních. Jedná se o optimální podmínky pro emergenci, tak jak o ni usilují některé směry UI. Tento typ ukládání paměti

(tedy neúplný obsah) můžeme aplikovat v oblasti řešení problému, rozhodování, vytváření úsudku či anticipace. Pokud by se podařilo tento model obhájit, lze jeho pomocí vysvětlit, proč se člověk neřídí striktně logickými pravidly, tedy ideálním a kýženým stavem, který mu přisuzují logici.

Na základě takových poznatků můžeme spekulovat, že fenomén **fantazie** je jen využíváním paměťových obsahů v jejich ranném nebo dokonale konsolidovaném stádiu. Fantazie pracuje s obsahy, jejichž uložení umožňuje více stupňů volnosti při jejich manipulaci.

8 PAMĚŤOVÉ SYSTÉMY A REPREZENTACE ZNALOSTÍ

Paměťový systém je zařízením, které je schopné ukládat informace a zpětně je vyvolat. Například neuronové sítě mohou fungovat jako asociativní paměťový systém. Asociativní paměť ukládá informace ve smyslu asociace a korelace s předchozími informacemi. Velikou výhodou je schopnost kategorizace, jejíž simulace je u klasických počítačů obtížná. Pokud je systému prezentováno několik podnětů, podobných si v určité vlastnosti, kterou dokáže systém obsáhnout v některé ze svých vah či souboru vah, vzniká schopnost tvorby třídy a také schopnost rozpoznat nový neznámý objekt, který má podobnou vlastnost (Caudill&Buttler, 2000).

Mnoho asociativních pamětí je uloženo v distribuované podobě. Záleží na způsobu reprezentace, ale u neuronových sítí bývá většinou používáno vah propojení jako paměťových atomů, jejichž složením vznikají vyšší celky, které se dají interpretovat jako reprezentace (Caudill&Buttler, 2000). Distribuovaným uložením informace získává systém robustnost, jelikož ztráta několika spojů nevede ke kompletní ztrátě uložené informace. Zdá se, že tento typ paměti je základní podmínou tvorby inteligentního systému (Caudill&Buttler, 2000).

U klasických počítačů je problematické uložit informaci asociativně. Souvisí to s používanými metodami zpracování informace, ale také způsobem zápisu. Do paměti RAM se ukládají informaci náhodně a na pevný disk také.(Caudill&Buttler, 2000). Což znamená nevýhody.

8.1 CAM

Mnozí badatelé pracující v oblasti *case based reasoning* (výzkum analogií) vypichují důležitost **účelu** a vybízejí k tomu, aby byly počítačové paměti indexovány způsobem, který posiluje vyhledávání analogů podle momentálních cílů (Kolodner, 1993,s. 92). Navrhují, aby se pro budování expertních systémů vyvinuly „obecně aplikovatelné indexující slovníky", které by se daly aplikovat pro všechny domény. Zda je takto indexovaná i lidská paměť lze rozhodnout jen psychologickými experimenty (Thagard, 2001).

Paměť člověka umožňuje vybavit pouze část požadované informace, přičemž následně dochází i k vybavení podobných či doplňujících informací. Tyto možnosti paměť

RAM nenabízí. Proto se objevuje požadavek po novém typu paměti, schopné ukládat a pracovat s uloženou informací asociativně. Zmíněné výhody nabízí obsahově adresovatelná paměť - CAM (Konar, 1999). Ta ukládá a vyhledává uložená data podle jejich obsahu. Nejjednodušším příkladem je databázový systém, který přiřadí informacím místo v paměti podle jejich blízkosti či podobnosti s předchozími informacemi. U klasických počítačů je realizace následující. Vstupní informace je rozbita pomocí speciální funkce, jejímž výsledkem je hodnota, která určí, kde bude tato informace uložena (Caudill&Buttler, 2000). Konekcionismus realizuje obsahové uložení paměti právě díky asociativním možnostem neuronových síti. Využitelnost paměti CAM leží v oblasti sémantiky. Pokud je informace v systému uložena formou reprezentace, která obsahuje kontextové okolí, je takový systém daleko lépe schopen vytvořit si rámec pro řešený úkol nebo pracovat s neúplnou informací. Zda je model paměti CAM dostačující pro vznik „pochopení" prostředí systémem se ukáže v budoucnu. Setkáváme se i s tvrzením, že nestačí pouze robustní forma uložení jedné modality, ale že je třeba propojení vstupů více modalit do jedné reprezentace, aby systém dokázal pracovat s obsahy prostředí.

8.2 Memory surface

Goldschlager ve svém modelu abstrahuje od vnitřní struktury mozku a soustřeďuje se pouze na důležité body této struktury a tok informací mezi nimi. Simuluje centrální část mozku, nazývanou *memory surface*, která má na starost zpracování preprocesovaných signálů z receptoru. Základní funkční jednotkou modelu jsou sloupce, které jsou spojeny v orientovaném grafu. Svou funkcí se nejvíce podobají impulzním perceptronům. Komunikace mezi sloupci probíhá formou sledu pulzů s proměnlivou frekvencí podélně váhových hran grafu. Některé sloupce mají spojení na předzpracované vjemy – reprezentují tak vstup do *memory surface*. Každý sloupec je charakterizován komunikačními a paměťovými charakteristikami. Vždy, když do sloupce dorazí pulz, je dále vyslán párovou hranou k té, kterou přišel. Pulzy přicházející vstupními hranami se během krátkých časových úseků sčítají a nakonec je vyprodukován sled pulzů s frekvencí f, která je přímo úměrná získané sumě. Paměťová charakteristika sloupce je záznamem jeho komunikačních změn. Komunikační parametry podléhají krátkodobé a dlouhodobé dynamice – z krátkodobého hlediska dochází k „únavě", tedy ke snižování výstupní frekvence při delší stimulaci. Dlouhodobá paměť je zabezpečena změnami vah aktivních vstupů (Kozej, 2004).

Reprezentace znalostí je v modelu povrchu paměti *(memory surface)* realizována pomocí konceptů složených z obrazců (patterns). Každý obrazec je tvořen množinou sloupců a jejich (relativních, normalizovaných) aktivit, které vyjadřují důležitost daného sloupce v obrazci. „Říkáme", že obrazec je aktivní silou s, jestli se jeho momentální aktivita liší od předchozí s-násobně (Kozej, 2004)..

Tento typ architektury je vylepšenou verzí neuronových sítí. Tam je komunikace mezi neurony reprezentována váhami propojení mezi neurony, ale komunikace probíhá pouze na úrovni pálí/nepálí. Informace (či paměť) systému tak vzniká na úrovni sítě nebo části, která je aktivní během podnětu. *Memory surface* však nabízí vznik informace na úrovni jednoho neuronu. Je to způsobeno právě pulzním charakterem aktivity neuronu, nabízející možnost zakódovat informaci do sledu pulzů. Také možnost rozdílných frekvencí pulzů umožňuje kvalitativní změnu této architektury oproti klasické neuronové síti.

8.3 Mentální reprezentace

Jedna z možností, kterou lze rozdělit přístupy v oblasti zpracovávání informací, používá jako kriterium způsob vnímání a následné uložení informace v paměti inteligentního systému, respektive její přítomnost či nepřítomnost. Je zřejmé, že mezi inferenčním a ekologickým přístupem k percepci existují rozdíly. Podle jednoho je vnímání naučeným odhadem, podle druhého získávání informací. Z filosofického hlediska je inferenční přístup velmi blízký idealismu, tedy tvrzení, že percepce jsou myšlenky formované myslí, která vnímá okolní svět. Ekologický přístup má naopak blízko k realismu či prezentacionismu, ve kterém jsou vjemy svázány přímo prostředím (Sternberg, 1999). Již jsme se zmínili o směru prezentacionistů v kognitivních vědách. Jeho zastánci argumentují, že veškeré potřebné informace (tedy informace o jevech i o jejich význam a vazbách) jsou obsaženy v prostředí a systém musí být pouze vybaven mechanismy pro jejich zpracování, aby byl považován za inteligentní. Tvrdí, že inteligence je již obsažena v samotném prostředí způsobem jeho průběhu (omezeného například fyzikálními zákony). Nedostatky přístupu se začínají objevovat až pokud se snažíme vysvětlovat některé schopnosti člověka, které pro své vysvětlení používají operace s nepřítomnými předměty či jevy světa, popřípadě jejich deformaci, kombinace nad úrovní fyzikálních zákonů shrnutelné pod pojmy fantazie

a kreativita.

Druhý přístup, který našel širokou odezvu zvláště v kognitivní psychologii, se nazývá reprezentacionismus. Jak již název napovídá, tento směr připouští možnost vytváření kopie reality v paměti inteligentních systémů. Použití slova kopie by ale bylo velmi nepřesné. Nejedná se o pouhé otisky, ale o tvorbu samostatných aktivních celků, které jsou schopny v sobě obsáhnout a postihnout větší část reality než jeden předmět či jev (jako v případě kopie). Mentální reprezentace nejsou obrazem reality, jsou vnitřní reprezentací možnosti (Dupoux, 2001). Právě díky schopnosti být univerzálním zástupcem skupiny jevů či objektů (stejně jako třída či kategorie) je mentální reprezentace povahy abstraktní a hypotetické. Její oprávněnost vychází z jednoduchých zjištění. Pokud bychom měli reprezentovat každou část či jev reality jako samostatný a unikátní prvek, který má specifické vlastnosti, musel by reprezentační systém mít kapacitu, která by překračovala možnosti operovat s nimi. Již v lidském paměťovém systému nacházíme základní předpoklady pro tvorbu úspor v procesu reprezentace, které nejen umožňují zpracovávat a operovat s obrazem (reprezentací) reality v abstraktní rovině, ale i schopnost tvorby a užívání jazykového kódu při práci s reprezentacemi. Představa, že je každá kopie reality uložena samostatně, evokuje představu poněkud zvláštního jazyka, můžeme-li ho v této úrovni ještě takto nazývat, pokud by se takovém případě vůbec vytvořil. To je ale pouze rovina spekulace.

Asi nejlepší knihou o mentálních reprezentacích v kontextu kognitivních věd je přehledová publikace Miluše Sedlákové z roku 2004, která zpracovává problematiku s podrobností, s jakou se jí autorka věnovala celý profesní život.

Již jsme se zmínili o mentální reprezentaci v souvislosti s lidským kognitivním aparátem. Jak je to ale u umělých systému? Může nám tento způsob uložení informace pomoci?

V obecné rovině určitě ano. Výhody, které přináší pro lidský kognitivní aparát, platí0 v oblasti simulace dvojnásob. Výpočetní a paměťové schopnosti jsou u umělých systémů daleko omezenější, než u člověka (UI nahrazuje nedostatky hrubou výpočetní silou). Právě možnosti reprezentace (abstrakce od jednotlivin reality) a její použitelnost v následném procesu kategorizace, generalizace apod. činí mentální reprezentaci dostatečně efektivní a univerzální pro oblast simulace.

Mentální reprezentace není pouhým uložením surových informací z prostředí (transdukovaných pomocí libovolné kvantifikace senzorů). Jedná se o sofistikovanější formu. Reprezentace musí být uloženy v takové podobě, aby způsob, kterým jsou vyjádřeny, byl plně pochopitelný pro systém, který je zpracovává. Kromě toho musí

splňovat způsob uložení (kódování) ještě podmínku **využitelnosti**. Měl by obsahovat takové typy vyjádření, které co nejvíce korespondují z mechanismy, které jej využívají. V psychologii i umělé inteligenci je tento požadavek pouze vysněným cílem. Nedostatek informací o mechanismech pracujících s reprezentacemi nám neumožní navrhnout způsob uložení reprezentace vhodnou formou.

Mentální reprezentace je teoretická entita, reprezentovaná pouze tím, co o ní řekne daná teorie (Crane, 2002)	
Redukcionistická definice	X reprezentuje Y pouze a jedině za podmínek ….
Konceptuální definice	Červená je to, co vidí normální člověk jako červenou.
Naturalistická definice	Červená je světlo vlnové délky 545 nm.
Fodorova definice	Základními znaky MR jsou podobnost a kauzalita.

Tab.3 *Způsoby nazírání mentální reprezentace*

V současných teoriích mentální reprezentace se setkáváme s potřebou definovat v mentální reprezentaci také prvky, které ji nekonstituují, ale naopak, jsou jejím protikladem. Pokud je to myšleno pouze jako negace vlastností, které reprezentace obsahuje (logické NOT), vystačili bychom si při tvorbě reprezentace pouze s formální logikou. Proces vymezení je ale složitější. Chceme-li do reprezentace včlenit i prvky, které vedou k její nepřítomnosti, potřebujeme k tomu více než formální aparát. Kontext a obsah jsou opět nutnou součástí při jejím vymezování.

Některé přístupy ke zkoumání mentálních reprezentací se ji snaží izolovat a zkoumat samostatně. Pokud ale zkoumáme mentální reprezentaci samostatně, mimo rámec okolního světa, rezignujeme tím na neisserovský cyklus vnímání či brunswikovskou čočku, jež v sobě okolí obsahují. Reprezentace je vždy vázána a vychází ze svého základu - prostředí. Pokud tento fakt pomineme a zkoumáme ji jako součást uzavřeného systému, může nabývat takových forem a vlastností, které při zpětné použití v reálném prostředí nejsme schopni obhájit (například Winogradovo SHRDLU).Dalšími tématy a otázkami, kterými se při zkoumání mentální reprezentace můžeme setkat jsou: Existuje mentální reprezentace emoce? Existují opravdu mentální reprezentace prvního řádu? Jedná se o uvědomění bez porozumění?

Jaká je tedy využitelnost mentálních reprezentací pro simulaci inteligence?

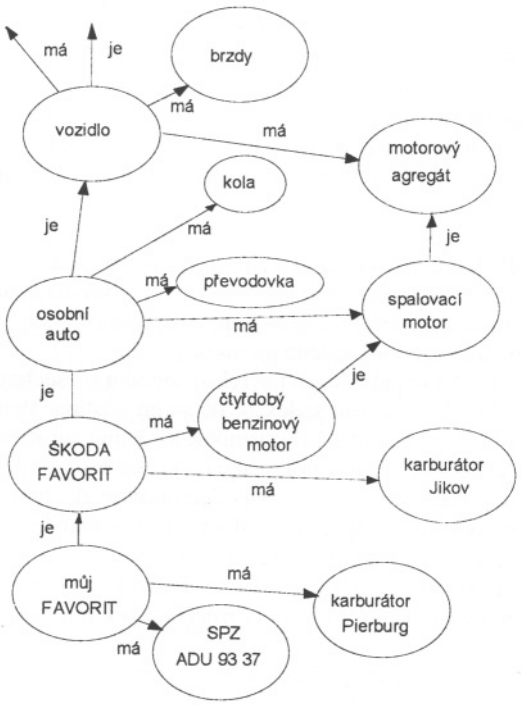
Ať je to dáno zčásti tím, že vzešla z psychologie, či její samotnou kvalitou, mentální reprezentace se svou strukturou nejvíce blíži takovému způsobu uložení informace v paměti, která při jejím zpětném použití umožní rekonstrukci kontextu a významu. Je třeba postoupit v oblasti tvorby umělých systémů (vhodnou volbou architektur), které dokáží plně a neredukcionisticky využit potence, jež tento paměťový systém (či spíše způsob uložení znalosti) nabízí.

8.4 Rámce

Řešení problému reprezentace poznatků bylo ve významné míře aktualizováno úvahou, kterou Minsky pod názvem *Matter, Mind, and Models* (Hmota, mysl a modely) publikoval roku 1968 v jím sestaveném knižním sborníku Semantic Information Processing (Zpracování sémantické informace). Sborník je prvním publikačním výstupem problematizujícím původní spoléhání výzkumníků v oblasti umělé inteligence na „hrubou sílu" stále výkonnějších počítačů a na hypotézu inteligence jako jediného integrálního principu, řídícího (lidské) jednání (Minsky, 1968). Tento posun vedl ke hledání specifických reprezentačních struktur zapamatování poznatků a specifických souborů procedur pro jednotlivé intelektuální aktivity (např. porozumění přirozenému jazyku, plánování postupů, řešení problémů, vidění apod.).

Za vyvrcholení snahy (která dominovala umělé inteligenci v 70. letech minulého století) najít pokud možno univerzální a efektivně zpracovatelné struktury reprezentace poznatků, můžeme pokládat návrh tzv. rámcové *(frame)* reprezentace. Minsky s touto ideou přišel v roce 1974, kdy navrhl mnohem komplexnější, tzv. tvarovou (gestalt) psychologií motivované chápání toho, co jsou a jak jsou asi v mysli organizovány poznatky.

Projevem snahy o vytvoření obecné reprezentace, použitelné pro široký okruh problémů, je tzv. teorie rámců *(theory of frames)*. Její základní myšlenkou je, že vstoupíme-li do nové situace, vybereme z paměti takovou strukturu, která jí odpovídá (podle předchozí zkušenosti) a srovnáme ji s aktuálně vnímaným světem. Rámce jsou tedy datové struktury reprezentující znalosti pomocí typických příkladů.



Obr. 28 *Sémantická síť - příklad*

Reprezentace znalostí založená na rámcích se považuje za alternativu reprezentace sémantickými sítěmi (viz. obrázek). Zdá se však, že pojem sítě je natolik univerzální, že by bylo možno na jeho základě teorii rámců vybudovat. V původní formulaci Minského se rámec považuje za základní, prototypovou strukturu, uloženou v paměti a vytvářející jakousi „kostru". Rámec je tedy charakterizován jako datová struktura k reprezentaci stereotypní situace (jako je např. pobyt v jistém druhu obývacího pokoje či návštěva večírku k oslavě narozenin). Ke každému rámci je připojeno několik druhů informací: některé se týkají toho, jak rámce použít, jiné toho, co se dá očekávat jako příští událost, a další zase informují, co dělat, nejsou-li očekávání splněna.

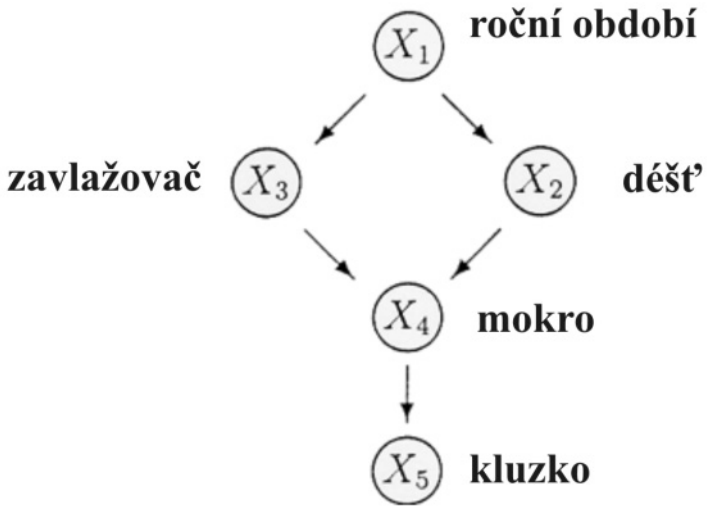
JMÉNO RÁMCE: Škoda Favorit
položky:
ČÍSLO RÁMCE: 1
IS-A: osobní auto
MOTOR: čtyřdobý benzínový
PŘEVODOVKA: manuální
KARBURÁTOR: Jikov
... atd.

Obr. 29 *Rámec - příklad*

Omezení použití rámců se objevuje v každém prostředí, kde dochází k dynamickým změnám. Jedná se o zachycování prvků a jejich vztahů v prostředí a rozpoznávání jejich důležitosti ve vztahu k pozorovateli. U člověka jsou za tyto úkoly odpovědné mechanismy pozornosti, vědomí, orientačně pátrací reflex, motivace a mnoho dalších. U umělého systému však máme k dispozici velmi málo prostředků jak tyto procesy nasimulovat. Janrelt rozděluje problém tvorby rámců na dva aspekty. Prvním je problém predikce, tedy schopnosti rozlišit, co je pro danou situaci relevantní a co ne. Druhým je kvalifikační problém, který určuje podmínky, za kterých může určitá akce nastat či nikoliv (Janrelt,1987). Některé aspekty tvorby rámců či definování toho, co je scéna, mohou vést k omezením, které nám neumožňují přechod mezi jednotlivými úrovněmi detailu v prostředí. Pokud používáme určité atomy vnímání (slova, čáry, objekty, geony, scény), narazíme na problém, pokud se snažíme reprezentovat nižší části (viz Nejmenší jednotky) či vyšší celek jako je scéna. Uplatňovaným procesem při přechodu do vyšších úrovní je generalizace či indukce. V tomto procesu musí docházet k redukci informace, či úrovně detailů základních objektů ve scéně. Pokud postupujeme metodou zobecňování, musí docházet i k redukci jednotlivých sémantických významů těchto atomů. Pracujeme-li pak s celou scénou, jak je možné, že nám redukovaný význam stále stačí k pochopení, či manipulaci s touto scénou? Pokud bychom abstrahovali stále více, časem původní objekt ztrácí zcela svůj význam v nové „makroscéně". Fascinující je způsob generalizace u člověka. Má ve většině případů dostatečné množství významových informací, aby mohl provést myšlenkovou operaci. Klíčovou se jeví schopnost mít v paměti uloženy i předchozí jednotlivosti, skrze které došlo ke generalizaci. V procesu indukce nám tato výhoda umožní zvratnost operace (přechod od indukčního vývodu zpět k informacím, které tvořily podklad pro indukci). Při dedukci je uchovávání starých informací spojeno s monotonií (která byla zmíněná jako neefektivní vlastnost inteligentního systému). Předchozí výtky ohledně monotónnosti umělých systémů tedy v popisovaném procesu nepůsobí jako omezení, ale jsou spíše usnadněním pro jeho následnou zvratnost. Jak bylo ale zmíněno výše, lidská mysl používá spíše pravidla mentální logiky, která je založena na významu, takže ji v této úrovni nemůžeme srovnávat s činností umělých systému založených syntakticky.

8.5 Bayesianské sítě

Jednou ze základních výhod Bayesianských sítí (které jsou vylepšenou variantou sítí sémantických), způsobenou jejich kauzální organizací, je schopnost reprezentovat a reflektovat změny v konfiguraci. Každá lokální rekonfigurace mechanismů v prostředí může být přeložena pouze s minimálními nároky na modifikaci sítě jako isomorfická rekonfigurace topologie sítě. Chceme-li ze sítě funkčně odstranit jeden objekt (reprezentovaný jedním uzlem), stačí zrušit všechna spojení, která k němu vedou (Wilson&Keil, 1999). Pokud ale potřebujeme spojení obnovit, jak budeme vědět, že tento uzel je stále součástí sítě? Schopnost flexibility je často uváděna jako základní vlastnost, která odděluje uvažující agenty od reaktivních a umožňuje uvažujícím zvládat nové situace kontinuálně, bez potřeby přetrénování nebo adaptace.



Obr. 30 *Bayesianská síť – příklad*

Bayesianské sítě mohou být použity pro modelování termodynamiky přestav díky tomu, že nahradí tradiční pravděpodobnosti nestandardními, které se dokáží nekonečně blížit jedničce nebo nule (Goldszmidt&Pearl, 1996)

9 TVORBA VÝZNAMU

9.1 Jazyk

Následující kapitola může působit poněkud heterogenně. Samotné téma jazyka je však svou povahou tak různorodé, že je nemožné jej uchopit způsobem, který by jej dokázal plnohodnotně popsat. Pokus o ucelené vyjádření jeho struktury a možných funkcí, byť jen pro oblast simulace by byl mnohovrstevnatý, proplétající se, sebeodkazující a vše, jen ne jednoduchý. I přesto bych rád některé vlastnosti jazyka zmínil, pro jeho klíčovou úlohu v oblastech, kterými se zabývá tato práce. Následující pasáže na sebe plynule nenavazují, jedná se spíše o soubor postřehů, které mají jedno společné téma.

Jazyk je organizován jako hierarchie se zvyšující se komplexitou. Objevují se nová pravidla významu a vztahy, které nemohou být vyjádřeny jako vlastnosti elementů, jež tvoří nižší úrovně, ale vycházejí z toho, jak jsou tyto elementy propojeny (Hogan, 1998). Příkladem může být genetický kód - jediný triplet nedokáže popsat stavbu člověka, ale celá DNA už ano.

Pokud byl jazyk utvářen jako komunikační nástroj mezi subjekty, nemůžeme jej používat jako objektivní hledisko, nebo jako věc samu o sobě.

Současné používání jazyka (libovolného) jako způsobu kódování programů či způsobu komunikace mezi jednotlivými stroji se setkává s neschopností zachycení významu pouhým kódováním do jazyka. Opět narážíme na arbitrárnost jazyka a potřebu referenčního kódování pro vznik významů. Je známo, že člověk dokáže číst knihy a rozumět jim, i když nejsou doplněny referenčním kódem (například obrazem). U umělého systému tato možnost neexistuje. Člověk má možnost si při četbě vytvořit obrazový průběh (kód) přečteného textu a skrze něj pochopit význam. „Obrazová paměť" používaná během čtení byla získána ještě před samotnou četbou (osobní historie). U umělého systému se nesetkáváme s tvorbou referenčního kódu, který by umožnil vznik významu. Podrobněji o tom hovoří kapitola o ukotvení symbolů.

Pokud je okolní svět prostorový, není možné vytvořit jeho prostorovou reprezentaci pouze jazykovým kódem, který je neprostorový, tedy neumožňuje věrně zachytit prostorové vztahy. Pomoci může matematika, která dokáže některé aspekty prostoru a hmoty kvantifikovat, či najít funkční souvislost (komplexní kauzální souvislost), ale jestliže popíšeme věci kauzálně, potřebujeme znát vždy počátek kauzality a tím extrémně zatěžujeme reprezentační mechanismus. Také nemůžeme kauzálně zachytit věci, jejichž kauzální historii neznáme. Matematika (či jiná formalizace) je nevhodná,

pokud potřebuji popsat rozmístění skříní ve svém pokoji, jehož kauzální historie je pro mě neznámá. I kdyby mi byla známa, bude nad mé výpočetní schopnosti. Obrazový kód v sobě takové prostorové vztahy a vlastnosti obsahuje. Jazykový kód může být aplikován až na obrazový, může umožnit jeho efektivní a ekonomickou redukci pomocí fragmentování, ale zmizí tím prostorové vztahy.

Počítačová a lidská řeč nejsou stejné, proto se mohou jen velmi slabě ovlivňovat. Jde spíše o to, že počítačová řeč je z lidské odvozená a je pouze jejím malým výsekem, takže možnosti ovlivňování jsou možné pouze ve směru lidská - počítačová (Boden, 1988).

Moderní lingvisté rozlišují mezi otevřenými a uzavřenými třídami slov. Otevřené třídy mohou být doplňovány, aniž by bylo třeba měnit základy jazyka, např. podstatná jména (opět paralela s formální logikou). Uzavřené třídy slouží k tvorbě gramatické struktury, např. členy, předložky. Jdou obtížněji rozebrat na části než otevřené třídy. Lidé s mozkovým poškozením je obtížněji rozlišují. Z hlediska počítačové teorie, Chomsky a moderní lingvisté tvrdí, že lidská mysl, jako výpočetní zařízení, obsahuje oddělené moduly pro syntaktickou a sémantickou analýzu (Sternberg, 1999).

9.2 Ukotvení symbolů

Problém, který nebyl v historii kognitivní vědy nikdy příliš brán v potaz, ale který je pro simulaci inteligence podstatný, souvisí s ukotvením symbolů. Situace je následující. Jak již bylo zmíněno v této práci mnohokrát, současné stroje zpracovávají své informace na syntaktické úrovni. Vyplývá to z jejich algoritmického způsobu operace s daty. Pokud ale systém potřebuje interagovat s reálným světem a pomocí zpracování interních reprezentací „o něm" provádět operace, které nejsou předprogramované (či netriviální), potřebuje zvládnout i sémantickou úroveň symbolů. To znamená soustředit se na problematiku tvorby významu. (Pfeifer&Scheier, 2001). Několik základních informací o tvorbě významu již bylo nastíněno v kapitole o mentálních reprezentacích. Problematika byla také zpracována samostatně.

Harnad nabízí řešení grounding problému vytvořením mechanismu, rozdělujícího smyslové vstupy do tříd podle jejich důležitých vlastností a přiřazující jim odpovídající symboly v reprezentaci. Symboly by byly pojmenováním významů nesených ve skutečnosti tímto mechanismem. Jedná se opět o přístup, který se snaží postulovat význam tím, že s ním explicitně počítá v mechanismech, které by měly tento proces umožnit.

Symboly a manipulace se symboly (které jsou častěji ukotvené pomocí tvaru než pomocí významu) bývají hodnoceny jako by význam měly, čímž konstituují symbolický systém jak jej známe. Takový způsob výkladu však nedefinuje symbolický systém jako vnitřní reprezentační systém. Jedná se parazitický způsob ukládání symbolů. Můžeme to přirovnat ke slovům v knize, která také nejsou vnitřní reprezentací, ale pouze médiem pro záznam symbolů, které nabývají významu až tehdy, kdy jsou zpracovány našimi mozky během čtení. Takže pokud jsou symboly uloženy v systému jako „vnější" reprezentace (oproti vnitřní v našem mozku), nelze mu přisoudit význam. Kognice není pouhá manipulace se symboly (Harnad, 1990).

Obvyklým názorem symbolistů je, že význam symbolů je ustaven skrze „správné" propojení symbolů se světem.

Ale ono „správné" propojení je jenom jinak nazvaný problém kognice, který je velmi obtížně řešitelný. Mnoho symbolistů také věří, že kognice je pouhou manipulací e symboly, která je uskutečňována skrze nezávislý funkční modul připojený na vstupní zařízení, umožňující mu „vidět" svět objektů, kterému odpovídají symboly. Opět poněkud vágní definice, která vychází z podivného principu „správného propojení" či „vidění", které je pro ustavení symbolů nutné, ale není blíže vysvětleno. Jedná se o podcenění převodu prostředí do formy, která je zastupitelná pomocí symbolů, tedy o zjednodušování problému ukotvení (Harnad, 1990). Teorie sémantiky musí být zásadně neredukcionistická, jelikož tvorba obsahu je založená spíše na expanzi.

Pro podrobnější vysvětlení ukotvení symbolů odkazuji na Harnadův článek (Harnad, 1990).

9.3 Kontext

Dlouhodobým problémem v oblasti simulace inteligence je schopnost pochopení. Tedy vytvoření kontextu, znalost významů podnětů a jejich souvztažnosti či kauzality (Konar, 1999). Potřeba studia a vymezení kontextu v rámci kognitivní vědy byla dlouhou dobu odsouvána na vedlejší kolej.

V třicátých letech se objevuje v Bartlettových pracích tématika individuálního a sociálního kontextu, kterým autor přikládá velkou důležitost. Trvalo ale dalších 30 let, než se objevila potřeba zkoumání kontextu v práce Ulrika Neissera (1976). Způsoby nazírání kontextu se v pracích následujících autorů značně liší. Vyplývá to ze samé povahy termínu, jelikož jeho vymezení je úzce spojeno se subjektivitou. První pokusy jej začleňovaly do mentálního rámce, kteroho jedinec využívá při své činnosti. Každý

jedinec prožívá tento fenomén prostřednictvím strukturované reprezentace svých znalostí. Používají se také výrazy schéma nebo reprezentace.

Pokud patřičná soustava neexistuje, je nutné ji vytvořit. Vzrůstá-li počet znalostí, které mají společný rámec, integrovaný do soustavy těchto poznatků, vzrůstají i možnosti práce s kontextem. Bohužel v případě takového vymezení vzniká tendence ztotožnit interní mentální soustavu se samotným kontextem, což by nebylo úplně přesné (Sternberg, 1999). Kontext lze také nazírat jako fenomén, který se vyskytuje i mimo inteligentní systém. Bronfenbernner hovoří o kontextu v rovině prostředí či v rovině sociálních vztahů.

Bartlettovým přínosem pro studium lidské paměti je odklon od nazírání paměti jako mechanického skladu informací. V jeho pojetí je paměť dynamickým procesem, ve kterém jsou přání, hodnoty, zážitky a zkušenosti jedince použity tak, aby umožnily vytvoření významu pro přicházející stimul. Nazýval tento proces „touhou po významu", čímž chtěl zdůraznit součinnost paměti s celým kognitivním aparátem, umožňující aktivní způsob zpracování vjemů a tendenci organismu je včlenit do své zkušenosti. To spíše vypovídá o procesech kategorizace a generalizace, ale kontext s danou problematikou úzce souvisí. Vymezení hranic kontextu z hlediska jeho zkoumání je velmi obtížné. Výzkumy, které slouží pro objasnění kontextu jsou většinou zaměřené spíše na vliv pozornosti, způsoby indukce, popřípadě řešení problému a otázky kontextu jsou až druhotné. Přitom existence kontextu je předpokladem pro vznik výše zmíněných jevů, které by bez něj nemohly proběhnout v sémantické rovině a nebylo by tak možno zajistit vznik významu, potažmo začlenění nových informací do stávajících (Sternberg, 1999).

Známým experimentem, který ukazuje na odlišnou tvorbu kontextu i pokud jedinci přijímají stejné informace, je Prohlídka domu. Zadání instrukce ZO je rozdílné pouze v jednom bodu. Jedna skupina je instruována tak, že by měla prohlížený dům vykrást a druhá koupit. Výsledná struktura znalostí o domě vykazovala u obou skupin signifikantní rozdíly. Tento druh přijímání a začleňování informací se nazývá **znalostní efekt**. Výzkum se dá ale opět zařadit spíše mezi příklady vlivu anticipace na prožívání. Fenomén kontextu se nám (z hlediska kognitivní psychologie) objevuje v každé fázi cyklu vnímání, což činí danou problematiku velmi obtížně vymezitelnou (Sternberg, 1999).

Zajímavý je vztah mezi asociacemi a kontextem. Kritika asocianismu poukazovala na to, že pouhými asociačními zákony nejsme schopni popsat procesy abstraktního myšlení, operace a manipulace s reprezentacemi a kreativitu. Ovšem otázka je polože-

na poněkud jinak. Asociační zákony slouží k ustavení struktury, popřípadě způsobu uložení paměťových obsahů takovým způsobem, který dokáže zachytit časové, prostorové a podobnostní vztahy. Tímto získáváme pouze v médiu uložené reprezentace prostředí (v praxi například paměti CAM). Pokud chceme hovořit o aktivním systému, potřebujeme také funkční komponentu. A to je chybějící část asocianismu. Pokud bude systém doplněn o operátory či mechanismy, které budou schopny zpracovávat uložené reprezentace s přihlédnutím k jejich kontextu (tedy asociačním vazbám), vytváříme tak systém, mající schopnosti (sporná je kreativita), které jsou čistému asocianismu upírány a které vedou k ustavení kontextově orientovaného systému. Sporná zůstává otázka, zda je paměť založená na asociacích schopná splnit požadavky ukotvení symbolů. Důležité pro práci operátorů je vytvářet vazby a vzájemně si vyměňovat informace (paralelismus) během práce s reprezentacemi. I pokud pracujeme s reprezentacemi ve formě asociací, neznamená to automaticky, že kontext bude vznikat na vyšších abstraktních úrovních, protože je nutné projít procesem generalizace, kde je třeba kontext ustavit znovu, jelikož získáváme kvalitativně nové reprezentace.

ZÁVĚR = SÉMANTIKA

Jako červená nit se vine celou touto prací jedno stěžejní téma, poukazující na zásadní nedostatek při tvorbě inteligentních systémů. Ve všech z probíraných přístupů se objevuje požadavek, aby systém, vytvořený na jeho základech, obsahoval schopnost „porozumět" prostředí, ve kterém se pohybuje. Mnohokrát se badatelé, zastávající ten či onen přístup, snažili vytvořit architektury a mechanismy různé úrovně složitosti, ale pokaždé můžeme o výsledku říci, že je nedostačující pro plnohodnotné napodobení obecné schopnosti inteligence.

Jedním ze záměrů této práce bylo prozkoumat dané přístupy právě z hlediska jejich možností a limit při simulaci inteligence. Nutno říci, že výsledky jsou stále uspokojující pouze v rovině simulace speciálních schopností.

V budoucnu může být perspektivní cestou pokus o tvorbu architektury a mechanismů, umožňujících umělému systému možnost inteligentně interagovat s okolním prostředím bez nutnosti zásahu svého tvůrce. Takový systém nemůže být pouze mechanickým manipulátorem se vstupy prostředí, ale naopak, měl by obsahovat schopnost vytvářet si reprezentační systém, který je mu vlastní a zachovává dostatečnou míru shody s prostředím, v němž se pohybuje. Mít možnost vytvářet význam prostředí, chápat obsah, vnímat svět sémanticky.

Tato práce může sloužit jako dobrý odrazový můstek pro další studium problematiky simulace inteligence. Obsahuje shrnutí základních poznatků, umožňujících lepší orientaci v oblasti kognitivních věd. Pokud to bude možné, autor této práce by se rád věnoval problematice i nadále a směřoval své příští studium právě do oblasti tvorby významu, do oblasti sémantických umělých systémů.

LITERATURA

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Methuen Press.

Barrow, J. D. (1996). *Teorie všeho*. Praha: Mladá fronta.

Boden, M. (1977). *Artificial intelligence and natural man* (2nd ed.). Cambridge, MA: MIT Press.

Boden, M. (1988). *Computer Models Of Mind*. Cambridge: Cambridge University Press.

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, MA: MIT Press.

Brustoloni, Jose C. (1991). *Autonomous Agents: Characterization and Requirement*s (Carnegie Mellon Technical Report CMU-CS-91-204). Pittsburgh: Carnegie Mellon University.

Caudill, M., &Butler, C. (2000). *Naturally intelligent systems* (5th ed.). Cambridge: MIT Press.

Crane, T. (2002). *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representation*. Harmondsworth: Penguin Books.

Dupoux, E. (2001). *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*. Cambridge, MA: MIT Press.

Franklin, S.,&Graesser A.(1996) Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. *In Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*. London: Springer-Verlag.

Gazzaniga, M., Ivry, R., & Mangun, G. (1998). *Cognitive Neuroscience: The Biology of the Mind*. New York: W.W. Norton and Co.

Goldszmidt, M.,&Pearl J.(1996) Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence, 84,* 57-112.

Greenspan, S. I. (1996). *The growth of the mind and the endangered origins of intelligence*. Reading, MA: Addison-Wesley.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D, 42*, 335-346.

Haugeland, J. (1997). *Mind Design II: Philosophy, Psychology, Artificial Intelligence* (2nd ed.). Cambridge, MA: MIT Press (A Bradford Book).

Havel, I. M. (2001). Přirozené a umělé myšlení jako filosofický problém. In Mařík, O. (Ed.). (2001). *Umělá Inteligence (3)*. Praha: Academia.

Hayes-Roth, B. (1995). An Architecture for Adaptive Intelligent Systems. *Artificial Intelligence: Special Issue on Agents and Interactivity, 72*, 329-365

Hendriks-Jansen, H. (1996). *Catching ourselves in the act: Situated activity, interactive emergence, evolution, and human thought.* Cambridge, MA: MIT Press (A Bradford Book).

Hofstadter, D. R. (1999). *Gödel, Escher, Bach : an eternal golden braid.* New York : Basic Books.

Hogan, J. P. (1998). *Mind Matters: Exploring the World of Artificial Intelligence.* New York: Del Ray.

Janlert, L. E. (1987). Modeling change: The frame problem. In Z. W. Pylyshyn (Ed.).(1987) *The robot's dilemma: The frame problem in artificial intelligence.* Norwood, NJ: Ablex.

Johnson-Laird, P.N. (1980). Mental Models in Cognitive Science. *Cognitive Science, 4,* 71-115.

Jonák, Z. (2000). Pojem "informace" ve světě sdíleného pojetí skutečnosti. *Ikaros [online], 2000, 2,* Retrieved 2000-02-01 from http://ikaros.ff.cuni.cz/ikaros/2000/c02/veda.htm

Kolodner, J. (1993). *Case-Based Reasoning.* San Mateo: Morgan Kauffman Publishers

Konar, A. (1999). *Artificial Intelligence and Soft Computing Behavioral and Cognitive Modeling of the Human Brain.* New York: CRC Press.

Kosslyn, S. M., &Koenig, O. (1992). *Wet Mind: The New Cognitive Neuroscience.* New York: Free Press.

Kotek, Z. a kol. (1983). *Teória automatického riadenia II.* Bratislava: SNTL, Alfa.

Kozej, P. (2004). *Kognícia bez mentálnych procesov.* Unpublished master's thesis. Univerzita Komenského, Fakulta matematiky, fyziky a informatiky, Bratislava, Slovensko.

Kučerová, H. (2002). *Teorie informace.* Retrieved 2004-05-05 from http://info.sks.cz/users/ku/UIS/inform1.htm

Lorenz, K. (1981). *Foundations of ethology.* London: Springer-Verlag.

Lucas, J.R. (1961). Minds, Machines and Goedel. *Philosophy, 36,* 112-127.

Luger, G.F. (1994). *Cognitive Science: The Science of Intelligent Systems.* Boston, MA: Academic Press.

Mataric, M. (1997). Learning social behaviors. Practice and Future of Autonomous Agents. *Special issue, R. Pfeifer and R. Brooks (Eds.). Robotics and Autonomous Systems,*
20, 191-204.

Mařík, O. (Ed.). (1993). *Umělá Inteligence (1)*. Praha: Academia.

Mařík, O. (Ed.). (1997). *Umělá Inteligence (2)*. Praha: Academia.

Mařík, O. (Ed.). (2001). *Umělá Inteligence (3)*. Praha: Academia.

Mařík, O. (Ed.). (2003). *Umělá Inteligence (4)*. Praha: Academia.

Minsky, M. (1967). *Computation: Finite and Infinite Machines*. Englewood Cliffs, NJ: Prentice-Hall.

Minsky, M. (1968). *Semantic Information Processing*. Cambridge, MA: MIT Press.

Minsky, M., and Papert, S. (1969). *Perceptrons*. Cambridge. MA: MIT Press.

Minsky, M.(1986). *The Society of Mind*. New York: Touchstone.

Newell, A., and Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Peregrin, J. (1997). *Teorie všeho, čeho všeho?*.
Retrieved 2004-09-08 from
http://jarda.peregrin.cz/mybibl/HTMLTxt/362.htm

Peregrin, J. (2002). *Člověk, pro kterého zítra již znamenalo včera*.
Retrieved 2004-09-08 from
http://jarda.peregrin.cz/mybibl/HTMLTxt/447.htm

Peregrin, J. (2003). *Filosofie a jazyk*. Praha: Triton.

Peregrin, J. (2004). *Umělá inteligence bude "zbastlená" - a my jsme počítače, které o tom ani nebudou vědět*.
Retrieved 2004-09-08 from
http://jarda.peregrin.cz/mybibl/HTMLTxt/448.htm

Pěchouček, M. (n.d.). *Úvod do filosofie umělé inteligence*
Retrieved 2004-02-08 from
http://cyber.felk.cvut.cz/gerstner/teaching/kui/kui-phil.htm

Pfeifer, R. ,& Scheier, C. (2001). *Understanding Intelligence*. Cambridge, MA: MIT Press.

Piaget, J. (1998). *Psychologie inteligence*. Praha, Portál 1998

Pícha, M. (2001). *Silná umělá inteligence*. Unpublished master's thesis. Masarykova Univerzita, Filosofická fakulta, Brno, Česká Republika.

Pick, H.L. Jr., P. van den Broek, & D.C. Knill (Eds.) (1992). *Cognition: Conceptual and methodological issues*.Washington, DC: American Psychological Association

Pinker, S. (1997). *How the Mind Works*. New York: W. W. Norton & Company.

Posner, M. I.,&Keele, S.V. (1968). On the Genesis of abstract ideas. *Journal of Experimental Psychology, 77(3,1)*, 353-363.

Pstružina, K. (1998). *Svět poznávání: k filozofickým základům kognitivní vědy*. Olomouc: Nakladatelství Olomouc

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and
organization in the brain. *Psychological Review, 65,* 386-408.

Ross, B.H.: (1996). Category representations and the effect of interacting with instances. *Journal of experimental Psychology:Learning, Memory and Cognition, 22*, 1249-1265.

Russell, S. J., &Norvig, P. (1995). *Artificial intelligence: A modern approach*. Upper Saddle River, NJ: Prentice Hall.

Sedláková, M. (2004). *Vybrane kapitoly z kognitivni psychologie : mentalni reprezentace a mentalni modely* . Praha: Grada.

Shannon, C. E., &Weaver, W. W. (1948). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Sloman, S. A.,& Rips, L. J. (1998): Similarity as an explanatory construct. *Cognition, 65 (2-3)*, 87-101.

Smith, J.C. (ed.).(1990). *Historical Foundations of Cognitive Science*. Dordrecht: Kluwer.

Smullyan, R. (2003). *Navěky nerozhodnuto. Uvod do logiky a zábavny průvodce ke Godelovym objevům*. Praha: Academia.

Sternberg, R. J. (Ed.).(1994). *Encyclopedia of human intelligence*. New York and Toronto: Macmillan.

Sternberg, R. J. (1996). *Kognitivní psychologie*. Praha: Portál.

Sternberg, R. J. (Ed.).(1999). *The nature of cognition*. Cambridge, MA: MIT Press.

Sternberg, R. J. (Ed.).(2000) *Handbook of intelligence*. Cambridge and New York: Cambridge University Press.

Sternberg, R. J. (Ed.).(2001). *Complex cognition*. NewYork: Oxford University Press.

Sutherland, S. (1989). *Macmillan dictionary of psychology*. London, New York : Macmillan. / Cit. podle Crick, F. (1997). *Věda hledá duši: Překvapivá domněnka*. Praha: Mladá fronta.

Šíma, J., Neruda R. (1996). *Teoretické otázky neuronových sítí*. Praha: Univ. Karlova.

Thagard, P. (2001). *Úvod do kognitivní vědy*. Praha: Portál.

Vysoký, P. (2004). C. E. Shannon – průkopník informačního věku. *Vesmír, 83*, 472-475.

Wiedermann, J. (2001). Turing Machine Paradigm in Contemporary Computing [Abstract]. *Mathematics Unlimited - 2001 and Beyond*. London: Springer-Verlag.

Wiener, N. (1947). *Cybernetics*. Cambridge, MA: MIT Press.

Wilson, R. A., & Keil, F. C. (Eds.). (1999). *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.

Winograd, T. (1976). Towards a procedural understanding of semantics. *Revue Internationale de Philosophie, 3,* 117-118.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.