

Boris Aberšek

Metka Kordigel Aberšek

FILOZOFSKA ANALIZA MEHANIZMOV KOGNITIVNEGA MODELIRANJA IN UMETNE INTELIGENCE V IZOBRAŽEVANJU



Univerzitetna založba
Univerze v Mariboru



Univerza v Mariboru

Fakulteta za naravoslovje
in matematiko

Filozofska analiza mehanizmov kognitivnega modeliranja in umetne intelligence v izobraževanju

Avtorja

Boris Aberšek

Metka Kordigel Aberšek

November 2025

Naslov <i>Title</i>	Filozofska analiza mehanizmov kognitivnega modeliranja in umetne inteligence v izobraževanju <i>Philosophical Analysis of Mechanisms of Cognitive Modeling and Artificial Intelligence in Education</i>
Avtorja <i>Authors</i>	Boris Aberšek (Univerza v Mariboru, Fakulteta za naravoslovje in matematiko) Metka Kordigel Aberšek (Univerza v Mariboru, Pedagoška fakulteta)
Recenzija <i>Review</i>	Stanislav Avsec (Univerza v Ljubljani, Pedagoška fakulteta) Mara Cotič (Univerza na Primorskem, Pedagoška fakulteta)
Jezikovni pregled <i>Language editing</i>	Hana Podjed
Tehnična urednika <i>Technical editors</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba) Marina Bajić (Univerza v Mariboru, Univerzitetna založba)
Oblikovanje ovitka <i>Cover designer</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba)
Grafika na ovitku <i>Cover graphic</i>	Abstrakt, avtor: PublicDomainPictures, pixabay.com, 2012
Grafične priloge <i>Graphic material</i>	Aberšek, Kordigel Aberšek, 2025

Založnik **Univerza v Mariboru**
Published by **Univerzitetna založba**
 Slomškov trg 15, 2000 Maribor, Slovenija
<https://press.um.si>, zalozba@um.si

Izdajatelj **Univerza v Mariboru**
Issued by **Fakulteta za naravoslovje in matematiko**
 Koroška cesta 160, 2000 Maribor, Slovenija
<https://www.fnm.um.si>, fnm@um.si

Izdaja
Edition Prva izdaja

Vrsta publikacije
Publication type E-knjiga

Dostopno na
Available at <http://press.um.si/index.php/ump/catalog/book/1057>

Izdano
Published at Maribor, november 2025



© Univerza v Mariboru, Univerzitetna založba
/ University of Maribor, University of Maribor Press

Besedilo/ *Text* © Aberšek, Kordigel Aberšek, 2025

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva-Nekomercialno-Brez predelav 4.0 Mednarodna. / *This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License.*

Uporabnikom je dovoljeno reproduciranje brez predelave avtorskega dela, distribuiranje, dajanje v najem in priobčitev javnosti samega izvirnega avtorskega dela, in sicer pod pogojem, da navedejo avtorja in da ne gre za komercialno uporabo.

Vsa gradiva tretjih oseb v tej knjigi so objavljena pod licenco Creative Commons, razen če to ni navedeno drugače. Če želite ponovno uporabiti gradivo tretjih oseb, ki ni zajeto v licenci Creative Commons, boste morali pridobiti dovoljenje neposredno od imetnika avtorskih pravic.

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

CIP – Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor

004:54(0.034.2)

ABERŠEK, Boris

Filozofska analiza mehanizmov kognitivnega modeliranja in umetne inteligence v izobraževanju [Elektronski vir] / avtorja Boris Aberšek, Metka Kordigel Aberšek. – 1. izd. – E-knjiga. – Maribor : Univerza v Mariboru, Univerzitetna založba, 2025

Način dostopa (URL):

<https://press.um.si/index.php/ump/catalog/book/1057>

ISBN 978-961-299-076-3 (PDF)

doi: 10.18690/um.fnm.4.2025

COBISS.SI-ID 255831555

ISBN 978-961-299-076-3 (pdf)
978-961-299-077-0 (mehka vezava)

DOI <https://doi.org/10.18690/um.fnm.4.2025>

Cena
Price Brezplačni izvod

Odgovorna oseba založnika prof. dr. Zdravko Kacič,
For publisher rektor Univerze v Mariboru

Citiranje Aberšek, B., Kordigel Aberšek, M. (2025). *Filozofska analiza*
Attribution *mehanizmov kognitivnega modeliranja in umetne inteligence v*
izobraževanju. Univerza v Mariboru, Univerzitetna založba. doi:
10.18690/um.fnm.4.2025

Kazalo

Predgovor.....	1
Izvleček.....	3
Uvodnik in zahvala	15
1 Uvod.....	17
1.1 Povezovanje kognitivne znanosti in filozofije.....	23
1.2 Zakaj umetna inteligenca potrebuje filozofijo?	24
1.3 Kognitivno modeliranje in generativizem	34
1.3.1 Razvoj sodobnega izobraževanja skozi čas.....	34
1.3.2 Od digitalne pedagogike do generativizma	37
1.3.3 Generativno učenje.....	40
1.3.4 UI in izobraževanje.....	42
1.4 Zgodovinski oris umetne inteligence	44
1.5 Sklep	52
2 Osnovna filozofska in kognitivna izhodišča.....	57
2.1 Zgodovinski razvoj filozofije duha.....	58
2.2 Novejši filozofski pogledi na problem duha in telesa.....	62
2.2.1 Vprašanje redukcije.....	66
2.2.2 Trenutno stanje redukcije v kognitivni znanosti.....	67
2.2.3 Vprašanje emergence.....	69
2.3 Zgodovinski oris modeliranja v kognitivni znanosti.....	81
2.4 Diskretni sistemi nasproti zveznim	83
2.4.1 Emergenca in redukcija v dinamični kognitivni znanosti	84
2.5 Dinamični sistemi in »utelešena kognicija«	95
2.6 Utelešena kognicija	97
2.7 Situacijska kognicija in konekcionizem.....	99
2.7.1 Prednosti in slabosti dinamičnih modelov	99
2.8 Inteligenca	101
2.8.1 Inteligentnost, ustvarjalnost in nadarjenost	104
2.8.2 Umetna inteligenca	107
2.9 Sklep	115
3 Kognitivna znanost	119
3.1 Razvoj kognitivne znanosti.....	120
3.1.1 Rojstvo kognitivne znanosti.....	121
3.2 Kognitivna teorija učenja.....	124
3.2.1 Model procesiranja informacij.....	125
3.2.2 Izvajanje kognitivnega procesa.....	127
3.2.3 Pridobivanje novih spoznanj (kaj je učenje?)	128

3.3	Učenje in kognitivno modeliranje.....	130
3.3.1	Informacijsko-teoretična ali kibernetična didaktika	131
3.4	Kognitivne arhitekture	141
3.4.1	Klasična kognitivna znanost in simbolni modeli	142
3.4.2	Kognitivno modeliranje.....	143
3.4.3	Razvoj kognitivne arhitekture.....	146
3.4.4	Filozofija uma in STEAM izobraževanje	148
3.4.5	Tehnološka in inženirska pismenost	151
3.4.6	Učne strategije temelječe na možganskih procesih	155
3.5	Sklep	160
4	Nevroznanost.....	163
4.1	Nevroni	163
4.1.1	Dinamični nevron	165
4.1.2	Akcijski potencial.....	166
4.1.3	Od električnih do kemijskih reakcij.....	167
4.2	Kognitivna nevroznanost	169
4.2.1	Kognitivna nevroznanost in vzratno inženirstvo	171
4.2.2	Nevrokognitivna znanost in nevro-morfno inženirstvo.....	172
4.3	Nevroznanost, učenje in spomin	175
4.3.1	Plastičnost.....	175
4.3.2	Organizacija spomina	177
4.4	Celostni pristop k raziskovanju v kognitivni nevroznanosti	179
4.5	Sklep	180
5	Simbolni sistemi in konekcionistični modeli	181
5.1	Predhodniki konekcionizma	183
5.2	Genetski algoritmi za simulacijo evolucije (razvoja)	185
5.3	Mrežni sistemi.....	187
5.4	Konekcionistične mreže	190
5.4.1	Perceptron	192
5.5	Konekcionizem in simbolni sistemi	194
5.6	Dinamika konekcionističnih modelov	194
5.7	Prepoznavanje vzorca in znanje.....	198
5.8	Ponovno rojstvo – novi konekcionizem	199
5.9	Sodobne nevrnske mreže.....	200
5.10	Umetne nevrnske mreže.....	202
5.11	Paradoks moderne računske tehnologije.....	203
5.12	Primer: matematična uganka.....	204
5.13	Nevronske mreže in umetni možgani.....	206
5.14	Gradnja silicijevih možganskih vezij	207
5.14.1	Silicijeva mrežnica	208
5.15	Računalniška nevroznanost.....	209
5.16	Kognitivizem in konekcionizem	211
5.16.1	Simboli ali mreže v kognitivni znanosti.....	211
5.16.2	Bionevronske mreže in konekcionistični modeli.....	213
5.17	O interdisciplinarnosti v kognitivni znanosti	217
5.18	Sklep	221
6	Kognitivne hibridne arhitekture	223
6.1	Hipoteza fizikalnega simbolnega sistema.....	223
6.2	Manipulacije s simboli.....	227
6.2.1	Primer – Eliza	228
6.2.2	Kako Eliza programsko deluje	230

6.3	Od simbolnih sistemov do jezika misli.....	244
6.3.1	Intencionalni realizem in vzročnost na podlagi vsebine	245
6.4	Računalniški model duha (zveza med sintakso in semantiko)	247
6.5	Primeri simbolnih arhitektur.....	250
6.5.1	SOAR	252
6.6	Kognitivni hibridni sistemi.....	255
6.6.1	ACT-R: kratka zgodovina.....	257
6.6.2	Hibridna arhitektura na primeru ACT-R.....	262
6.6.3	ACT-R: teorija	264
6.6.4	Primer: algebra.....	266
6.6.5	Stopnje usvajanja veščin	267
6.6.6	Računsko procesiranje v ACT-R	269
6.6.7	Simbolne in podsimbolne zveze	270
6.6.8	Simbolne in podsimbolne strukture v deklarativnem modulu.....	271
6.6.9	Simbolne in podsimbolne strukture v procesnem modulu	273
6.6.10	Končna refleksija o simbolnih podsimbolnih razlikah	274
6.6.11	ACT-R in KIS.....	277
6.6.12	Modeliranje možganov v ACT-R.....	278
6.6.13	Kognitivno modeliranje.....	281
6.7	Sklep	283
7	Dinamični sistemi	287
7.1	Definicija stanja sistema in prostor stanja.....	289
7.1.1	Prostor stanja	289
7.1.2	Osnove teorije nihanj	293
7.2	Modeliranje dinamičnih sistemov	295
7.2.1	Odnos med modelom in originalom.....	296
7.2.2	Matematični model.....	297
7.3	Teorija dinamičnih sistemov in kognitivna znanost.....	299
7.3.1	Prva rešitev – simbolno računski centrifugalni regulator.....	301
7.3.2	Druge rešitev – fizikalni dinamični centrifugalni regulator.....	302
7.3.3	Kritika prve rešitve	305
7.4	Hipoteze dinamičnih sistemov v kognitivni znanosti.....	306
7.4.1	Matematične zveze.....	308
7.4.2	Matematični model nevrona: Hodgkin-Huxleyjev model.....	309
7.4.3	Matematični model nevrona: Fitzhugh-Nagumov model.....	314
7.5	Primer dinamičnega modeliranja kognitivnih zmožnosti.....	315
7.5.1	Motorična kontrola: osnovni model.....	315
7.5.2	Motorična kontrola: Haken–Kelso–Bunz model	317
7.5.3	Model možganskih aktivnosti.....	321
7.6	Ocenjevanje modelov dinamičnih sistemov	331
7.7	Možnosti realizacije umetnih možganov	333
7.7.1	Izhodišča	333
7.7.2	Modeli posnemanja bioloških možganov.....	334
7.7.3	Kvantno računalništvo.....	337
7.7.4	Teorija univerzalnega računanja.....	339
7.7.5	Projekt modri možgani	341
7.8	Sklep	344
8	Resničnost, prostor in čas	347
8.1.1	Poenotena resničnost	355
8.2	Problem duha in telesa: osnovne rešitve	357
8.3	Redukcija.....	357
8.3.1	Modeli teoretične redukcije.....	357

8.3.2	Novi trendi v redukciji: algoritemska redukcija.....	367
8.4	Sklep	369
9	Sklep	371
	Literatura	381
	Stvarno kazalo.....	391

Predgovor

Ta monografija temelji na disertaciji za pridobitev doktorata znanosti na Univerzi v Mariboru, je v ponedeljek, **14. julija 2014**, po pooblastilu rektorja univerze, prof. dr. Danijela Rebolja, na podlagi odločitve Doktorske komisije, javno zagovarjal

Boris Aberšek

This monograph, based on the dissertation to obtain the degree of doctor at the University of Maribor, was publicly defended by

Boris Aberšek

on Monday, the 14th of July, 2014, on the authority of rector, prof. dr. Danijel Rebolj, on account of the Doctorate Board.

Izvleček

Tehnologija je dvoje, je vzrok in posledica za vse hitreje spreminjajočo se družbo. Ko danes govorimo o tehnologiji, govorimo predvsem o inteligentnih sistemih, o tako imenovanih kibernetičnih sistemih, pri katerih sistem vsebuje, če govorimo v jeziku tehnologije, *strojno opremo*, stroje in naprave (človeško telo), in *programsko opremo* – inteligenco – duha, kar bi v jeziku humanistike lahko poimenovali tudi *občutke*, *intencionalna stanja*, *čustva* ipd. Mnogi problemi, ki so povezani s sodobnimi tehnologijami, so zato povezani tudi z (umetno) inteligenco in učenjem. Za reševanje problemov v zvezi z umetno inteligenco potrebujemo predvsem splošna znanja o svetu, družbi in ljudeh, kar je *filozofski problem*. Zato je za vse, ki morajo sprejemati odločitve s področja uporabe umetne inteligence, izjemno pomembno, da *razumejo* osnovne splošne mehanizme inteligence – filozofijo duha, tj. kako deluje naravna, kako deluje človeška inteligenca, kako se človek uči in prilagaja okolju, kar je še posebej pereče na področju vzgoje in izobraževanja.

V tej knjigi smo se pretežno posvečali področju filozofije duha in učenju, torej enemu od njenih izjemno pomembnih elementov, kjer je vse od najzgodnejših začetkov pojava človeka kot mislečega »stroja« pa vse do danes prisotna dvojnost njegovega pojasnjevanja, ki se zrcali v dveh smereh:

- *Misterianistični*, ki zagovarja tezo, da je duh nekaj edinstvenega v svoji neponovljivosti in da je misterij v nekih lastnostih, svojstvenih samo študijam

duha. Izhajajoč iz tega lahko sklepamo, da je duševnost povsem drugačna od vsega ostalega v svetu in mora zato imeti poseben ontološki status.

- *Naturalistični*, ki zagovarja tezo, da duh lahko, ali ga celo moramo »naturalizirati,« in to tako, da so lahko duhovna stanja in procesi opisljivi z jezikom znanosti, npr. fizike, kognitivne znanosti, nevroznanosti ali jezikom drugih naravoslovnih ved. Po tem pristopu sprejemamo, da je človek kot bitje del narave ter da so procesi, ki se dogajajo v njem, vključno z duševnimi procesi, prav tako del te narave.

Oba pristopa skušata najti resnico, vendar uporabljata popolnoma drugačne metode in orodja ter interpretirata svoje dosežke na popolnoma različnih osnovah in izhodiščih. V raziskavi *zagovarjamo naturalističen pristop*, temelječ na sodobnih naravoslovnih metodah in znanostih ter redukcionizmu v teoriji duha, s poudarkom na strojnem učenju kot paradigmatiskem primeru algoritemskega reducibilnega mentalnega procesa z možnostjo aplikacije na človeka. Tako kot lahko pri stroju, ki se uči, učenje skrbimo na algoritme oziroma fizikalne procese, lahko mentalne procese v človeku zapišemo kot dinamične modele in jih nanje tudi reduciramo. Takšen redukcionistični pristop v okviru filozofije, kognitivne znanosti, nevrobiologije, nevroračunalništva in umetne inteligence pogosto zastavlja vprašanje, ali in kakšna je zveza med razvojem kognitivnega nevromodeliranja (modeliranje – izdelava delovanja možganov s pomočjo sodobnih računalniških tehnologij) in nevroračunalništva (računalniške simulacije nevronske mreže oz. t. i. umetnih nevronske mreže, ki v kognitivni znanosti prevzemajo vlogo procesnih modelov možganskih in duševnih procesov). Do sedaj še ni bil dan natančen odgovor na tri, za raziskave na teh področjih bistvena vprašanja, in sicer:

1. Ali lahko celovito kognitivno nevromodeliramo in simuliramo delovanje možganov s pomočjo sodobnih računalniških tehnologij?
2. Ali lahko razvoj nevroračunalništva in umetne inteligence pripelje do nadomeščanja človeka in njegove naravne inteligence (*neposredno* – humanoidni robot – stroj, ki opravlja človeško delo, ali *posredno* – računalniški »učitelj«, inteligentni tutor/agent, ki opravlja le človekove mentalne funkcije)?
3. Ideja, ki se ponovno bolj intenzivno pojavlja v zadnjih desetletjih in ki jo zagovarjamo in dokazujemo tudi v tem delu, ali lahko človeka, tj. njegovo

mentalno življenje, reduciramo na dinamični sistem, skratka, ali obstaja takšen sistem diferencialnih enačb, s katerimi bi ga lahko zapisali.

V tej knjigi smo poskušali osvetliti odgovore predvsem na ta vprašanja.

Delo je zasnovano po zgodovinskem principu, kjer od samih začetkov filozofskega razmišljanja o duhu v delu zakoračimo preko bližnje preteklosti in aktualne sedanjosti v bližnjo in daljno prihodnost ideje o umetnem človeku in pri tem zagovarjamo naturalističen pristop, temelječ na sodobnih naravoslovnih metodah in znanostih, ter redukcionizem v teoriji duha s poudarkom na možnosti algoritemske redukcije mentalnih procesov v človeku na dinamične modele.

V *prvem poglavju* smo si zastavili raziskovalni problem, ki je imel temelje v znanstveni fantastiki preteklosti (Kordigel, 1994) in je postal resničnost danes, in zastavili nekaj izhodiščnih vprašanj o obstoju umetne inteligence (UI) danes, npr. generativne umetne inteligence (GUI) in velikih jezikovnih modelov, ter razvoju UI v prihodnosti in se pri tem osredotočil predvsem na vprašanje: Zakaj UI potrebuje filozofijo? Na temelju teh vprašanj smo definirali cilje in smernice raziskovalnih področij. Celotno prvo poglavje temelji na razmišljanjih izpostavljenih v *Recepcija slovenske znanstvenofantastične literature* (Kordigel, 1991) in delu *Znanstvena fantastika* (Kordigel, 1994).

V *drugem poglavju* smo se osredotočili predvsem na osnovna filozofska izhodišča filozofije o duhu in se pri tem na temelju zgodovinskega razvoja posvetili novejšim filozofskim pogledom na duha ter se dotaknil vprašanj redukcije in emergenc. Temu je sledil oris modeliranja, pri čemer je bilo izpostavljeno, da na področju modeliranja duha obstajajo trije prevladujoči pristopi oz. modeli: *diskretni*, *zvezni* in *dinamični* ali povedano z drugimi besedami: *simbolni*, *konekcioniistični* in *dinamični* modeli.

V *tretjem poglavju* se poglobimo v kognitivno znanost in povezavo le-te s kognitivno teorijo učenja ter s procesi in modeli procesiranja informacij. Vseskozi se pri tem poudarja, da je tako kot nevroznanost tudi kognitivna znanost izrazito interdisciplinarna in pogled s samo enega zornega kota, torej iz zornega kota ene posamične vede, ne more dati pravih rezultatov, najpogosteje pa daje celo popolnoma napačne in zavajajoče rezultate.

V *četrtem poglavju* smo se dotaknili področja nevroznanosti, ki se je predvsem zaradi izjemno močnega tehnološkega razvoja na področju empiričnih raziskav in metod zbiranja podatkov (fMRI, EMR, BOLD ipd.) v zadnjih desetletjih izrazito intenzivno razvijala. Zaključimo, da so možgani v svojem bistvu bionevronska mreža, katere osnovni gradniki so nevroni, zato jim posvetimo pomemben del tega poglavja in raziskujemo, kako se informacije v možganih (nevronih) prenašajo, kako se signali pretvarjajo od električnih do kemičnih reakcij in kako se pri tem ustvarja ustrezen akcijski potencial, ki definira dostopnost do določenih informacij in verjetnost njihovega priklica v spomin.

V naslednjih poglavjih se podrobneje lotimo različnih pristopov modeliranja duha. Tako se v *petem poglavju* osredotočimo na začetku na kognitivne fizikalne simbolne (diskretne) sisteme in nato v nadaljevanju na kognitivne konekcionistične modele. Nakazane so smernice razvoja in prikazane možnosti izgradnje silicijevih nevronske mreže kot osnove umetnih možganov. Izpostavimo paradoks, da danes še vedno poskušamo modelirati delovanje duha (delovanje možganov), ki izkazuje izrazito zvezni proces delovanja na konvencionalnih, digitalnih računalnikih, ki delujejo izključno po diskretnem principu.

Šesto poglavje je namenjeno kombinaciji konekcionističnih in simbolnih sistemov, ki jih imenujemo tudi hibridni sistemi. V nadaljevanju si ogledamo nekatere trenutno najbolj delujoče simbolne arhitekture. Poudarimo, da vsaj bližnja prihodnost zagotovo ni ne črna in ne bela, oz. ni ne mrežna, konekcionistična, ne fizikalno simbolna, temveč je nujna njuna kombinacija. To hipotezo dokazujemo na osnovi trenutno verjetno najbolj kompleksne in tudi najbolj delujoče kognitivne hibridne arhitekture, torej na primeru ACT-R ki je po svojem principu in po svojem delovanju danes še najbližje delovanju človeških možganov, kar potrjujejo tudi različne empirične raziskave v nevroznanosti.

V *sedmem poglavju* uporabimo teorijo dinamičnih sistemov in ob tem izhajamo iz teze, da je vesolje dinamični sistem. Upošteva je to, je tako tudi vsak sistem, podsistem ali element dinamičnega sistema dinamični sistem in to vse do najmanjših delcev. Iz tega sledi, da je tudi človek in vse, kar se nanj nanaša, dinamični sistem in da lahko za njegovo razlago uporabljamo isto teorijo, tj. teorijo dinamičnih sistemov, kar dokažemo z definiranjem možganskih procesov kot dinamičnega sistema. Nato splošno teorijo dinamičnih sistemov prevedemo na hipotezo dinamičnih sistemov v

kognitivni znanosti in poglavje zaključimo z zagovorom in dokazovanjem možnosti redukcije duha na dinamične sisteme ter dokazovanjem uporabnosti predlagane teorije tako na enostavnejših primerih kot tudi na kompleksnejšem primeru dinamičnega modela možganskih aktivnosti. Dotaknemo se vprašanja, ali lahko na diskretnih sistemih zares uspešno obravnavamo zvezne sisteme, kar možgani zagotovo so, ali smo z današnjo računalniško tehnologijo že dosegli njen »rob« in bo potrebno za naslednji korak narediti miselni in tehnološki preskok, t. i. »kvantni preskok«.

Delo zaključimo z *osmim poglavjem*, kjer se podrobneje osredotočimo na izhodiščna vprašanja, redukcijo in z njo povezano zavest. Zavest so v zgodovini tolmačili in si jo razlagali na različne načine. Po eni strani se nam zdi, da je zavest (duševnost) povsem drugačna od vsega ostalega v svetu in mora zato imeti poseben ontološki status. Po drugi strani pa sprejemamo, da je človek biološko bitje in da je del narave ter da so procesi, ki se dogajajo v njem, vključno z duševnimi procesi, naravni procesi. Ker v tem delu zagovarjamo naturalističen pristop, je še najbolj obetavna razlaga zavesti verjetno kvantna razlaga in možnost redukcije le-te na kvantni nivo, torej razlaga zavesti po teoriji kvantne mehanike. Ta razlaga pa je tudi osnova argumentacije možne redukcije, ki je izhodiščna teza in rdeča nit raziskave. Predstavljeno redukcijo, temelječo na dokazu možnosti kvantne razlage duha kot najbolj kompleksnega pojava v kogniciji, smo poimenovali algoritemska redukcija in je popolnoma nova vrsta redukcije. Je hibridna redukcija, ki tvori kombinacijo Nagelovske in funkcionalistične redukcije. Opišemo jo lahko kot dvostopenjski proces, pri katerem v prvem koraku za osnovo uporabim funkcionalistično redukcijo, v drugem koraku pa Nagelovsko redukcijo. V prvem koraku morajo biti kognitivna stanja opisana v pojmi njihovih vzročnih zvez (kar je prvi korak v funkcionalistični redukciji), v drugem koraku pa zapišemo diferencialne enačbe, ki opisujejo oboje, kognitivno in možgansko dinamiko, in morajo imeti enako obliko, torej morajo biti izomorfne (kar je prvi korak Nagelovske redukcije), to pa pomeni, da lahko duha reduciramo na možganske procese preko matematičnih algoritmov.

Za konec odgovorimo na še najpogostejše zastavljeno vprašanje: *Za kaj in kje mi bo to koristilo?* Naj na kratko povzamemo: kognitivno modeliranje je v šoli možno in potrebno tako v splošni vsakdanji praksi pri ustvarjanju (inteligentnih) učnih okolij, kar počnemo že danes, vsak dan, pri pripravi učnih načrtov in scenarijev, pa tudi pri pripravi različnih učnih gradiv. Na osnovi že znanih izsledkov kognitivne- in nevro

znanosti vemo, kako delujejo in se učijo naši možgani in kako poteka usvajanje znanj v učnem procesu. Ob upoštevanju teh znanj pa lahko na dokaj enostaven način oblikujemo sodobna učna okolja in s tem neposredno inoviramo učni proces in vanj vnašamo izsledke sodobnih raziskav.

Kognitivno modeliranje pa se lahko izvede tudi v bolj kompleksni obliki, npr. ko šolo ali pouk obravnavamo po teoriji dinamičnih in kompleksnih sistemov kot dinamičen proces, z namensko razvitimi diferenčnimi in diferencialnimi (gibalnimi) enačbami, s katerimi bolj podrobno opisujemo tako človekove kognitivne procese kot tudi procese v šoli, kot je na primer uvajanje UI v proces izobraževanja. To je vizija prihodnosti, ko bodo ta znanja bolj razširjena in splošno dostopna ter bodo tudi del izobraževalnega procesa učiteljev. Sedaj, v tej knjigi, pa je to nekakšna vizija prihodnosti.

No, in na koncu: če bi želeli izvesti skoraj nemogoče, razdelitev avtorstva v tem skupnem delu, razdelitev celovite prepletenosti in dolgih večernih diskusij, v eksplicitno označevanje, potem, kot je bilo že omenjeno, delo zares temelji na izhodiščih doktorata (Aberšek, 1012), za katerega sta bila zaslužna predvsem mentorja Bojan Borstner in Janez Bregant. A moram izpostaviti še svetovalca iz ozadja, Metko Kordigel Aberšek. To pomeni, da je v vsaki uporabljeni besedi mojega doktorata tudi delček nje. Ob tem lahko izpostavim še, da je bila ključna pri ustvarjanju predvsem Povzetka, Uvoda, Poglavja 1 in Poglavja 3, kar daje tej monografiji tudi pomembno uporabno vrednost, ki lahko služi tudi kot učno gradivo predvsem na podiplomskih študijskih programih.

Summary

Technology is both a cause and a consequence for an ever faster changing society. Speaking about technology today, we speak above all about intelligent systems, so called cybernetic systems, the system which contains, if we speak in technological language, *engine equipment*, machines and devices (the human body), and *program equipment* – intelligence – mind (in the language of humans), also *feelings*, *intention conditions*, *emotions*, etc. Many problems connected with modern technologies are therefore also connected with (artificial) intelligence (AI). To solve problems connected with artificial intelligence, above all we need common knowledge about the world, society, and human beings, which is a *philosophical problem*. Therefore, it is very important for everyone who has to take decisions in the field of artificial intelligence *to understand* the basics of general mechanisms of intelligence – philosophy of the mind, to understand the way natural and artificial intelligence functions especially in the area of education.

In this book we focus mainly on the field of philosophy of the mind and learning, one of its particularly important elements, where from the early beginning of thinking about a human as a thinking “machine” to the contemporary teaching about this problem, we continuously observe two ways of explaining:

- A *mysterian* one, which posits that the mind is something unique and is a source of some mysteries which are characteristic only of studies of the mind. Consequently, we could conclude that the mind is completely different from anything else in the world and must therefore have a special ontological position.
- A *naturalistic* one, which posits that the mind can, or even must be, “naturalised”, so that mental conditions and processes are describable with the language of science: physics, cognitive science or languages of other natural sciences. According to this approach we expect that a human being is a part of nature, and the processes which happen in the human being, including mental processes, are also a part of nature.

Both approaches try to find the truth, but both of them use totally different methods and tools and interpret findings from totally different theoretical backgrounds and origins. In this thesis, we advocate a naturalistic approach, based on current natural sciences, their methods, and reductionism in the theory of the mind with an emphasis on engine learning as a paradigmatic case of an algorithmically reducible mental process with the possibility of application on human beings. In this case, when the machine is learning, we can reduce this learning to algorithms or physical processes. In addition, mental processes in human can be reduced and written as dynamic models. Such reductionist approaches in the frame of philosophy, cognitive science neurobiology, neuro-computer science and artificial intelligence often pose the question of what connection there is, if any, between the development of cognitive science and neuro-computer science. So far, no final answer has been found to three key questions in this field:

1. Can we entirely (neuro)model and simulate brain functioning with the help of current computer technologies?
2. Can the development of neuro-computer science and artificial intelligence lead to the replacement of human beings and their natural intelligence (*directly* – a humanoid robot, engine or machine which does humans’ jobs; or *indirectly* – a computer “teacher”, an intelligent tutor, which performs only humans’ mental functions).
3. Can a human – his mental life – be reduced via mathematical algorithms to brain processes? For example, is there such a system of differential equations by which

the mind could be written, and which could be used to bridge a gap between two different sciences, i.e. psychology as a special science and physics as a basic science.

In this book we will try to answer these questions.

The book is designed along the historical principle, where from the early beginnings of philosophical thinking about the mind we continue through the past and present to the near and distant future of the idea of artificial human beings and argue for a naturalistic approach, based on current natural sciences and their methods, and reductionism in the theory of the mind with an emphasis on the possibilities of algorithmic reduction of mental processes in human beings on dynamic models.

In the *first chapter* we define the research problem, some basic questions about the existence of artificial intelligence (AI) and AI development in the future and focus above all on the question of why AI needs philosophy. On the basis of these questions we define the goals and guidelines of both research fields.

In the *second chapter* we focus mainly on basic philosophical questions about the mind and focus on the basis of historical development on the recent philosophical views of the mind. In doing this, we touch the questions of reduction and emergence. After that we focus on the field of mind modelling. In the field of mind modelling we can observe three main approaches: *discrete*, *continuous* and *dynamic* or, in other words: a symbolic, connectionist and dynamic system.

In the *third chapter* we speak about cognitive science and its connection with the cognitive theory of learning and models of information processing. Throughout, we point out that, as with neuroscience, cognitive science is also a particularly interdisciplinary science. Consequently, an observation from one single point of view, from the point of view of a single science, would give totally wrong and misleading results.

In the *fourth chapter* we touch on the field of neuroscience, which has developed very quickly in the last decades (a consequence of speedy technological development in the fields of empirical research and methods of data collecting (fMRI, EMR, BOLD, etc.)). We conclude this chapter with the fact that the brain is a bio-neuron net -- the

essential parts of which are neurons. The most important part of chapter three is dedicated to neurons. We explain how information is transmitted between neurons, how signals are transformed from electric to chemical reactions and how active potentials are created, all of which define the access to particular information and the probability of its recall into memory.

In the next chapters we discuss different approaches to the modelling of the mind in greater length. In the *fifth chapter* we examine the cognitive physical symbolic (discrete) systems and then continue with cognitive connectionist models. We indicate the guidelines of the development and present the possibilities of the building of neuron nets as the basis of an artificial brain. We point out the paradox of still trying to model the functioning of the mind (brain functioning) that shows a distinctive continuous process of functioning on conventional, digital computers, which function on a discrete principle.

The *sixth chapter* is focused on a combination of connectionist and symbolic systems, which we also call hybrid systems. After that we present some current best functioning symbolic architectures. We point out that the near future is definitely not black and white. It is not networks, neither connectionist nor physical symbol. The near future is inevitably a combination of both of them. We built the proof of this hypothesis on the basis of currently best functioning cognitive hybrid architecture, for example ACT-R, which is, in principle and according to its functioning, the closest case to the functioning of the human brain. That was confirmed by empirical research in the field of neuroscience.

In the *seventh chapter* we use the theory of dynamic systems and, in doing so, defend the theory that the universe is a dynamic system. Originating from this, we derive the conclusion that every system, subsystem or element of a dynamic system – down to its smallest parts is also a dynamic system. Consequently, a human being and everything in connection with him is also a dynamic system. That is why we can use the same theory, the theory of dynamic systems, for the explanation of a human being. We prove this with the definition of brain processes as a dynamic system. We continue with the translation of common theories of dynamic systems to the hypothesis of dynamic systems in cognitive science and conclude the chapter with the defence and proof of the possibility of the reduction of the mind with dynamic systems and proving the usability of the suggested theory in simple cases as well as

in the complex case of a dynamic model of brain activity. We also address the question of whether continuous systems (as the brain definitely is), as with discrete systems, can be treated successfully. This question is connected with the next question of whether a computer technology we currently possess or which is on the point of development is its edge and whether, for the next step, new thinking and a technological “jump over” is needed.

We finish the book with *chapter eight*, where we focus on the origin questions, reduction and with reduction connected consciousness. Throughout history consciousness has been explained in different ways. On one hand, consciousness seems to be completely different from anything in the world and must therefore have a special status. On the other, we can easily accept that a human being is a biological creature, a part of nature and the processes which are happening “in” him – including mental processes – are natural ones. Since we argue for a naturalistic approach, the most promising explanation of consciousness to us is the quantum explanation, and the existing possibility of its reduction to the brain – the explanation of consciousness according to the theory of quantum mechanics. This explanation is the basis of argumentation for possible reduction. The introduced reduction, which we named the algorithmic reduction (AR), is based on the proof of possibility of a quantum explanation of the mind – the most complex emergence in cognition. It is a hybrid reduction forming a combination of Nagel and functionalist reduction. We can describe the algorithmic reduction as a two phase process in which the first step uses the functional reduction and the second step the Nagel reduction. This two-step reduction process of cognitive skills may be what Kelso also suggested. Briefly, if we can find (and this may be what we've demonstrated and proved in the previous chapters) the same differential equations which control high-level (behavioural) and lower level (neuro-physiological) processes, it could be ascertained that the processes are isomorphic and therefore a cognitive behaviour can be reduced to the brain processes, or in other words, the mind to the body.

Finally, let us answer the most frequently asked question: What and where will it be helpful for us? In summary, cognitive modelling is possible and necessary in school, as well as in general everyday practice when creating intelligent learning environments, which we already implement today in everyday practice, when preparing lesson plans, and when preparing various teaching materials. Based on the

well-known findings of cognitive and neuroscience, we know how our brain works and learns and how knowledge acquisition occurs in the learning process. Taking this knowledge into account, we can easily shape modern learning environments today, thus directly innovating the learning process and bringing in modern research results.

Cognitive modelling can also be performed in a more complex form. For example, like a school, we treat lessons according to the theory of dynamic and complex systems as a dynamic process, with specially developed differential and differential (dynamic) equations, with which we describe human cognitive processes and processes in more detail in school. Such cognitive modelling, a vision of the future, when this knowledge will be more widespread and generally accessible and will also be part of the educational process of teachers, is written in this book, so it is a kind of vision of the future.

And finally, if we wanted to carry out the almost impossible work of dividing authorship in this joint work, dividing the entirety of the entanglement and lengthy evening discussions into explicit marking, then, as already mentioned, the work is really based on the starting points of the PhD thesis (Aberšek, 1012), for which the mentors Bojan Borstner and Janez Bregant were mainly responsible. Still, I must also point out that I had the advisor from the background, Metke, Kordigel Aberšek, This means there is a fragment of it in every word used in my PhD. At the same time, I can point out that it was key in creating the Summary, Introduction, Chapters 1 and 3, which also gives this monograph significant practical value and can serve as studying source, especially in postgraduate study programs.

Uvodnik in zahvala

Mladi možgani pridejo v šolo, da se učijo; starejši možgani pridejo v šolo, da jih učijo. Vsaj to je temeljna ideja. Utemeljitev za pisanje te knjige so bila vprašanja: ali nam kognitivna nevroznanost lahko pove kaj o tem, kako se mladi možgani učijo od starejših možganov in kako naj bi torej starejši možgani učili mlajše za dosego optimalnega rezultata?
(Geake, 2009)

Sistem znanja med mlajšimi in starejšimi možgani se spreminja tako, da se znanje porazdeli med ene in druge, in v idealnem sistemu vsi pridobivajo. Po analogiji lahko podobno rečemo za sistem prenosa znanja med enimi možgani (učitelji) in več možgani (učenci v razredu), kar se dogaja v tradicionalnih šolskih okoljih. Ker to velja tako za biološke kot nebiološke sisteme, se lahko zastavi vprašanje, kako bi bilo možno optimalno organizirati tovrsten prenos tudi med obema, torej med biološkimi in nebiološkimi sistemi, tj. v interakciji med naravnimi (učiteljevimi in učenčevimi) in »umetnimi« možgani, ki temeljijo na umetni inteligenci? Nadalje se lahko vprašamo, ali so raziskave na tem področju osnova za nadaljnji razvoj človeštva kot celotne družbe in izobraževalnega sistema kot temelja za razvoj te družbe. Na ta temeljna, izhodiščna vprašanja, ki bodo porajala vrsto novih, bomo iskali odgovore v knjigi, ki je pred vami.

Da sva imela možnosti priti do vseh v tej knjigi zapisanih vprašanj, spoznanj in potencialnih rešitev, bi se želela zahvaliti vsem, ki so naju kakorkoli spodbujali in nama omogočili dobiti ta spoznanja, predvsem pa Ministrstvu za izobraževanje,

znanost in šport RS, ESS, ARRS in številnim Phare, Tempus, Leonardo da Vinci in Erasmus+ projektom za ideje, raziskave in njihovo finančno podporo. Vsem najinim kolegom s Filozofske fakultete, Fakultete za naravoslovje in matematiko in Pedagoške fakultete Univerze v Mariboru kot tudi kolegom z Zavoda Antona Martina Slomška iz Maribora, za njihove prispevke in koristne komentarje. Hvala Bojanu Borstnerju in Janezu Bregantu za vse naše pogovore, v katerih smo si izmišljevali vprašanja in skupaj preverjali naše intuitivne odgovore v okviru mojega drugega doktorata *Filozofska analiza mehanizmov naravnega mišljenja in umetne inteligence*. Prav to delo pa je bila osnova za razmišljanje o tem, kako nekatera dokaj zamotana dejstva in spoznanja prevest v nekoliko bolj razumljiv jezik. Zato moram izraziti še svojo osebno zahvalo, saj mi je na tej poti, tako pri doktoratu kot tudi pri nastajanju te knjige, na vsakem koraku vedno stala ob strani kolegica Metka Kordigel Aberšek. Skupaj se nama je porodila ideja za nastanek te knjige, ki je skupno delo, vsak delček pa je tudi rezultat skupnega razmišljanja. Zato ji gre še posebna hvala, brez njenih prizemljenih nasvetov bi zagotovo zaplaval v svet, ki bi ga razumeli le redki. Vsak potrebuje svoje sidro in Metka je bila to sidro zame. Skupaj pa se zahvaljujeva tudi vsem najinim študentom za navdih in pomoč.

Avtorja se želiva zahvaliti tudi vsem, ki so poskrbeli za kakovost te knjige. Nenazadnje, pa gre zahvala predvsem staršem in otrokom ter učiteljem, ki so naju navdihovali, podpirali in nama dali priložnost, da sva to, kar sva.

1 Uvod

Tehnologija je dvoje, je vzrok in posledica za vse hitreje spreminjajočo se družbo. Ko govorimo o tehnologiji, govorimo o inteligentnih sistemih, o t. i. kibernetičnih sistemih, pri katerih (naj bodo to živalski, mehanski, ekonomski, socialni, človeški ipd.) sistem vsebuje, ob vsem ostalem, če govorimo v jeziku tehnologije, *strojno opremo* – hardware, stroje in naprave (človeško telo) in *programsko opremo*, software – inteligenco – duha ter mehanizem povratnih informacij (sistem dobiva informacije o tem, kaj se v njem dogaja, in se na podlagi teh informacij uči ukrepati v novih okoliščinah). V jeziku humanistike bi to lahko poimenovali tudi *razum*, *etiko*, *zavest*, *botenje* ipd. Obstoje tega mehanizma povratnih informacij je osnova in osnovna zahteva za preživetje sistema. Kibernetični pristop dokazuje povezanost sistema v celoto in dokazuje, da sprememba v enem delu sistema povzroča spremembe, tako pričakovane kot tudi nepričakovane v drugih delih sistema.

Čeprav morda celo lahko razumemo povezavo med tehnologijo in družbo, pa kljub temu nismo sposobni v celoti vplivati na svojo prihodnost. Na dogodke večinoma le reagiramo, ne moremo pa jih vedno usmerjati. Na področju povezav med tehnologijo in družbo obstaja vrsta popolnoma diametralnih teorij. Jacques Ellul in Herbert Marcus menita, da postaja tehnologija vse bolj človekov gospodar in vse manj njegovo orodje. V nasprotju s tem pa drugi, kot npr. Lynn White, menijo, da je tehnologija nevtralna, skratka, da »*tehnologija odpira vrata, vendar se človek lahko odloča, ali bo vstopil ali ne!*« (Aberšek, 2012) Ta pogled takoj poraja naslednja vprašanja:

- Kdo odloča, katera vrata moramo odpreti?
- Ko in če vstopimo, ali tehnologija določa obliko sobe, v katero smo vstopili?
- Če je tehnologija enostavno skupek zamisli, se postavi še naslednje vprašanje: kdo določa cilje in ali ne nazadnje ne obstaja nevarnost, da zamisli same postanejo cilji?

Vsa ta vprašanja niso le zgodovinska vprašanja, ampak so tudi realni problemi današnjih dni, kar je razvidno tudi iz smeri in trendov razvoja in kontrole vseh tehnologij.

Mnogi problemi, ki so povezani s sodobnimi tehnologijami, so povezani tudi s strojno inteligenco (UI). Za reševanje teh problemov potrebujemo splošna znanja in odgovore o svetu, družbi, ljudeh, torej je to *globalen filozofski problem*. Zato je za vse, ki morajo sprejemati odločitve s področja UI, izjemno pomembno, da *razumejo* osnovne splošnega mehanizma inteligence – filozofijo duha, tj. kako deluje naravna, človeška inteligenca.

Na področju filozofije duha je vse od najzgodnejših začetkov pojava človeka kot mislečega »stroja« pa vse do danes, prisoten osnovni razkorak med razlagami, ki bi jih ekstremno lahko poenostavili na dve struji (Horst, 2007):

- *misterianistično*, ki zagovarja tezo, da je duh nekaj edinstvenega v svoji neponovljivosti, kjer je misterij v nekih lastnostih, svojstvenih samo študijam duha in duhu samemu, kot npr. t. i. »misterianizem«¹, ki ga je razvijal npr. Colin McGinn (1999);
- *naturalistično*, ki zagovarja tezo, da duha lahko ali ga celo moramo »naturalizirati«, in to tako, da so lahko mentalna stanja in procesi opisljivi z jezikom znanosti: fizike, nevroznanosti ali jezikom drugih naravoslovnih ved.

Oba pristopa skušata najti resnico, vendar uporabljata popolnoma različne metode in orodja ter interpretirata svoje dosežke na popolnoma različnih osnovah in izhodiščih. Če želimo poudariti to misel, lahko povzamemo misli A. Einsteina, ki je

¹ *Misterianizem* je filozofska pozicija, ki zagovarja tezo, da je *trd problem zavesti* (ang. »hard problem of consciousness«) onkraj meja naše razumljivosti.

trdil, da sta znanost (objektivnost) in duhovnost (subjektivnost) komplementarni druga drugi in zato potrebujemo obe. Zato filozofija uporablja obe, znanost in duhovnost, tj. objektivne in subjektivne izkušnje z namenom doseči uravnoteženost in harmonijo med racionalnim in intuitivnim razumom, med glavo in srcem. Resnična klasična filozofija se ubada z vprašanji, kot so: kako deluje veselje, zakaj veselje obstaja in kaj je smisel življenja. Razlika med znanstveniki, poimenujmo jih empirični znanstveniki (naravoslovci, tehniki ipd.), in filozofi ni v vsebini, temveč je v metodah in načinu razmišljanja. Za razliko od empiričnih znanosti, kjer je poudarek na opazovanju, zbiranju gradiva in klasifikaciji ter na izvajanju in interpretiranju eksperimentov ali razvijanju novih pristopov in sistemov, pristop filozofov temelji predvsem na *argumentiranju*, *pojmovni analizi* in *zgodovinskem vidiku* (Markič, 2010). Filozofi se sprašujejo o temeljnih predpostavkah (metafizičnih, epistemoloških in metodoloških), na katerih znanstveniki postavljajo hipoteze in načrtujejo eksperimente. Ob tem poskušajo podati sintezo različnih pristopov in sestavljajo pregled področja znanosti.

Raziskovalci s področja filozofije, kognitivne znanosti, nevrobiologije, nevroračunalništva in umetne inteligence se pogosto sprašujejo, ali in kakšna je zveza med razvojem kognitivnega nevromodeliranja (simuliranja delovanja možganov s pomočjo sodobnih računalniških tehnologij) in nevroračunalništva (računalniške simulacije nevronskih mrež oz. t. i. umetnih nevronskih mrež, ki v kognitivni znanosti prevzemajo vlogo procesnih modelov možganskih in duševnih procesov). Posplošeno, sprašujejo se o korelaciji človek – stroj in to glede prenosa idej, navdiha za modele ter o primernosti/uporabnosti teh modelov za najrazličnejše namene. S terminom »stroj« bom poimenoval vsak sistem, ki v sodobnem svetu skuša nadomestiti človeka, tako človeka kot fizično bitje (npr. stroj kot humanoidni robot, ki lahko na primer izvaja različne fizične aktivnosti), predvsem pa človeka kot duševno, mentalno, razmišljajoče bitje (npr. računalnik, ki simulira človeške kognitivne procese).

Do sedaj še ni bil dan natančen odgovor na tri, za raziskave na teh področjih bistvena vprašanja, in ta so:

- Ali lahko celovito kognitivno nevromodeliramo, tj. simuliramo delovanje možganov (glede korelacije med možgani in duhom) s pomočjo sodobnih računalniških tehnologij?

- Ali lahko razvoj nevroračunalništva in umetne inteligence, še posebej v obliki generativne umetne inteligence pripelje do nadomeščanja človeka in njegove naravne inteligence (*neposredno* – humanoidni robot, ki opravlja človeško delo ali *posredno* – računalniški »učitelj,« inteligentni tutor, ki opravlja le človekove mentalne funkcije)?
- Ideja, ki se ponovno bolj intenzivno pojavlja v zadnjih desetletjih, je, ali lahko človeka (duh in telo) obravnavamo kot dinamični sistem, skratka ali obstaja takšen sistem diferencialnih enačb, s katerimi bi lahko to zapisali.

Izhajajoč iz teh treh vprašanj bi lahko formulirali še četrto:

- Če je (bo) to vse možno, kakšna je korelacija med kognitivnim nevromodeliranjem in nevroračunalništvom?

Morda se komu zdita naslov in vsebina nekoliko odmaknjena od resničnih potreb današnje šole, kar kaže na očiten razkorak med znanostjo in aplikativnostjo oz. med teorijo in prakso (Kordigel Aberšek, Aberšek, 2021). Praksa bi nam upravičeno lahko zastavila vprašanje: *kje pa bi to lahko uporabili?* Poskusimo kar na začetku odgovoriti na to vprašanje in s tem rešiti dilemo, ki danes še ni tako očitna, zagotovo pa bo v prihodnosti eden od ključnih izzivov sodobne šole. Delo je sicer nastajalo v celotnem letu 2022/23 in je bilo dokončano že ob koncu leta 2023, vendar pa je potrebno izpostaviti, da smo ob koncu leta 2022 doživeli na tem področju pravo revolucijo, ki jo je potrebno vsaj omeniti. S ChatGPT se je zgodil pravi vdor umetne inteligence (UI) v šolsko področje in lahko upravičeno trdimo, da je vdor UI, imel podoben revolucionaren efekt kot pandemija leta 2020, ki nas je v naših razmišljanjih pri uvajanju digitalizacije premaknila kar za nekaj let v prihodnost. Umetna inteligenca nam zagotovo ne bo dala možnosti hitrega prilagajanja, kot smo ga imeli pri pandemiji. Način našega razmišljanja in prilagajanja se bo moral pospešiti, metaforično "s svetlobno" hitrostjo. In pri tem je pomembno, da vsaj poskušamo razumeti, kaj se dogaja. Prav temu pa je namenjena ta knjiga, ki izpostavlja tri ključne besede prihodnosti: *kognicijo*, *modeliranje* in *umetno inteligenco*.

Poskusimo osvetliti predvsem prva dva termina, *kognicijo* in *modeliranje* in ju poverzati v ključno sintagmo *kognitivno modeliranje*. Ko govorimo o *kogniciji*, ko govorimo o kognitivni znanosti, se poskusimo vrniti v nekakšno časovno izhodišče, k Platonu

in njegovi definiciji *ideje*. Zagovarjal je obstoj *sveta idej*, kjer je ideja (grško *Idea*, pogosto *Eidos*, redko *Ousia*) od človekovega spoznanja neodvisna bitnost. Ideje pri Platonu obstajajo neosebno in so neminljive. Po njegovem neminljiv človekov duh/kognicija teži k nadčutnemu svetu. Idejam lahko pripišemo 3 poglavitne značilnosti:

1. predmeti uma, mišljenja, za razliko od objektov čutnega zaznavanja,
2. oblike ali ideje dejansko bivajo (za razliko od objektov zaznavanja, ki so podvrženi obstajanju - genesis), zato jim pripišemo bitnost (ousia),
3. oblike ali ideje so večne, za razliko od sveta in stvari v njem, ki so minljive.

Skratka, človekova kognicija je nekaj, kar je človeku dano, je predmet uma, mišljenja in je zato bitna. Kognicija pa je tudi ozko povezana z delom Bloomove taksonomije (1956), to je z kognitivnim področjem.

Ko pa govorimo o *modeliranju*, izpostavimo *model*, kjer le-ta predstavlja pomanjšan ali poenostavljen posnetek realnosti, oz približek (abstrakcijo) realnosti. Pri teh poenostavitvah najpogosteje:

- izključimo nekoristno ali tisto, kar ima manjši vpliv,
- sistem podatkov predstavljamo lahko v formalni, matematični obliki ali le v obliki algoritma,
- upoštevamo bolj koristno in pomembno, manj koristno pa zanemarimo,
- uporabljamo lahko računalniške simulacije, npr. simulacije učnega procesa.

Kognitivno modeliranje srečujemo na različnih področjih, npr.:

- *Kognitivno modeliranje pri računalništvu* se ukvarja s simulacijo človeškega reševanja problemov in miselnega procesiranja v računalniškem/matematičnem modelu. Takšen model je mogoče uporabiti za simulacijo ali napovedovanje človeškega vedenja ali uspešnosti pri nalogah, podobnih modeliranim, in izboljšanje interakcije med človekom in računalnikom, kar je danes ključno pri najrazličnejših aplikacijah, npr. pri napovedovanju vremena.
- *Kognitivno modeliranje v psihologiji* opisuje, kako misli in zaznave ljudi vplivajo na njihovo počutje in vedenje. Kognitivni model je jedro kognitivne vedenjske

- terapije CBT (angl. Cognitive Behavior Therapy) in ima ključno vlogo pri pomoči terapevtov pri konceptualizaciji in zdravljenju težav njihovih strank.
- *Kognitivno modeliranje v izobraževanju* je tehnika prenosa znanja, pri kateri učitelj razkrije učencem svoj prikrit postopek procesiranja informacij z glasnim razmišljanjem med izvajanjem postopkov poučevanja, vključenih v nalogo za učenje.

Slednjemu, kognitivnemu modeliranju v šolstvu, posvetimo nekoliko več pozornosti, saj je ključno, za razumevanje v knjigi predstavljenih izhodišč in teorij trenutnih procesov v današnji šolski praksi. V obstoječi šoli so ogrodje za kognitivno modeliranje predmetnik in učni načrti na makro nivoju in učne priprave na mikro nivoju. V vseh dokumentih je temeljni proces povezan s formalizacijo procesa zapisa, kako učitelj razkriva učencem svoj način (prikrite) obdelave informacij, in to z razmišljanjem med izvajanjem poučevanja. To pomeni zapisati sosledje miselnih korakov na poti reševanja zastavljenega problema. Kot primer omenimo le splošno znani zapis Evklidovega algoritma (ki bo razložen v nadaljevanju na sliki 3.3), ki je v matematiki učinkovita metoda za izračun največjega skupnega delitelja (GCD) dveh celih števil, največjega števila, ki ju deli brez ostanka in jo lahko izrazimo v obliki diagrama poteka po sliki 3.3 ali pa v obliki formaliziranega računalniškega zapisa rekurzivne različice $\text{gcd}(r_{N-1}, 0) = r_{N-1}$, ki temelji na enakosti GCD zaporednih ostankov in pogoju ustavitve izračuna:

```
function gcd(a, b)
  if b = 0
    return a
  else
    return gcd(b, a mod b)
```

Naj na kratko povzamemo: kognitivno modeliranje v šoli je možno in potrebno tako v splošni vsakdanji praksi, pri ustvarjanju inteligentnih učnih okolij, kar že danes izvajamo pri vsakdanji praksi, pri pripravi učnih načrtov, kot tudi pri pripravi različnih učnih gradiv in učnih okolij. Na osnovi že znanih izsledkov kognitivne in nevro znanosti vemo, kako delujejo in se učijo naši možgani in kako poteka usvajanje znanj v učnem procesu. Ob upoštevanju teh znanj pa lahko na dokaj enostaven način oblikujemo sodobna učna okolja že danes, in s tem neposredno inoviramo učni proces in vanj vnašamo izsledke sodobnih raziskav.

Kognitivno modeliranje pa se lahko izvede tudi v bolj kompleksni obliki, npr. za šolo ali pouk kot dinamičen proces z namensko razvitimi diferenčnimi in/ali diferencialnimi (gibalnimi) enačbami, s katerimi bolj podrobno opisujemo človekove kognitivne procese in procese v šoli. To je vizija prihodnosti, ko bodo ta znanja bolj razširjena in splošno dostopna ter bodo tudi del izobraževalnega procesa učiteljev.

1.1 Povezovanje kognitivne znanosti in filozofije

Kognitivna znanost je interdisciplinarno področje raziskovanja duševnih procesov, v katero so kot osrednje discipline vključene filozofija, psihologija, računalništvo, nevroznanost, lingvistika in družbene vede. Kognitivna znanost skuša z razvijanjem interdisciplinarnega pristopa premostiti težave zaprtosti posamičnih strok, ki so ta in podobna področja raziskovale že prej, in priti do bolj celostnega vpogleda. Filozofija se v skupno področje kognitivne znanosti vključuje predvsem na dva načina:

- *Prvi* je vidik filozofije in teorije znanosti. Ker ne moremo govoriti o enotni znanosti, saj ima vsaka disciplina svoje značilnosti, so to filozofije posameznih znanstvenih disciplin (filozofija matematike, biologije itd.), ki se jim je v zadnjih desetletjih pridružila še filozofija kognitivne znanosti. Sem sodijo med drugim razprave o naravi in ravneh razlage, redukciji v znanosti, realizmu, antirealizmu in konstruktivizmu, metodologiji interdisciplinarnega raziskovanja, prvoosebne, drugoosebne in tretjeosebne raziskovanje in možnostih njihovega povezovanja ter vprašanja o temeljnih teoretskih postavkah, kot sta npr. računanje in reprezentacija. Filozofi se ukvarjajo tudi s teoretskimi osnovami kognitivne znanosti. Gre za poskus oblikovanja temeljnega teoretskega okvira in orodij, ki so potrebni za to, da bi razvili znanost, ki izhaja iz predpostavke o fizičnih osnovah duševnih procesov. Vključuje razprave, povezane z vprašanji o kognitivni arhitekturi: računski funkcionalizem, reprezentacijska teorija duha in jezik misli kot filozofski temelji klasične kognitivne znanosti in simbolnih modelov, konekcionizem, porazdeljene reprezentacije in vprašanje poljudne filozofije, utelešena kognicija in teorija dinamičnih sistemov.
- *Drugi* vidik predstavlja kognitivna filozofija, ki obsega tista področja, kjer se raziskovanja v kognitivni znanosti prekrivajo s filozofijo ter neposredno

prispevajo k razumevanju pojavov in reševanju vprašanj, ki so bila prej izključno v njeni domeni. Gre predvsem za vprašanja filozofije duha, pa tudi epistemologije, filozofije jezika, filozofije znanosti, estetike in etike. (Churchland, 1986, 2007)

1.2 Zakaj umetna inteligenca potrebuje filozofijo?

Ideja o inteligentnih strojih je stara, sega nekako v leto 1940, ko je Isaac Asimov pričel objavljati kratke zgodbe o robotih v ameriških revijah *Super Science Stories* in *Astounding Science Fiction*, ki jih je potem leta 1950 izdal v knjigi *I Robot*. To delo je ena od prvih študij interakcije med človekom in robotom ter moralo in etiko. (Aberšek, 2013)

Resnejše znanstveno delo na področju umetne inteligence, ki je takrat še niso imenovali s tem terminom, pa se je pričelo z obdobjem pojava računalništva, natančneje s pojavom programabilnih računalnikov (Zuse, 1993). Menimo lahko, da se je filozofija umetne inteligence pričela s Turingovim člankom *Computing Machinery and Intelligence* (Turing, 1950) in s Shannovo razpravo (Shannon, 1950), kako bi moral biti programiran stroj, da bi lahko igral šah. Od teh prvih začetkov je razvoj umetne inteligence skokovito napredoval vse do današnjih dni in do vprašanja: »Ali stroj lahko misli?« To je v povezavi z osnovnimi definicijami umetne inteligence, ki jih je mnogo in so bolj ali manj natančne; npr. po Copelandu: »Umetna inteligenca je *znanost o izdelavi strojev*, ki so sposobni narediti stvari, za katere je po naših merilih potreben um.« (Marvin Minsky, 1968, ustanovitelj laboratorija za UI na MIT-ju, (v Copeland, 1993/2007: 1). Na generalnem nivoju pa lahko ugotavljamo, da je umetna inteligenca sestavljena dualistično, da se v njej prepletata dve smeri, dva dela:

- *epistemološki del*;
- *hevristični del*.

Epistemološki del je filozofska predstava sveta in njegovih problemov, katerih prepoznavanje in reševanje izhajata iz njihovih teoretskih osnov (npr. vedeti moramo, kaj moramo gledati, da to vidimo), medtem ko je *hevristični del* mehanizem sklepanja, ki na osnovi informacij in najrazličnejših matematičnih modelov rešuje problem in odloča o potencialnih rešitvah.

Pri svojem dosedanjem raziskovalnem delu sem se pretežno ukvarjal s hevrističnim pristopom reševanja problema. Moje osnovno izhodišče, lahko ga poimenujemom tudi inženirski pristop k reševanju problema, je povezano predvsem s problemom in njegovo rešitvijo, ki lahko bolj ali manj zadovoljivo reši ali odpravi zadani problem. Govorimo o optimiranju, o iskanju optimalnih rešitev (ne le ene, katerekoli, temveč »prave« rešitve). Pri tem sta bili vedno v ospredju vprašanji: *Kako izdelati nek računalniški program – ekspertni sistem in katere metode umetne inteligence pri tem uporabljati za doseg zastavljenega cilja*. Izid tega dela je bil, med drugim, na znanju temelječ ekspertni sistem (angl. Knowledge based expert system) STATFAG, namenjen optimiranju zobniških gonil. V tej programski opremi, ekspertnem sistemu (ES), sem združil ekspertna znanja – podatkovne baze, inženirska orodja (matematične modele za preračunavanje zobnikov predvsem na podlagi sodobnih metod, kot je npr. metoda končnih elementov); avtomatsko generiranje potrebnih geometrijskih podatkov, kot je npr. oblika zob in zobnikov; mreže končnih elementov ipd., orodja za zajemanje in predstavljanje podatkov in hevristiko ter pravila in metode umetne inteligence, katere primarni predstavnik je bil genetski algoritem. Ekspertni sistem je omogočal avtomatizirano in predvsem izredno hitro izračunavanje potrebnih parametrov za pripravo optimalnih rešitev zobniških gonil. Programska oprema je bila v prvem koraku namenjena predvsem olajšanju inženirskega dela (veliko izračunov v izredno kratkem času, kar je omogočalo suvereno izbiranje najboljše rešitve), v kasnejši fazi pa sem jo dogradil za uporabo tudi pri študiju. Tako je ekspertni sistem omogočal študentom in mlajšim, manj izkušenim strokovnjakom učenje in študij vplivov posameznih parametrov na končni videz zobnikov in zobniških gonil. (Aberšek, Flašker, 2004)

Osnovni namen tega dela je dodati dosedanjemu hevrističnemu delu tudi njegov filozofski del razumevanja naravne in umetne inteligence, predvsem na področju filozofije duha. Hipoteza, ki jo bom v tem delu zagovarjal in dokazoval, je, da je duh *lahko*, ali še zahtevneje, *mora biti* »naturaliziran,« tako da se (psihofizična redukcija) mentalna stanja in procesi reducirajo v nekaj, kar bi se dalo izraziti z jezikom znanosti – fizike, nevroznanosti ali s kakšnim drugim naravoslovnim jezikom (npr. najpomembnejše lastnosti razuma, kot so zavest, hotenje, normativnost, reducirati v to, kar počnejo možgani). To bom poskušal osvetliti iz različnih zornih kotov.

Če se nekoliko poglobimo v zastavljeno problematiko in izhajamo iz premise, da je filozofijo duha možno opazovati naturalistično in da je možna redukcija, potem se pri tem takoj postavi vprašanje: *Kaj naj bo naravoslovno – duh kot tak, zavest²...?* In pri tem, ali splošno shemo lahko razumemo kot *pozitivistično zahtevo* (duh je lahko prilagodljiv) ali kot *normativno zahtevo* (duh mora biti prilagodljiv). Na osnovi tega bi lahko definirali tri mejne primere, kako so naravoslovno usmerjeni znanstveniki v preteklosti razumeli svet narave in problem duha v njem (Horst, 2007):

- *Redukcionistično*: Sem prištevamo naturaliste duha, katerih glavni zagovornik je bil Galileo s svojo znanstveno metodologijo analize in sinteze (angl. MRC – method of resolution and composition).
- *Normativno*: S predpisi urediti razmerje med duhom in telesom, torej najti zakone, ki bi omogočili napovedovanje in kontrolo. Glavni predstavnik te strujev naravoslovju je bil Newton, v filozofiji pa Locke.
- *Duh po bioloških izražih*: Sem sodi sociobiologija – razvojna teorija, ki jo je ustvaril Darwin.

Pri sodobnih filozofih duha se izpostavljata pri tem dve temeljni filozofski vprašanji:

- Ali je mogoče mentalne fenomene, kot so občutki, intencionalna stanja (prepričanja, želje, namere, mišljenja) in čustva, zreducirati na razlago s termini in jezikom naravoslovnih znanosti?
- Ali mentalni fenomeni ustrezajo stanjem in procesom, ki jih opisujejo naravoslovne znanosti – fizika, kemija, biologija, nevroznanost?

V novejši zgodovini (druga polovica dvajsetega stoletja), ko so ta vprašanja prišla v ospredje, je bilo kar nekaj razlag in odgovorov, zagotovo pa vse do danes ni prišlo na tem področju do nikakršnega konsenza. Naj na kratko povzamemo le nekatere (Kos, 1973):

² *Zavest* je termin, ki označuje raznolikost vidikov povezav med duhom in svetom, s katerim je v povezavi (van Gulick, 2004). Definirana je kot osebna: je zavedanje, sposobost doživeti izkušnjo ali čutiti; budnost, imeti smisel za individualnost ali ekskluzivni nadzorni sistem nad duhom. (Farthing, 1992, Bregant, 2010).

- *Psibofizični redukcionizem (reduktivni fizikalizem)*, ki je dosegel svoj razmah sredi šestdesetih in zaton v začetku sedemdesetih let, trdi, da lahko mentalne fenomene prikazujemo s fizikalnimi dejstvi in jih lahko zreduciramo na fizikalne zakonitosti. Poenostavljena metafizična/ontološka redukcija: psihološki zakoni → individualni nevron v možganih → molekula → fizika atomov. Če to velja, potem so vsi psihološki zakoni v principu zreducirani na zakone fizike osnovnih delcev. Pri tem se postavi naslednje vprašanje: Kako daleč še lahko poenostavljamo (reduciramo) nevroznanost, da ne izgubimo preveč oz. se preveč ne oddaljimo od bistva problema? Tukaj prihaja do razkoraka med redukcionisti samimi.
- *Psibofizični antiredukcionizem (nereduktivni fizikalizem)* zagovarja tezo, da so mentalni pojavi odvisni zgolj od fizičnih, niso pa z njimi identični, zaradi česar prvih ne moremo reducirati na druge. Tako mentalni pojavi ohranijo pristen ontološki status, psihologija pa ostane neodvisna znanost s svojo posebno metodologijo raziskovanja.

Na metafizično vprašanje o tem, kakšen je odnos med fizičnim in psihičnim (problem duha in telesa), so filozofi že davno pred rojstvom kognitivne znanosti iskali odgovore v smeri dualizma; duh in telo sta dve različni stvari (npr. Platon, Descartes); monizma – bodisi materializma/fizikalizma, po katerem je vse, kar obstaja, fizično (zagovarjajo atomisti, npr. Lukrecij, La Mettrie) bodisi idealizma, po katerem je vse, kar obstaja, duševno (Berkeley) ali nevtralnega monizma/psihofizičnega paralelizma, po katerem je to, kar obstaja, nekaj nevtralnega, duševno in fizično pa sta dva vidika le-tega (npr. Spinoza, Russell, Dammasio, 2008). Slednje se imenuje tudi paralelizem. Znotraj vsake teh skupin najdemo nato številne nadaljnje različice, ki še mnogo bolj natančno opredeljujejo posamezna stališča. Tu se bomo omejili na funkcionalizem in eliminativizem, ki sta najbolj zaznamovala diskusije med filozofi kognitivne znanosti. Videli bomo, da je sodelovanje med filozofi in empiričnimi znanstveniki kot nekakšno vzpenjanje po spirali, kjer enkrat filozofska razmišljanja vplivajo na empirična raziskovanja, drugič pa empirična raziskovanja vplivajo na filozofska razmišljanja. (Markič, 2008)

Kot smo že omenili, je v začetnem obdobju kognitivne znanosti zavzemala osrednje mesto analogija med digitalnim računalnikom in umom. Tako um kot digitalni računalnik sta simbolna informacijska sistema. Filozofi so z računsko različico

funkcionalizma, pri kateri gre za kombinacijo reprezentacijske teorije (Fodor 1975, 1987) in računalniške analogije, teoretsko podprli klasično kognitivno modeliranje. Funkcionalistična teorija je filozofski poskus razločevanja kritičnega dela, tj. načina, kako prepoznavamo in razvrščamo duševna stanja, npr. kaj je tisto skupno, zaradi česar so različne bolečine bolečine. Funkcionalizem pravi, da so tisto, kar je skupno vsem bolečinam in kar jih naredi bolečine, njihove funkcionalne vloge. Pri tem so mišljene vzročne povezave od subjektivih zaznavnih dražljajev do ostalih duševnih stanj in odgovorov, ki se kažejo v vedenju. Z drugimi besedami lahko rečemo, da duševna stanja določamo glede na njihovo vzročno vlogo.

Reprezentacijska teorija temelji na pristopu, po katerem s propozicionalnimi naravnostmi predstavljamo duševna stanja. Izraz propozicionalne naravnosti, ki ga je uvedel Bertrand Russell, se nanaša na povedi, kjer v glavnem stavku uporabljamo glagole, kot so *verjeti*, *želeli si*, *upati si*, *bati se*, ki jim sledi odvisni stavek, uveden z veznikom *da*. Npr.: *Želim si, da bi šli jutri na izlet*. Propozicionalne naravnosti so postale standardna oblika za opredelitev duševnih stanj. Omogočajo nam, da imamo lahko isto naravnost do različnih propozicij ali pa različne naravnosti do iste propozicije. V računski reprezentacijski teoriji, kot jo je razvil Fodor (1975, 1987), so propozicionalne naravnosti funkcionalne (računske) relacije do duševnih reprezentacij. To, da so v posameznem propozicijsko naravnem stanju, pomeni, da imamo reprezentacijo vsebine propozicije prikazano v funkcionalni vlogi, ustrezni tej naravnosti. Npr. misliti: »*Deževalo bo. Torej grem noter,*« (Fodor, 1987: 17) je prikazati duševno reprezentacijo, ki pomeni: »*Grem noter,*« povzročeno na določen način s prikazovanjem duševne reprezentacije, ki pomeni 'deževalo bo.' Druga trditev reprezentacijske teorije duha govori o naravi duševnih procesov: Duševni procesi so vzročna zaporedja prikazovanj duševnih reprezentacij. (Markič, 2008)

- *Funkcionalizem* (1960–1980). Funkcionalisti poudarjajo, da karakteristične lastnosti niso strukturne, temveč funkcionalne (srce je krvna črpalka, torej je naravno srce enako umetnemu srcu).

Funkcionalizem je kot izhodišče kognitivne znanosti za svojo temeljno podlago vzel poljudno (zdravorazumsko) psihologijo. To je preprosta teorija, ki so jo za pojasnjevanje in razumevanje vedenja ljudi uporabljali že starogrški filozofi in jo v vsakdanjem življenju uporabljamo še danes. Temelji na spoznanju, da želje in

prepričanja povzročajo obnašanja. V običajni zdravorazumski psihologiji imajo duševna stanja tako vzročno vlogo kot tudi semantično vsebino in sosledje misli v veliki meri ohranja resničnostno vrednost. Fodor je z reprezentacijsko teorijo duha in s pomočjo računalniške metafore pokazal mehanizem, ki zmore posredovati med vzročnimi in semantičnimi lastnostmi. V računalniku so vzročne lastnosti simbola povezane s semantičnimi lastnostmi preko sintakse. Sintaksa je ena od fizikalnih lastnosti in si jo lahko v približku zamišljamo kot abstrakten atribut njene oblike. Sintaksa je tako reducirana na obliko, in ker oblika potencialno določa vzročno vlogo, si lahko zamislimo primerke simbola, ki bodo delovali vzročno zaradi svoje sintaktične strukture. »Sintaksa simbola lahko določa vzroke in posledice njegovih uprimerjanj na podoben način, kot lahko oblika ključa določa, katero ključavnico bo odprl«. (Fodor, 1987: 19) Po drugi strani pa je računalnik interpretiran avtomatičen formalni sistem – stroj za manipuliranje s simboli. Če stroj sledi v sistemu definiranim sintaktičnim pravilom pri tvorjenju novih formul, potem bodo interpretirane formule ohranjale svojo semantično vrednost. Po analogiji med računalnikom in duhom so tudi duševne reprezentacije simboli, ki imajo vsebino in so hkrati vzročno odgovorni za vedenje.

Klasična kognitivna znanost temelji na dveh predpostavkah. Prva je prepričanje, da vsako inteligentno obnašanje predpostavlja zmožnost ustrezne predstavitve sveta in da razlaga obnašanja brez te predpostavke ni mogoča. Zato je eden od osrednjih pojmov klasične kognitivne znanosti pojem reprezentacije. Druga predpostavka odraža temeljno značilnost funkcionalističnega pristopa, da določanje duševnih stanj in procesov poteka neodvisno od nevrofizioloških stanj in procesov (materialnega udejanjanja). Tipe duševnih stanj razvrščamo na osnovi njihovih vzročnih (funkcionalnih) vlog in ne na osnovi njihovega fizičnega uprimerjanja. Vrnimo se k primeru bolečine. Da bi za neko duševno stanje lahko rekli, da je bolečina, je dovolj, da določimo njegovo vzročno vlogo, ki je izogibanje ali odstranitev izvora poškodbe organizma. Pri tem bolečino pojmujeemo kot bolečino zato, ker ima tako funkcionalno vlogo, ne pa zaradi določenega nevrološkega stanja. Za funkcionaliste so posamezna duševna stanja, ki imajo isto vzročno vlogo, le primerki določenega tipa duševnega stanja. Kognitivni znanstveniki govorijo o treh ravneh opisa duha oz. možganov: semantični, sintaktični ter o ravni strukturalne arhitekture (fizični ravni). Za razlago kognitivnih procesov sta pomembni semantična in sintaktična raven, medtem ko raven strukturalne arhitekture možganov, ki je v domeni nevroznanosti, podaja predvsem omejitvene pogoje. Kajti čeprav je sintaktična raven fizično

udejanjena, je ne moremo reducirati (v običajnem pomenu) na fizično raven. Namesto o redukciji kognitivisti govorijo o udejanjanju ali implementaciji podobno, kot je računalniški program implementiran v strojni opremi. Iskanje ustreznih psiholoških razlag je podobno iskanju ustreznih programov, ki jih izvajajo možgani. (Markič, 2010)

Prevladujoče stališče funkcionalizma glede problema telesa in duha je šibkejša različica teorije identitete – identiteta primerkov (Davidson, 2004) oziroma nereduktivni fizikalizem. (Horgan, 1997) Po tej teoriji je vsak primerek duševnega stanja primerek nevralnega stanja, vendar teorija ne zahteva, da bi tipe oz. vrste duševnih stanj izenačili s tipi oz. vrstami nevralnih stanj. Tako je npr. vedno, kadar sem v določenem duševnem stanju (npr. v tem trenutku čutim bolečino v roki), to duševno stanje identično z določenim možganskim stanjem, obenem pa sem ob drugi priložnosti lahko v istem duševnem stanju, a v drugačnem možganskem stanju. Ta teorija ima precejšnjo podporo v psihološki praksi. Psihologi pripisujejo isto duševno stanje (npr. bolečino) človeku v različnih obdobjih, čeprav vemo, da se možgani skozi čas spreminjajo, ali pa različnim ljudem, čeprav obstajajo razlike med možgani različnih ljudi. Običajno se pripisujejo ista duševna stanja tudi članom različnih vrst, pri katerih se možgani še bolj razlikujejo med seboj. V filozofskem jeziku rečemo, da duševna stanja supervenirajo nad fizičnimi stanji oziroma da so duševna stanja določena s fizičnimi stanji. Prav argument iz večvrstne realizacije je bil glavni argument proti močnejši identiteti vrst teorij oziroma reduktivnemu fizikalizmu. (Place, 1956) Po tej teoriji so duševna stanja identična s stanji možganov: vsi primeri določenega tipa duševnih stanj (npr. bolečina) so identični s primeri koreliranih tipov nevronske stanj (npr. vzbujenje C-vlaken). V zadnjem času ta teorija spet pridobiva podporo, predvsem zaradi novih spoznanj v nevroznanosti. (Bickle, 1998)

Analogija med računalnikom in duhom je po eni strani omogočila postavitev hipoteze, da so duševne reprezentacije fizično udejanjene oz. realizirane v obliki simbolne kode v možganih, in s tem pomagala odgovoriti na vprašanje, kako da so duševna stanja, predvsem propozicionalne naravnosti (prepričanje, želja, namera itd.), ne samo fizično mogoča, ampak tudi dejansko povzročajo vedenje. Po drugi strani pa je analogija pripeljala do hipoteze o računalniku kot mehanskem modelu misli in omogočila empirično preverjanje psihološke teorije. Temeljno merilo za

sprejem ali zavrnitev hipotez je možnost njene računalniške uresničitve, izdelave modela.

Klasična kognitivna znanost je zdravorazumsko psihologijo jemala kot izhodišče za nadaljnje raziskovanje kognitivnih pojavov. Med pomembne dosežke šteje prav to, da je pokazala na možnost, kako so lahko prepričanja in želje kot vzročno učinkujoča stanja uprimerjana v fizičnem mehanizmu. S tem je »izzvala« Renčja Descartesa in njegov argument za dualizem iz *Razprave o metodi* (Descartes, 1637/1957: 81–82), kjer je dokazoval, da je skoraj nemogoče, da bi stroj lahko deloval razumno. Če bi se raziskovanje in modeliranje razvijalo uspešno, bi uspešni modeli, razlage in napovedi podkrepili začetno izhodišče in teoretski pristop funkcionalizma.

Tako filozofska kot empirična raziskovanja so pokazala, da imata funkcionalistična teorija in pristop klasične kognitivne znanosti tudi precejšnje pomanjkljivosti. Filozofi so npr. izpostavili problem duševne vzročnosti – ali karakterizacije duševnih stanj kot funkcionalnih stanj, ki so udejanjena v možganih – ki zagotavlja, da bodo duševna stanja imela vzročno moč in ne bodo zgolj epifenomeni. (Kim, 1998, Bregant, 2004) S tem se je odprlo tudi vprašanje kvalitativnih doživljajev. Prav tako predstavlja grožnjo za funkcionalistični pogled na svet obstoj subjektivnih kvalitativnih izkušenj ali doživljajev – takšnosti oz. qualia, ki naj se jih ne bi dalo opisati v fizikalističnih pojmi. (Jackson, 1982) Problem je bila tudi ostra delitev na funkcionalno in strukturalno raven ter pristop od zgoraj navzdol, kar je omogočalo razlago prehodov med udejanjenimi stanji, ni pa dajalo vzročne razlage za nastanek določene kognitivne funkcije in ni dovolj upoštevalo nevrofizioloških omejitev. Pokazale so se tudi težave, ki se tičejo konkretnih neuspehov pri modeliranju, npr. prepoznavanje vidnih ali slušnih vzorcev, kategorizacija, gibanje v prostoru, razmišljanje s pomočjo ozadnjega (podzvestnega) znanja (angl. background knowledge) in na splošno vse tiste naloge, za katere ne poznamo eksplicitnih pravil.

– *Eliminativizem*. Gre za teorijo duha, ki zavrača obstoj mentalnega.

Nekateri filozofi duha v nasprotju s funkcionalisti menijo, da je napačno že samo izhodišče klasične kognitivne znanosti, saj je zdravorazumska psihologija primer napačne teorije in zato povsem neustrezna kot osnova znanstvene psihologije. Tako Paul Churchland (2007) dokazuje, da je poljudna psihologija napačno in zavajajoče pojmovanje vzrokov obnašanja in značilnosti kognitivne dejavnosti. Po njegovem

mnenju ne gre le za nepopolno interpretacijo obnašanja, kot bi se strinjali tudi funkcionalisti, temveč celo za napačno prezentacijo naših notranjih predstav. Zato naj ne bi bilo mogoče pričakovati ustrezne nevroznanstvene teorije, ki bi se ujemala s kategorijami našega zdravorazumskega okvira. Zaradi nemožnosti redukcije naj bi bil stari okvir zdravorazumske psihologije (prepričanja, želje, namere, občutki itd.) preprosto odstranjen in nadomeščen z razvito nevroznanstveno teorijo. Razlage vedënja bi se sklicevale na nevrofarmakološka stanja in nevralno delovanje v specializiranih področjih možganov oziroma na tisto, kar bo relevantno v novi teoriji, kar pa hkrati ne bo ustrezalo zdravorazumskim entitetam (Churchland, 2007). Kot primer takega odnosa v zgodovini znanosti eliminativistični materialisti največkrat navajajo flogiston, eter in čarovnice, ki nimajo mesta v sodobnih znanstvenih teorijah. Oglejmo si primer *flogistona*. Vse do Antoina-Laurenta de Lavoisiera (1743–1794) so bili znanstveniki prepričani, da se ob gorenju osvobaja duhu podobna substanca, ki so jo imenovali *flogiston*. Lavoisierova odkritja o sestavi zraka in vlogi kisika pri dihanju ter gorenju so prispevala k nastanku moderne kemije. Ker se je izkazalo, da pri gorenju sodeluje kisik iz atmosfere, je flogiston postal ne le nepopoln, ampak radikalno neustrezen opis dogajanja. V novi teoriji flogiston ni imel več nobene vloge in ker ni bil primeren za redukcijo ali identifikacijo s pojmi v novi kemiji, je bil odstranjen iz znanosti.

Eliminativist mora torej najprej pokazati, da je teorija, v kateri nastopajo določene entitete, neustrezna in jo je treba zamenjati z novo, boljšo teorijo. Običajno velja, da je nova teorija boljša, če z njo dobimo natančnejše napovedi in boljšo razlago na večjem področju. Pri tem se mora vsaj tako dobro kot stara teorija ujemati tudi s teorijami s sosednjih področij. Vendar to še ni dovolj. Eliminativist mora nadalje pokazati, da nova teorija ne ohranja starih teoretskih postavk. V novi teoriji se namreč lahko stare predpostavke tudi ohranijo, in to na dva načina: tako da se postavke stare teorije reducirajo na postavke nove teorije ali pa tako, da se teoretske postavke stare teorije bolj natančno opredelijo. Primer za drugo možnost je Kopernikova heliocentrična teorija, ki je sicer povsem drugače opredelila planete in njihovo gibanje kot stara Ptolemajeva teorija, a je kljub temu še vedno govorila o planetih.

Teoretske spremembe, pri katerih se entitete in procesi stare teorije ohranijo oziroma reducirajo na nove, se imenujejo ontološko konservativne teoretske spremembe. Teoretske spremembe, ki niso ontološko konservativne, pa so ontološko radikalne

in imajo za posledico odstranitev postavk stare teorije. Ramsey, Stich in Garon (1991) so dokazovali, da konekcionistični pristop h kognitivnemu modeliranju ni v skladu z zdravorazumskim okvirom in da so konekcionistične reprezentacije nezdružljive s prepričanju, željami in drugimi postavkami zdravorazumske psihologije. Za poljudno psihologijo je po njihovem ključno načelo propozicionalne modularnosti, ki je sestavljeno iz trditev, »da so propozicionalne naravnosti funkcionalno diskretna, semantično interpretabilna stanja, ki igrajo vzročno vlogo v ustvarjanju drugih propozicionalnih naravnosti in nazadnje v ustvarjanju obnašanja«. (Ramsey et al., 1991: 204) To, da so propozicionalne naravnosti funkcionalno diskretna stanja, pomeni, da je smiselna trditev, da je oseba pridobila ali izgubila posamezno prepričanje ali spomin (npr. Maja se je zjutraj zbudila in popolnoma pozabila, kam je spravila ključ, čeprav ni pozabila ničesar drugega). Propozicionalne naravnosti so semantično interpretabilne, če dopuščajo posplošitve zdravorazumske psihologije, ki se naslanjajo na semantične lastnosti prepričanj. Pri tem so predikati, ki izražajo te semantične lastnosti (npr. verjetje, da bo imel vlak zamudo), taki, da jih lahko projiciramo v prihodnost in jih uporabljamo v nomoloških posplošitvah (npr. vse ljudi, ki so prepričani, da bo imel vlak zamudo, uvrstimo v eno skupino in na podlagi tega pričakujemo, da bodo v njihovem vedenju neke zakonitosti). In končno, ta funkcionalno diskretna, semantično interpretabilna stanja igrajo vlogo v ustvarjanju obnašanja, saj posamezno prepričanje lahko navedemo kot vzrok določenega dejanja.

Argument Ramseyja, Sticha in Garona je bil pogojen; namreč, če se konekcionistični modeli in nevronske mreže izkažejo za dobre kognitivne modele, potem je to močan razlog za zavrnitev poljudne, zdravorazumske psihologije. Na tem mestu se ne moremo natančneje posvetiti temu argumentu, rečemo lahko le, da je tema spodbudila tako filozofe kot znanstvenike. Odprla se je široka diskusija, kako sploh razumeti poljudno psihologijo in jo povezati z dosežki znanosti. (Greenwood, 1991) Nekateri so prepričani, da poljudne psihologije ne moremo zavreči, in zato sklepajo o neustreznosti konekcionizma in nevronske mreže. Drugi opozarjajo, da so bili obravnavani modeli preveč enostavni in da so kompleksni konekcionistični modeli lahko združljivi s poljudno psihologijo ter da so pri nekaterih vprašanjih, npr. problem okvira, celo bolj obetavni.

Celotna problematika bo razdeljena na dva dela, torej na epistemološki in hevristični del.

1.3 Kognitivno modeliranje in generativizem

V 21. stoletju lahko z zagotovostjo postavimo dve izhodiščni eksistenčni premisi (Aberšek, 2013):

- *svet ni eden, svetova sta dva*³ in
- *inteligenca ni ena, inteligenci sta dve (bitje ni eno, bitji sta dve (človek + UI)).*

Kaj to pomeni? Razložimo na kratko. Danes ne obstaja samo *fizični, naravni svet* in iz njega izpeljan socialni svet, svet ki si ga je ustvaril človek skozi zgodovino, ko je svojim željam in potrebam prilagajal *naravni svet*. Obstaja tudi *kibernetski svet*, svet, ki ga je pričel kreirati pretežno s pojavom računalništva in informacijske družbe (v okviru tretje industrijske revolucije) človek sam, ko se je postavil v vlogo *kreatorja*. V zadnjem desetletju postajajo meje med tema svetovoma vse bolj zabrisane. Prav tako tudi ne obstaja več *samo ena inteligenca*. Po letu 2022 se je celotno človeštvo pričelo zavedati, da obstaja poleg *naravne inteligence*, ki je nastajala skozi milijone let naravnega razvoja, tudi umetna inteligenca, *generativna umetna inteligenca* (GUI), ki jo je človek razvil v nekaj zadnjih desetletjih. Dejstvo pa je, da obe inteligenci danes sobivata na istem planetu, sobivata v istem sistemu, v istem prostor-času.

1.3.1 Razvoj sodobnega izobraževanja skozi čas

Z današnjim nivojem vedenja, bi lahko sodoben razvoj človeštva in njegovega vpliva na svet (naše naravno okolje) v okvirno zadnjih dveh stoletjih poimenovali s terminom »ustvarjalno uničenje«. S tem opisujemo dinamiko tehnoloških diskontinuitet, ki so poganjale tako imenovane valove tehnoloških inovacij (industrijske revolucije) v naši sodobni dobi (McKraw, 2007). Za vsak val v tej seriji je značilen nabor tehnologij, ki so bile vodilne tehnologije svoje dobe. *Prvi val* (1785 do 1845), ki je sprožil industrijsko revolucijo, je bil zaznamovan z uporabo vodne energije za proizvodnjo tekstila in železa v gradbeništvu. *Drugi val* (1845 do 1900) je bil zaznamovan s parno energijo, jeklom in širitvijo železniških sistemov, nato *tretji val* (1900–1950) z električno energijo, kemikalijami in motorjem z notranjim zgorevanjem. Ko je vsak val tehnoloških sprememb prišel, je dinamika ustvarjalnega

³ Obstaja naravni svet, svet 1, v katerem bivamo ljudje in roboti, vidne oblike umetnega življenja, in kibernetski svet, svet 2, svet znotraj računalniških sistemov, v katerem bivajo samo človeku nevidne/breztelesne oblike umetnega življenja, kot je npr. ChatGPT.

uničevanja prinesla s seboj nove načine dela in življenja, vendar je tudi uničila nekaj tega, kar je bilo prej, in predvsem drastično vplivala na človekovo naravno okolje. Prvi trije valovi inovacij so preoblikovali zahodne družbe, izpodrinili *fizične delavce* in jih nadomestili s strojno avtomatizacijo in *delavci znanja*. Kmetijska gospodarstva so postala industrijska, nato postindustrijska. V njih je bilo zaposlenih vedno več delavcev znanja. Generacija za generacijo študentov, učiteljev, intelektualcev, strokovnjakov, pisateljev, raziskovalcev, ustvarjalcev – ti so bili gonilna sila človeške ustvarjalne ekonomije znanja.

Četrti val (1950–1990) je bil obdobje petrokemije, elektronike in letalstva. Označuje tudi začetek umetne inteligence kot področja raziskav. Izraz "umetna inteligenca" je bil skovan leta 1956 na srečanju poletnega raziskovalnega projekta Dartmouth, ki sta ga gostila Marvin Minsky in John McCarthy. *Peti val* (1990–2020) je bila digitalna doba, ki so jo opredelila omrežja, ki so jih omogočili izum interneta (1983) in World Wide Web (1989) Tima Berners-Leeja, ki je uporabnikom omogočil krmarjenje po njem. Prvi splet - zdaj imenovan Web 1.0 - je bil statični prostor, sestavljen iz strani HTML (hipertext označevalni jezik), imenovan "*branje in pisanje*" spleta, ker je bilo to vse, kar so uporabniki lahko storili z njim. Toda izboljšan Web 2.0, ki je sledil med letoma 1999 in 2004, je zaznamoval pomembno spremembo v funkcionalnosti in uporabniški izkušnji, saj ga je uvedba HTTP (protokol za prenos hiperteksta) preoblikovala iz statične v socialno, uporabniku prijazno platformo (Berners-Lee et al., 2001). V primerjavi s prejšnjim omrežjem statičnih spletnih strani z informacijami, ki tečejo v eno smer, lahko uporabniki zdaj sodelujejo in delijo informacije, komentirajo, pošiljajo mnenja, objavljajo bloge in komunicirajo. S tem se je začelo obdobje socialnega spleta ali spleta 2.0.

V 1990-ih je prišlo tudi do razvoja tehnik strojnega učenja in izboljšanih arhitektur nevronske mreže, ki so uporabljale ogromne količine podatkov (tako imenovane "velike podatke"), ki so jih ustvarila družbena omrežja Web 2.0, vključno s Facebookom (2004), YouTubom (2005), Twitterjem (2006) in Instagramom (2010). Web 3.0, tako imenovani semantični splet, ki ga je leta 1989 zasnoval tudi Tim Berners-Lee kot mrežo podatkov, ki jih lahko berejo stroji –, je ustvaril protokole in tehnologije, ki so strojem omogočile branje metapodatkov, kar jim omogoča, da razumejo naše uporabniške preference in pretekle osebne spletne izkušnje. Ta splet je uporabljal tudi generativno umetno inteligenco za ustvarjanje vsebin, prilagojenih

individualnim željam in vedenju, ter za razvoj bolj izpopolnjenih iskalnikov in sistemov priporočil (Aberšek, Flogie, Pesek, 2023).

Šesti val se je pričel decembra 2015, ko je skupina, ki je vključevala Sama Altmana, Grega Brockmana, Reida Hoffmana, Elona Muska in druge, ustanovila *OpenUI*. Temeljni cilj te skupine je bil ustvariti algoritme varne *splošne umetne inteligence* (angl. *Artificial General Intelligence, AGI*), ki naj bodo v korist človeštva. Koncepti in terminologije AGI (tudi močna UI) se precej razlikujejo, vendar je ključna beseda v teh razpravah "*splošna*". Medtem ko imamo veliko specializiranih umetnih inteligenc na določenih področjih, ki so sposobne opravljati naloge enako dobro ali bolje od ljudi, še nimamo umetne inteligence s splošno inteligenco, ki bi bila tako kompleksna, da bi lahko dosegala ali celo presegala inteligenco ljudi v širokem spektru kognitivnih nalog. Časovnica za doseganje AGI je predmet stalne in živahne razprave več pomembnih osebnosti v skupnosti UI, vključno z Elonom Musk, Shaneom Legg, Dariom Amodeijem in Samom Altmanom, in se vse bolj oddaljuje od v izhodišču zastavljene, leta 2020 (Henshall, 2024). Raziskava večje skupine 2.778 raziskovalcev predvideva, da bodo do leta 2028 dosegli več ključnih mejnikov, da bodo stroji brez pomoči presegli inteligenco ljudi v vseh možnih nalogah in to za 10% do leta 2027 in za 50% do leta 2047. Izpostavimo pa lahko še Kurzweilove napovedi o tem, kdaj bo UI dosegla človekovo inteligenco (Kurzweil, 2006).

Leta 2017 je skupina Googlovih raziskovalcev predstavila novo vrsto arhitekture transformatorjev⁴ (Vaswani et al., 2017), ki je omogočila razvoj velikih jezikovnih modelov (angl. large language model - LLM), in s tem sprožila razcvet *generativne umetne inteligence* (GUI). Leta 2018 je OpenUI objavil članek, ki pojasnjuje, kaj so generativni predtrenirani transformatorji (GPT) (Radford & Narasimhan, 2018), originalni pogovorni klepetalnik GPT-1 pa ni bil deležen velike medijske pozornosti. Posodobljene različice GPT so bile nato izdane v rednih časovnih presledkih, pri čemer je število parametrov naraščalo v milijarde, kar je predstavljalo vse večje in bolj zapletene in predvsem bolj zanesljive modele, ki so lahko obdelali vse več podatkov. Toda šele novembra 2022, z izdajo ChatGPT-3.5, je generativna umetna inteligenca dosegla zavest javnosti. Nenadoma je bila dinamika ustvarjalnega

⁴ Transformatorji so vrsta arhitekture nevronske mreže, ki preoblikuje ali spremeni vhodno zaporedje v izhodno zaporedje. To počnejo z učenjem konteksta in sledenjem odnosov med komponentami zaporedja.

uničenja otipljiva in Suleymanov prihajajoči val je bil viden s prostim očesom (Suleyman & Bhaskar, 2023)⁵.

1.3.2 Od digitalne pedagogike do generativizma

Z zamenjavo omrežja statičnih spletnih strani z informacijami, ki tečejo v eno smer, lahko uporabniki v spletu 2.0 sodelujejo in delijo informacije, komentirajo, ustvarjajo spletne strani, pošiljajo mnenja, objavljajo bloge in komunicirajo. Žal pa mnogi izobraževalci še vedno delujejo v okviru operativnih modelov, ki bi jih lahko poimenovali Izobraževanje 1.0 (z uporabo filozofije spleta 1.0), to je bolj analogno behavioristično izobraževanje kot izobraževanje 2.0, digitalno, konstruktivistično in konektivistično izobraževanje.

Digitalna pedagogika

Učinkovito poučevanje, ne glede na modaliteto, mora vedno temeljiti na pedagogiki. Iz raziskav vemo, da na učni proces vpliva vrsta dejavnikov, vključno s predhodnim znanjem, organizacijo znanja, motivacijo za učenje, razvojem kompetenc, intencionalno prakso in ciljno usmerjenimi povratnimi informacijami, vključujočim in podpornim učnim okoljem ter samousmerjenim učenjem (Ambrose et al., 2010; Sharples, 2019). *Digitalna pedagogika*, kot temelj izobraževanja 2.0, se osredotoča predvsem na to, kako uporabljati digitalne tehnologije za podporo, omogočanje in olajševanje procesa učenja in poučevanja (Aberšek, 2014, Kordigel Aberšek, Aberšek, 2021). Povezani digitalni svet, v katerem zdaj živimo, se je začel z internetom in svetovnim spletom, vendar je bil zgrajen na dolgi zgodovini povezanega izobraževalnega računalništva. Njegovi temeljni teoriji učenja sta bili *konstruktivizem* in *konektivizem*, slednji se je pojavil v zgodnjem obdobju spleta 2.0 in izpostavlja vse pomembnejšo vlogo tehnologije v izobraževanju pospremljeno z pedagogiko, osredinjeno na učence.

⁵ Mustafa Suleyman, soustanovitelj DeepMind in Inflection UI, vodja umetne inteligence pri Microsoftu, je svojo knjigo za leto 2023 naslovil *The Coming Wave* (Suleyman & Bhaskar, 2023). V njem je trdil, da svet ni pripravljen na val močnih novih tehnologij umetne inteligence in je "problem zadrževanja" opredelil kot izziv za našo dobo. Suleyman še zdaleč ni bil prvi, ki je uporabil to metaforo za opis pretresa, ki sledi uvedbi prelomne nove tehnologije v družbo.

Konektivizem temelji na ideji, da se učimo, ko vzpostavljamo povezave med različnimi "vozlišči" informacij ter ustvarjamo in vzdržujemo povezave tako, da oblikujemo znanje (Downes, 2005). Poudarja pomembno vlogo, ki jo ima tehnologija v učnem procesu, in enostavnost, s katero tehnologija omogoča učencem dostop do informacij v digitalni dobi. Konektivizem je zato osredinjen na učence, saj prenaša učne odgovornosti z učitelja na učenca tako, da bi slednji razumeli informacije, ki jih odkrijejo v digitalnem svetu. Ta interakcija z digitalnim svetom pomeni, da konektivizem obravnava učenje tudi kot družbeni proces, ne le kot nekaj, kar je izključno notranje in individualno, ampak nekaj porazdeljenega in dinamičnega (Stodd et al., 2023, str. 106). Ob tem predlaga, da učenci z učenjem krmarijenja po mreži virov in mrežah virov skupaj z vrstniki prevzamejo odgovornost za svoj intelektualni razvoj. To vključuje razvoj kritičnih digitalnih veščin, vključno z digitalno pismenostjo, mrežno pismenostjo in kritičnim ocenjevanjem informacij, ki prihajajo iz različnih in pogosto nasprotujočih si virov. V tem kontekstu postane vloga učitelja predvsem ustvarjanje in oblikovanje učečih se skupnosti ter ustvarjanje spodbudnih učnih okolij. Konektivizem je zato postavil temelje za pristope, ki jih lahko poimenujemo tudi *izobraževanje 4.0* in izobraževanje, ki ga omogoča umetna inteligenca. Leta 1966 si je Patrick Suppes, pionir računalniškega učenja, zamislil prihodnost, v kateri bodo učenci imeli dostop do ogromnih količin znanja ob pomoči računalniških mentorjev. Napovedal je, da bodo milijoni šolarjev imeli dostop do tega, kar je sin Filipa Makedonskega Aleksander užival kot kraljevo pravico: osebne storitve učitelja, ki je tako dobro obveščen in odziven kot Aristotel (Suppes, 1966). Kdaj, kako in ali sploh bo to možno v bližnji prihodnosti za celotno populacijo pa je naslednje temeljno vprašanje, na katero bo stroka poskušala poiskati odgovor v naslednjih letih.

Generativizem

Danes je vse bolj jasno, da bo prihodnost izobraževanja opredeljena tudi s sodelovanjem z GUI na vseh področjih in da bo moralo priti do intenzivnega sodelovanja med obema vrstama inteligence, človeške in umetne za ustvarjanjem nekega novega družbenega reda vključno z izobraževanjem (Pratschke, 2024). GUI je pripravljena spremeniti prakso na vseh področjih. Ko se praksa spremeni in GUI postane del procesa raziskovanja in odkrivanja, bo preoblikovala tudi naše družbo z izobraževanjem na čelu. Toda kaj to pomeni v praksi? Kako lahko sodelujemo z GUI pri oblikovanju učenja? *Generativizem opisuje simbiotski pristop k oblikovanju in*

zagotavljanju učenja v sodelovanju z GUI. Temelji na načelu učenja kot procesa in na nekaterih najpomembnejših in najvplivnejših učnih teorijah in pristopih našega časa:

- **konstruktivizmu**, ki poudarja potrebo po aktivnem učenju za konstruiranje pomena; **konektivizmu**, ki poudarja vlogo digitalnega omrežja, v katerem so učenci aktivni akterji pri tem učenju;
- **socialnem učenju**, ki poudarja pomen skupnosti in sodelovanja;
- **Izkustvenem učenju**, ki poudarja vlogo učenja z delom in razmišljanjem o tej izkušnji; in **generativnem učenju**, ki predstavlja učenje kot proces ustvarjanja smisla z uporabo shem.

Sooblikovanje in izvajanje procesa izobraževanja z GUI je opredeljujoča značilnost generativizma kot prakse, pri čemer se znanje ustvarja ob delu v sodelovanju z GUI, z učnimi dejavnostmi, ki so sooblikovane, olajšane in ocenjene z GUI. Generativizem lahko povzamemo kot:

1. sooblikovanje učne izkušnje v sodelovanju z GUI;
2. soizvajanje učnih dejavnosti in ocenjevanja v sodelovanju z GUI; in
3. ocenjevanje učenja kot procesa v sodelovanju z GUI.

Generativizem uporablja uveljavljene digitalne okvire kot temelj, ki zagotavlja metodološki pristop, kognitivno modeliranje, ki ga potrebujemo za oblikovanje procesa učenja v in za učna okolja, ki jih podpira GUI. To bo imelo pomembne posledice za širjenje znanja, saj bo način poučevanja tesno prepleten z vsebinami, s tem, kaj poučujemo. Učitelji morajo učence pripraviti na prihodnost, v kateri je ustvarjanje, raziskovanje in kreiranje strokovnega znanja na področju sodelovanja med človekom in umetno inteligenco izjemno pomembno. Z več inteligencami kot viri strokovnega znanja, se odmikamo od vloge samostojnega učitelja kot nosilca vsebinskega znanja k modelu vodnika, usmerjevalca in tvorca ustvarjalnih učnih okolij. Zasnova učenja in kognitivnega modeliranja mora odražati to spreminjajočo se vlogo učitelja od nosilca vsebinskega znanja in kognitivnega modeliranja učnih priprav in scenarijev, do tistega, ki spodbuja proces konstruiranja znanja, tako pri ljudeh (učencih) kot tudi pri učnih okoljih (UI). Ta vloga bo verjetno postala bolj dorečena, ko bomo prilagodili izobraževanje temu sodelovanju, saj mora biti temeljni cilj izobraževanja učeče pripraviti na svet, v katerem bodo živeli. Učenci

morajo biti sposobni delati z GUI, ne le kot z orodjem za učinkovitost na delovnem mestu, temveč kot sredstvom za raziskovanje in ustvarjanje znanja tako na svojem študijskem področju kot tudi v vsakodnevnem življenju. Ob tem se nam lahko zastavi vrsta vprašanj, kot:

- Kako lahko oblikujemo izobraževalne izkušnje, ki spodbujajo kritično ustvarjanje smisla, ki je potreben pri delu s to nastajajočo in še vedno globoko pomanjkljivo tehnologijo?
- Kako izkoristiti generativne in družbene privilegije GUI, hkrati pa še vedno vcepiti temeljne družbene vrednote ob veččinah kritičnega razmišljanja, kreativnosti in raziskovanja?

Ena od potencialnih možnosti bi lahko bil tudi okvir *digitalnega kognitivnega modeliranja*, ki je popolnoma primeren za generativne in družbene dostopnosti GUI. Digitalno modeliranje se v tem kontekstu nanaša na posnetke, ustvarjene za uporabo v katerem koli digitalnem mediju, vključno z računalniki, telefoni in tabličnimi računalniki. V širšem pomenu pa termin digitalno modeliranje izdelkov, zajema oblikovanje digitalnih izdelkov ali storitev kot celote, vključno ne le z uporabniškim vmesnikom in uporabniško izkušnjo, temveč tudi s splošno strategijo, funkcionalnostjo in značilnostmi izdelka.

1.3.3 Generativno učenje

Študentje, ki delajo z GUI, morajo razviti sposobnost razumevanja, kaj in kako GUI ustvarja, in to integrirati v širši kontekst svojega znanja, pri čemer morajo GUI videti kot pomočnika s katerim morajo sodelovati s položaja kritične distance in ob tem razvijati kritično mišljenje in kritično presojo (Pratschke, 2024). To učenje mora biti osredotočeno na človeka, vendar ob pomoči GUI. Digitalne prakse združujejo pogled konektivizma na študente kot agente v digitalnem omrežju, z njihovo osredotočenostjo na konstruktivizem (znanje si mora vsak ustvariti sam) in na aktivno učenje, kjer pa mora biti proces sam po sebi socialen in odvisen od interakcije z drugimi v omrežju, torej sodelovalen. Ponovno si lahko zastavimo temeljna vprašanja, ki nas bodo pripeljala do odgovorov:

- Kako lahko oblikujemo učenje, ki spodbuja učence, da sodelujejo z GUI v takem aktivnem in družbenem procesu ustvarjanja smisla?
- Kako lahko presežemo zgodnjo fazo eksperimentiranja in produktivno uporabljamo GUI za razvoj kritičnega in razvijajočega se strokovnega znanja?

Za odgovore na ta vprašanja lahko izbiramo med številnimi okviri za oblikovanje poučevanja/učenja, kot na primer model ADDIE, razvit na državni univerzi Florida leta 1975. Izpostavlja pet ključnih faz v oblikovanju poučevanja: *analizo, oblikovanje, razvoj, izvajanje in vrednotenje*.

Obstaja pa tudi veliko drugih modelov oblikovanja, najbolj priljubljeni pa so

- Gagnejevih 9 dogodkov poučevanja, ki učni proces organizira kot 9-stopenjski proces;
- model zaporednega približevanja (SAM), ki je bil razvit kot alternativa ADDIE;
- Merrillova prva načela poučevanja, ki definirajo univerzalna načela, ki so skupna vsem učinkovitim modelom oblikovanja (osredotočenost na naloge; aktivacija; predstavitev; uporaba; integracija); in
- model Dicka in Careyja, znan tudi kot model sistemskega pristopa, ki predpostavlja, da vsi elementi učne izkušnje tvorijo medsebojno povezan sistem (inštruktor, študenti, učna gradiva in dejavnosti, tehnike poučevanja in učno okolje), ki opredeljuje rezultat.

Najbolj znan in eden od prvih okvirjev učenja in poučevanja pa je seveda Bloomova taksonomija (Bloom et al., 1956; Gogus, 2012) in vseprisotna hierarhija, ki jo prepoznajo vsi učitelji. Bloomov okvir ni toliko okvir za oblikovanje učenja, v večji meri predstavlja stopnje razvoja v treh ključnih učnih domenah: kognitivni, afektivni in psihomotorični. Kako bi se ta proces lahko razvil, ko se učenec ukvarja s pogovornim strokovnjakom GUI? Kakšen je proces, s katerim izpopolnjujemo te veščine? Katere (učne) dejavnosti moramo oblikovati, da jih razvijemo? Didaktične rešitve imajo pri tem ključno vlogo. V Tabeli 1 je tako prikazan prehod Bloomove taksonomije izobraževanja 1.0 do izobraževanja 4.0.

Generativna preobrazba

Digitalna in generativna preobrazba nedvomno spreminjata nabor spretnosti, ki so potrebne za uspeh v svetu. Standardni digitalni ekosistem fizične ustanove (šole, fakultete) vključuje učne sisteme upravljanja (angl. Learning Management System - LMS), ki se včasih imenuje virtualno učno okolje (VUO). Vključuje pa tudi vsa povezana zunanja orodja, na katera se institucija naroči. Ta mreža digitalnih orodij, platform in virov, ki raste od začetka leta 2000, je jedro digitalnega ekosistema institucije, ki ponuja osebno ali kombinirano ponudbo.

Zdaj živimo v svetu, kjer je generativna umetna inteligenca povsod okrog nas. To pomeni, da je strokovno znanje o kateri koli temi, kadarkoli, na kateri koli napravi, za vsakogar, ki ima dostop do ekosistema umetne inteligence, vedno dostopen le '*en klik stran*'. GUI je zdaj integriran v večino orodij, platform in naprav, ki jih uporabljamo in vidimo, ko se ozremo naokrog. Tehnologija se še naprej izboljšuje, raznolikost modelov še naprej raste, hkrati pa se hitro nadaljuje integracija GUI v druga orodja Web 4.0, kot so humanoidni roboti. Jasno je, da so raznolikost modelov GUI, ali celo v prihodnosti umetna splošna inteligenca (USI), multimodalnost (delo v smeri svetovnega modela) in integracija (v programsko in strojno opremo) ključne teme za učitelje. GUI, Industrija 4.0 oz. Družba 5.0, je digitalno mrežo orodij Web 2.0, preoblikovala v ekosistem, poln več inteligenc in prisotnosti v obliki avatarjev, chatbotov in robotov. To je bistvena razlika od digitalne dobe (*Peti val* (1990–2020): *ekosistem umetne inteligence* je namesto skladišča orodij *kibernetski svet*, kjer je umetna inteligenca stalno prisotna, s katero je mogoče soustvarjati znanje. Zato moramo k vključevanju GUI pristopiti drugače kot k digitalni tehnologiji v preteklosti, predvsem zato, ker ekosistem umetne inteligence ne izboljšuje le tradicionalnega izobraževanja, temveč nam omogoča, da ustvarimo nekaj povsem novega. *In to, ali to želimo ali ne, sploh ni več vprašanje. Obdobje šoka, frustracije in depresije je že zdavnaj za nami.*

1.3.4 UI in izobraževanje

Učenje je veliko več kot le potek dela, ki ga je treba narediti učinkovitejšega. To je potovanje v ustvarjanje smisla in odkrivanje, ki ga zaznamujejo trenutki čudenja in naključja. Je kognitiven in afektiven proces. Generativno učenje torej ni odvisno le od tega, kako so informacije predstavljene učencem (tj. metod poučevanja), ampak

tudi od tega, kako jih učenci poskušajo osmisлити (tj. učne *strategije* ali *učne dejavnosti* za spodbujanje tega smisla). Ta proces je ključnega pomena: generativno učenje je tudi aktivno učenje, ki ga navdihuje konstruktivizem, saj od učencev zahteva, da se ukvarjajo z znanjem v obliki gradiv (od učitelja, besedila ali rezultatov GUI) in da ta rezultat integrirajo v oblikovanje lastnega znanja, tako da ga lahko nato uporabijo v drugem kontekstu. Ta proces gradnje znanja, ki ga je Wittrock imenoval *ustvarjanje smisla*, je temeljnega pomena pri delu z GUI, saj naj bi spodbudil kritično razmišljanje, preizpraševanje rezultatov in zavestno sodelovanje s procesom ustvarjanja smisla.

Kljub vsemu pa je treba razmislek o učinkovitosti, ki nakazuje, da lahko orodja GUI optimirajo učni proces, dopolniti z nekaterimi pomisleki. Raziskave kažejo, da čeprav so pomočniki UI lahko privlačni, morda ne pomagajo pri učenju in bi, nasprotno, kompetence učečih lahko nazadovale in ne napredovale (Darvishi et al., 2024). V tem kontekstu, je zagotovo najprej treba omeniti *kognitivno lenobo*⁶, ki bo zagotovo sledila, ko bo UI delila prostor - tako fizični kot virtualni - s sošolci (ljudmi). Vendar, ko bodo učitelji usvojili didaktične kompetence v zvezi z uporabo umetne inteligence, utemeljene na sodobni pedagogiki (generativizmu), bodo motnje v izobraževanju, ki so jih mnogi napovedali, veliko manj verjetne.

GUI je še vedno nastajajoča tehnologija v zgodnjih fazah razvoja, kar pomeni, da je treba z njo ravnati previdno. Ti temelji se lahko uporabijo kot izhodišče za izgradnjo novega modela izobraževanja, ki je osredotočeno na človeka, vendar ga opredeljuje sodelovanje z generativno umetno inteligenco. Ponuja nam potencial, da povečamo lastno inteligenco na doslej nepredstavljljive načine, hkrati pa prebujajo naše najgloblje strahove, da bi postali zastareli. Je edini obetavni prehod v Družbo 5.0, v super pametno družbo.

Posledice za izobraževanje bodo globoke. Učitelj ne bo več edini nosilec znanja, pisna naloga kot dokaz o naučenem ni več izvedljiva in učilnica ni več središče dejavnosti.

⁶ Kognitivna lenoba nas sili k temu, da se izognemo iskanju možnih rešitev (ostanemo na Kahnemanovem intuitivnem Sistemu 1, in ne vklopimo Sistema 2, poglobljenega, statističnega načina razmišljanja) in s tem zapremo vrata ustvarjalnosti. Kognitivna lenoba povzroča blage kognitivne motnje, vključuje težave s spominom, jezikom in presojo.

Zaključek

Medtem ko pogovor s stroji ni povsem nov, ima dejstvo, da lahko vsak nestrokovnjak zdaj govori neposredno z računalnikom, kot da bi govoril s človekom, nekaj nepričakovanih rezultatov, ki nas prisilijo, da razmišljamo o delu s tehnologijo na nov način. Od novembra 2022 lahko uporabniki po vsem svetu neposredno komunicirajo s temi »nezemeljskimi« inteligencami, postavljajo vprašanja in se pogovarjajo o vsaki temi pod soncem, ne da bi bilo potrebno tehnično znanje ali usposabljanje. Govor in jezik pa je sinonim za človeško civilizacijo – pravzaprav je naša sposobnost jezikovnega izražanja in govora tista, ki nas ločuje od živali. Svoje zgodbe pripovedujemo v človeškem jeziku, v govorni zgodovini in jih zapisujemo v pisni zgodovini. Generativna umetna inteligenca je močna prav zaradi svoje sposobnosti, da posnema ta naš najbolj izpopolnjen človeški izum (Bozkurt, 2023). Preboj na področju generativne umetne inteligence je zato zabrisal meje med človekom in strojem – ali, natančneje, naše dožemanje teh meja. V spominu na obdobje Elize⁷: inženir Blake Lemoine je bil julija 2022 odpuščen iz Googla, potem ko je javno trdil, da je LaMDA (jezikovni model za dialoške aplikacije) podjetja dosegel zavest (Tiku, 2022).

1.4 Zgodovinski oris umetne inteligence

Že starogrški filozofi so si postavljali vprašanja, kot so: kako se učim, pomnim, zaznavam? Kakšna je vloga jezika? Ali duša preživi smrt telesa? Kakšen je odnos med duševnimi in telesnimi procesi? Ali smo svobodni v svojih odločitvah? Kako vemo, da nekaj vemo? Pomembni antični filozofi Platon, Aristotel, Leukip, Demokrit, Epikur in Lukrecij so v svojih delih nanje odgovarjali na različne načine, delo pa so nadaljevali Descartes, Spinoza, Hume, Kant in mnogi drugi.

Na kratko in zelo poenostavljeno bi lahko stališča do narave duha prikazali z dvema sklopoma domnev:

- *Prvi pristop*: Po eni strani se nam zdi, da je duševnost povsem drugačna od vsega ostalega v svetu in mora zato imeti poseben ontološki status.

⁷ ELIZA je zgodnji računalniški program za obdelavo naravnega jezika, ki ga je med letoma 1964 in 1967 na MIT razvil Joseph Weizenbaum. Ustvarjen za raziskovanje komunikacije med ljudmi in stroji. Za bolj podrobno razlago glej (Aberšek, 2014).

- *Drugi pristop:* Po drugi strani sprejemamo, da je človek kot bitje del narave ter da so procesi, ki se dogajajo v njem, vključno z duševnimi procesi in fizičnimi procesi, prav tako del te narave.

Na podlagi naturalističnega, tj. redukcionističnega ali antiredukcionističnega pogleda lahko zaključimo, da potrebujejo kognitivna znanost in mentalna stanja obrambo pred dualizmom, če želimo obdržati globalni in nepristranski filozofski pogled na zastavljeni problem. Če želimo razumeti duh, način človeškega mišljenja in razmišljanja po kriterijih današnje znanosti, se moramo osredotočiti na fizikalistični pogled na znanost, ki upošteva, da so mentalna stanja v nekem pogledu fizična stanja, čeprav jih ne moremo (jih ne znamo) reducirati na fizična stanja, torej upošteva splošno *razlagalno vrzel* v psihologiji. (Horst, 2007: 28–31)

Naj na kratko povzamemo zgodovinski hevristični razvoj umetne inteligence:

Začetki 1943–1952:

- 1943: J. McCulloch, W. Pitts: model umetnega nevrona;
- 1949: D. Hebb: formulacija pravila modificiranja vezi med dvema nevronoma;
- 1950: A. Turing: Turingov test, strojno učenje, genetski algoritmi, podprto učenje;
- 1951: Minsky in Edmonds: prva nevronska mreža sestavljena iz 40 nevronov (vakuumске cevi).

Zgodnji entuzijazem in velika pričakovanja 1952–1969:

- 1952: Samuel (IBM): igra dama;
- 1956: J. McCharty (MIT): darthmoutska konferenca, The Dartmouth Summer Research Project on Artificial Intelligence;
- 1956: Newell, Shaw in Simons: prvi AI program, Logic Theorist (LT) – kratek dokaz teorema iz *Principia Mathematica*;
- 1956: Newell in Simon (GPS): prvi program, ki je upošteval človeški način razmišljanja;
- 1958: McCharty (MIT): programski jezik LISP;
- 1960–1962: Widrow in Hoff: ADALINE;

- 1962: F. Rosenblat: perceptron, dokaz konvergence;
- 1965: J. Weizenbaum: program ELIZA;
- 1965: Robinson: pravilo resolucije;
- 1966: Quillian: semantične mreže;
- 1969: Minsky, Papert: perceptron omejitev NN.

Strezničev 1969–1973:

- rezultati zgodnjih sistemov so se izkazali za slabe na širšem razponu problema ali pri težjih nalogah;
- zgodnji programi so vsebovali malo ali sploh nič znanja, njihov uspeh pa je temeljil na enostavnih sintaktičnih manipulacijah;
- npr. strojno prevajanje (spodbudila ga je potreba po prevodih ruskih del (1957) pri tekmovanju v vesoljski tehniki – Sputnik), je temeljilo na sintaktičnih transformacijah in menjavanju besed na osnovi angleške in ruske slovnice;
- drugi velik problem je bila prilagodljivost oz. prožnost mnogih problemov, ki so jih poskušali reševati z uporabo umetne inteligence;
- pred razvojem teorije izračunljivosti so menili, da se bo »skaliranje« na večje probleme rešilo s povečanjem računske moči;
- 1969: Minsky in Papert: *Perceptron*. (Minsky, Papert, 1969)

Na znanju temelječi sistemi 1970–1979:

- 1975: Minsky: teorija okvirjev (angl. frames theory);
- napredek v obdelavi naravnih jezikov;
- PROLOG;
- sistemi: DENDRAL, MYCIN;

Umetna inteligenca je industrija 1980–1990:

- od nekaj mio. USD v letu 1980 do milijarde USD v leti 1988;
- 1982 McDermott: DEC R1 ekspertni sistem;
- 1980 povratak nevronske mreže (Werbos – backpropagation algorithm – delta pravilo).

Umetna inteligenca 1990–2030

Področje umetne inteligence (UI) s konca 20. in začetka 21. stoletja je končno doseglo nekaj ciljev, ki si jih je zastavilo na samem začetku razvoja (Newel, Minsky). Umetna inteligenca se je pričela uspešno uporabljati na celotnem tehnološko razvojnem področju, čeprav pod različnimi imeni, ki so se tekom let vedno bolj poenotila pri uporabi termina umetna inteligenca. Nekaj uspeha je bilo posledica povečane računalniške moči, nekaj pa je bilo doseženega z osredotočanjem na posebne izolirane probleme in njihovim preučevanjem z najvišjimi standardi znanstvene odgovornosti. Kljub temu je bilo na tem področju le malo soglasja o razlogih za neuspeh umetne inteligence pri izpolnitvi sanj o inteligenci na človeški ravni (Kurzweil, 2006). Vsi ti dejavniki skupaj so pripomogli k razdrobitvi umetne inteligence na konkurenčna podpodročja, osredotočena na posebne probleme ali pristope, včasih celo pod novimi imeni, ki so prikrivala očrnen pedigre termina umetna inteligenca.

Mejniki in Moorov zakon

11. maja 1997 je Deep Blue postal prvi računalniški sistem za igranje šaha, ki je premagal aktualnega svetovnega šahovskega prvaka Garija Kasparova. Super računalnik je bil specializirana različica ogrodka, ki ga je izdelal IBM, in je bil sposoben obdelati dvakrat več potez na sekundo kot med prvim dvobojem (ki ga je Deep Blue izgubil), celo do 200.000.000 potez na sekundo. Leta 2005 je robot Stanford zmagal na tekmovanju DARPA Grand Challenge z avtonomno vožnjo 131 milj po neobdelani puščavski poti.

Ti uspehi niso bili posledica neke nove revolucionarne paradigme, temveč predvsem preproste uporabe inženirskih orodij in veščin ob izjemnem povečanju hitrosti in zmogljivosti računalnikov. Pravzaprav je bil računalnik Deep Blue 10-milijonkrat hitrejši od računalnika Ferranti Mark 1, ki ga je Christopher Strachey naučil igrati šah leta 1951. To dramatično povečanje se meri z Moorovim zakonom, ki pravi, da se približno na vsaki dve leti hitrost in pomnilniška zmogljivost računalnikov podvojita, kar je posledica podvojitve zmogljivosti visoko zmogljivih integriranih vezij. Temeljni problem »surove računalniške moči« bo tako počasi premagovan.

UI v 21. Stoletju

Posamični namenski algoritmi in agenti, ki so jih prvotno razvili raziskovalci in razvijalci UI, so se začeli združevati v večje sisteme. Umetna inteligenca je rešila veliko zelo težkih problemov in njihove rešitve so se izkazale za uporabne v celotni tehnološki industriji, kot so podatkovno rudarjenje, industrijska robotika, logistika, prepoznavanje govora, Googlov iskalnik in mnogi drugi. Toda področje umetne inteligence je bilo za ta napredek deležno le malo ali nobenih zaslug za te uspehe v devetdesetih letih prejšnjega stoletja in v začetku leta 2000. Številne največje inovacije umetne inteligence so bile zmanjšane na status le še enega predmeta v orodju računalništva. Številni raziskovalci umetne inteligence v devetdesetih letih prejšnjega stoletja so svoje delo namerno imenovali z drugimi imeni, kot so informatika, na znanju temelječi sistemi, kognitivni sistemi ali računalniška inteligenca. Deloma je to morda tudi zato, ker so menili, da je njihovo področje bistveno drugačno od umetne inteligence, a tudi nova imena so pomagala pri pridobivanju sredstev ali kot je leta 2005 poročal New York Times: "Računalniški znanstveniki in programski inženirji so se izogibali izrazu umetna inteligenca v strahu, da bi nanje gledali kot na sanjače."

Minsky je verjel, da so bili odgovor na to osrednji problemi umetne inteligence, kot je zdravorazumsko sklepanje, zanemarjeni, medtem ko je večina raziskovalcev sledila stvarjem, kot so komercialne aplikacije nevronske mreže ali genetskih algoritmov. Za Raya Kurzweila (2006) je bil problem računalniška moč in je z uporabo Moorovega zakona napovedal, da se bodo stroji s človeško inteligenco pojavili do leta 2029. Jeff Hawkins je trdil, da raziskave nevronske mreže ignorirajo bistvene lastnosti človeških možganov in dajejo prednost enostavnim modelom, ki so bili uspešni pri reševanju preprostih problemov. Bilo je še veliko drugih razlag in za vsako je potekal ustrezen raziskovalni program.

Od sedaj do bližnje prihodnosti – 2010—2030:

V prvih desetletjih 21. stoletja so bili dostop do velikih količin velikih podatkov, cenejši in hitrejši računalniki ter napredne tehnike strojnega učenja uspešno uporabljeni za številne probleme v znanosti, tehniki, gospodarstvu in mnogih drugih panogah. Uporaba velikih podatkovnih baz je začela segati tudi na druga področja, kot so modeli usposabljanja v ekološki industriji in različne aplikacije v ekonomiji.

Napredek pri globokem učenju (zlasti razvoj globokih konvolucijskih nevronske mreže in ponavljajoče se nevronske mreže) je spodbudil napredek in raziskave na področju obdelave slik in videa, analize besedila ter prepoznavanja vzorcev in govora.

Simon in Newell zapišeta:

»Na svetu so stroji, ki mislijo, ki se učijo in ustvarjajo. Celo več, te njihove sposobnosti se bodo intenzivno razvijale, dokler v bližnji prihodnosti ne bo obseg problemov, ki jih bodo sposobni obdelovati in reševati, enak ali celo večji, kot je obseg, ki ga je sposoben reševati človeški um.«
(Kurzweil, 1998: 69)

Danes se lahko ponovno zastavi logično vprašanje in raziskovalna hipoteza: Ali je človeštvo že tako daleč in ali so stroji že prestopili človeške meje?

Po novejši literaturi je strojna inteligenca sprejeta znotraj področja sistemskih razvijalcev in je namenjena končnim uporabnikom, vendar je to večinoma še vedno slabo raziskano področje in še to, kar je raziskano, je po navadi strogo varovana skrivnost in ni splošno dostopno. Če na kratko izpostavimo današnje usmeritve raziskovanja na področju UI, lahko povzamemo naslednje:

Globoko učenje

Globoko učenje je veja strojnega učenja, ki modelira visokonivojske abstrakcije v podatkih z uporabo globokega grafa s številnimi plastmi obdelave. V skladu z univerzalnim aproksimacijskim izrekom globina ni potrebna, da bi nevronska mreža lahko aproksimirala poljubne zvezne funkcije. Kljub temu obstaja veliko težav, ki so skupne plitvim omrežjem (kot je prekomerno opremljanje), ki se jim globoka omrežja pomagajo izogniti. Kot take so globoke nevronske mreže sposobne realistično ustvariti veliko bolj zapletene modele v primerjavi s svojimi plitvimi dvojniki.

Najsodobnejše arhitekture globokih nevronske mreže in generativne umetne inteligence (GUI) se lahko včasih celo kosajo s človeško natančnostjo na področjih, kot je računalniški vid, algoritmi za obdelavo jezikov (veliki jezikovni modeli), ki jih poganjajo pametni iskalniki, pa zlahka premagajo ljudi pri odgovarjanju na splošna

nepomembna vprašanja (kot sta IBM Watson ali Google LaMBDA, ChatGPT ali DeepSeek), nedavni razvoj globokega učenja pa je prinesel osupljive rezultate pri različnih tekmovanjih z ljudmi (Aberšek, Flogie, Pesek, 2023).

Velike podatkovne baze

Velike podatkovne baze se nanašajo na zbirko podatkov, ki jih v določenem časovnem okviru ni mogoče zajeti, upravljati in obdelati z običajnimi programskimi orodji. Gre za ogromno zmogljivosti odločanja, vpogledov in optimizacije procesov, ki zahtevajo nove modele obdelave. V dobi velikih podatkovnih baz veliki podatki pomenijo, da se namesto naključne analize (vzorčne raziskave) za analizo uporabijo vsi podatki. Značilnosti obdelave velikih podatkovnih baz (ki jih predlaga IBM) so: obseg, hitrost, raznolikost, vrednost, resničnost. Strateški pomen tehnologije obdelave velikih podatkov ni obvladovanje ogromnih podatkovnih informacij, temveč specializacija za te pomembne podatke. Z drugimi besedami, če velike podatkovne baze primerjamo z industrijo, je ključ do doseganja dobičkonosnosti v tej panogi povečanje na primer procesne zmogljivosti podatkov in uresničitev dodane vrednosti podatkov z obdelavo.

Splošna umetna inteligenca (SUI) ali močna UI

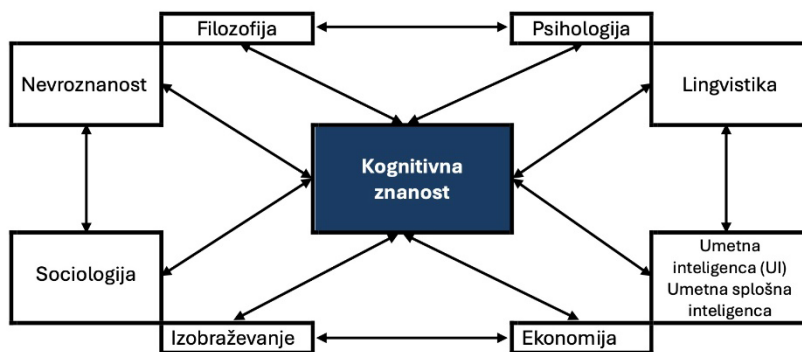
Splošna umetna inteligenca je zmožnost reševanja kakršnega koli problema, in ne le iskanje rešitev le za določen problem (GUI). Splošna umetna inteligenca (angl. artificial general intelligence — AGI) je program, ki lahko uporabi inteligenco pri najrazličnejših težavah na skoraj enak način kot ljudje. Splošna umetna inteligenca se imenuje tudi *močna UI* ali sintetična inteligenca v nasprotju s *šibko UI* ali *ožko UI*. (Akademski viri rezervirajo "močno umetno inteligenco" za stroje, ki so sposobni izkusiti zavest in čutiti).

Temeljne modele, kot so veliki modeli umetne inteligence, usposobljeni za ogromne količine neoznačenih podatkov, ki jih je mogoče prilagoditi širokemu naboru nalog na nižji stopnji, so začeli razvijati leta 2018. Modeli, kot so GPT-3, ki ga je izdal OpenAI leta 2020, in Gato, ki ga je izdalo podjetje DeepMind leta 2022 ali LaMBDA, ki ga je razvil Google, so bili opisani kot pomembni mejniki na poti do umetne splošne inteligence. Danes, po letu 2022 sta splošno najbolj znana Chat GPT in

DeepSeek. Slednji je ugledal luč sveta nekako ob koncu leta 2024 in začetku leta 2025.

Izhodišča kognitivnega modeliranja

Kognitivna znanost je zrasla iz gibanja kibernetike v petdesetih letih dvajsetega stoletja in je od takrat doživela številne paradigmske spremembe. Študiji kognitivne znanosti, ki so v zadnjih dvajsetih letih zacveteli po vseh pomembnih svetovnih univerzah, so večkrat obarvani s kakšno od konstitutivnih disciplin (na primer kognitivna lingvistika, kognitivna nevroznanost, kognitivna antropologija ipd.), vse bolj pa postaja jasno, da edino enakopravna obravnava vseh področij zagotavlja celostno obravnavo duševnih procesov.



Slika 1.1: Struktura kognitivne znanosti

Struktura kognitivne znanosti je prikazana na sliki 1.1.

V zadnjih desetletjih doživlja kognitivna znanost skokovit razvoj. Še posebej zahvaljujoč napredku nevroznanosti počasi začenjamo slutiti, da se je mogoče znanstveno lotiti raziskovanja duševnih pojavov in celo zavesti, torej področja, ki je bilo še do nedavnega rezervirano samo za mistike. Svoja spoznanja skušajo kognitivni znanstveniki prenesti tudi v prakso, še posebej na področju učenja in poučevanja, procesov kolaborativnega dela in na področju računalniškega strojnega učenja in odločanja.

Konekcijonizem je nastal sredi osemdesetih let kot alternativa *klasični računalniški paradigmi* v kognitivni znanosti, ki je nastala v petdesetih letih in je temeljila na analogiji med digitalnim računalnikom in umom, po kateri je mišljenje neke vrste simbolno računanje. Konekcijonistični modeli, ki se zgledujejo zlasti po možganih in njihovi fiziološki in funkcionalni zgradbi, se od klasičnih, simbolnih modelov razlikujejo po nekaterih bistvenih značilnostih, kot so vzporedna obdelava podatkov, vsebinsko-asociativni pomnilnik in porazdeljene reprezentacije. Ker tako klasični kot tudi konekcijonistični modeli kažejo mnoge pomanjkljivosti, sta Horgan in Tienson (1996) predlagala nov teoretični okvir za kognitivno modeliranje, t. i. teorijo dinamičnih sistemov, ki se opira na konekcijonizem, vendar prevzema tudi nekatera temeljna spoznanja klasicizma, zlasti pomen sintakse ali skladnje. Dela s tega področja so namenjena filozofski obravnavi konekcijonizma in računalniških modelov duha, ki omogočajo osvetlitev dolgoletnih filozofskih vprašanj v novi luči: *Kaj je zavest? Kje so izvori znanja? Kakšni so mehanizmi zaznavanja, spominjanja, učenja? Kakšna je vloga jezika? Kako rešujemo probleme? Kakšen je odnos med fizičnim in psihičnim, med umom in telesom?* Vse bolj zmogljivi računalniki so namreč postali zelo močno orodje, ki omogoča empirično preverjanje teoretičnih zamisli in ustvarjanje modelov, ki v večji ali manjši meri ustrezajo človekovim kognitivnim funkcijam.

1.5 Sklep

V knjigi se posvečamo sedanjosti in prihodnosti in na podlagi tega bomo iskali in utemeljevali paralele med naravno in umetno inteligenco na epistemološkem in ne toliko na hevrističnem nivoju. Pri tem se bomo nekoliko posvetili paradigmam in strategijam učenja kot enemu od pomembnih delov inteligentnega delovanja, njihovim vhodnim podatkom in izidom. Če naj računalniški programi delujejo inteligentno, morajo vsebovati predstavo o svetu in o tem, kako si razlagamo te predstave in kako delujejo osnovne zakonitosti v njem. Vsak ustvarjalec umetne inteligence mora vedeti, kaj je to znanje in kako do njega pridemo. Prav zaradi tega se nekateri tradicionalni filozofski problemi pojavljajo tudi v umetni inteligenci. Pri tem se bomo osredotočili predvsem na paradigme, strategije in mehanizme učenja in usvajanja novih znanj (tako človeka, kot stroja), ki jih bomo poglobljeno obdelali v revidirani kibernetični didaktiki. Kibernetična didaktika temelji na kibernetičnih načelih in spoznanjih o krmiljenju in reguliranju tehničnih, organskih in družbenih sistemov. Med slednje uvrščamo tudi pedagoški sistem. Običajno je ta didaktika definirana kot znanost, ki raziskuje, kako s tehnološkimi in matematičnimi

pripomočki spodbuditi in voditi učne procese oziroma ravnanje nekega učnega sistema (v izhodišču je to predvsem človek, v naši revidirani kibernetični didaktiki pa bo to tako človek kot tudi stroj), da bi bili učni cilji doseženi optimalno. Ker kibernetična didaktika ne odraža trenutnih potreb in možnosti sodobnih spoznanj s področja umetne inteligence, jo bomo za naš konkreten primer *posodobili* in v njej upoštevali tudi vse interakcije stroj – človek oz. naravna – umetna inteligenca.

Učenje je generativna dejavnost in izobraževanje 4.0 ne sme pomeniti uporabe tehnologije za krepitev zastarelih načinov poučevanja in učenja. Inovacije, ki smo jih videli v digitalnem izobraževanju v zadnjih dveh desetletjih, tvegajo, da bodo te inovacije ogrožene, če uporabimo GUI za vrnitev k zastarelim modelom izvajanja izobraževanja. GUI bi moral namesto povečanja avtomatizacije v izobraževanju omogočiti več interakcije, aktivnejše učenje in več personalizacije. Zato se morajo učitelji osredotočiti na oblikovanje aktivnega, sodelovalnega, konstruktivističnega in konektivističnega učenja, ki spodbuja ustvarjanje smisla in kritično vključenost, ne pa ustvarjanje vsebine. Pedagogika umetne inteligence se pojavlja in jasno je, da je prihodnost izobraževanja v sooblikovanju z GUI. Mednarodne organizacije so začele ustvarjati okvire za vključevanje GUI (Miao & Holmes, 2023; Miao in Tawil, 2024), vendar je bilo zelo malo razprav o tem, kako dejansko vključiti to sooblikovanost v izobraževanje. Generativizem uporablja okvire iz digitalne dobe kot temelj za izgradnjo nečesa novega, ki ponuja pristop k oblikovanju in zagotavljanju učnih izkušenj, ki so družbene, sodelovalne, skupnostno usmerjene in osredotočene na človeka/osredinjeno na učenca.

Zastavljen problem je še vedno izziv in nanj vse do sedaj ni bil dan niti približen, kaj šele enoznačen odgovor.

»Potrebno bo še veliko utemeljevanja, saj je to v nekem pogledu filozofska terra incognita. Zato moramo poiskati nekakšne nove razlage, raziskati možnosti reduktivnega in/ali nereduktivnega fizikalizma, dualizma ter eliminativizma in poiskati primer, kjer bomo s pomočjo znanstvenega pluralizma raziskali različne pristope in teorije k razlaganju duha, zavesti, botenj in ne nazadnje delovanja celovite človeške inteligence. Samo na podlagi tega bomo lahko potegnili paralele med človeško in umetno inteligenco.« (Horst, 2007:195)

Tu pridemo do točke, ko si moramo zastaviti naslednje vprašanje: Kako vključiti koncept zavesti v razumevanje človeškega duha? Še vedno ni popolnoma znano, kako koncept zavesti dodati k razumevanju človeškega duha (Anderson, 2007). Anderson je že leta 1973 razvil ACT (angl. Adaptive Cognitive Theory) teorijo, ki je

unitarna teorija kognicije in temelji na človeškem asociativnem spominu (Anderson, Bower, 1980). Na osnovi te teorije je pripravil tudi ustrezno programsko podporo. Od takrat sta se teorija in programska podpora razvijali vse do današnje šeste verzije, ACT-R 6.0⁸.

Če skušamo koherentno interpretirati termin *zavest*, potem lahko rečemo, da npr. ACT-R poenostavljeno povedano predstavlja računalniško simulacijo človeških možganov z vsemi njenimi funkcijami in z njeno celotno strukturo. Skratka, če ACT-R ponazarja možgane, potem lahko logično predpostavimo, da je odgovor na našo osnovno dilemo, *ali lahko človeški duh oz. človeški um formaliziramo in zapišemo v simbolni obliki*, pritrdilen. Če pa ta trditev drži in če je zavest tista, ki razlikuje človeka od stroja, potem lahko sklepamo, da obstajajo tudi vse možnosti, da postanejo stroji še bolj (če ne popolnoma) podobni človeku. Skratka osnovni problem je, kaj je zavest oz. kako jo formalizirati, torej zapisati v simbolni ali kakšni drugi obliki (npr. kot dinamični sistem) ter nato kako koncept zavesti dodati k celotnemu konceptu človeškega duha oz. k človeškim možganom.

To bomo poskušali doseči s pomočjo ACT-R in z uporabo teorije dinamičnih sistemov ter z njima kognitivno modelirati določene miselne procese in na osnovi tega dokazovati ali zavračati osnovno hipotezo o možnosti oz. nemožnosti vključevanja zavesti in drugih višjerednih kognitivnih procesov v aplikacije umetne inteligence.

Izhajajoč iz povedanega bi torej osnovna filozofska dilema kot izhodišče zastavljenega problema lahko bila: *Ali lahko človeški duh, formaliziramo in reduciramo na jezik znanosti in če da, kar bomo tudi dokazovali, kakšna orodja za to potrebujemo?* Skratka, iskali bomo odgovore na to, kako duh formalizirati, zapisati v fizikalni obliki ter nato kako koncept duha dodati k celotnemu konceptu duh – telo.

Če in ko bomo poznali odgovor na to vprašanje, bo odgovor na izhodiščno in ključno vprašanje, ali se lahko umetna inteligenca razvije do stopnje človeške in ali stroji lahko postanejo podobni ljudem (ne samo po videzu, temveč predvsem po njegovi funkciji), enostaven. Pri tem vprašanju pa je ključno predvsem to, ali se bodo

⁸ Vsa literatura in program so dostopni na uradni spletni strani »gibanja«: <http://act-r.psy.cmu.edu/>

lahko stroji tudi zavestno odločali, čutili bolečine, kar pomeni, ali bomo pri UI in kognitivnem modeliranju lahko dosegli splošno umetno inteligenco oz. močno UI ipd.

Naj na koncu omenimo le še izhodiščno premiso:

Po letu 2022, po našem splošnem zavedanju o obstoju velikih jezikovnih modelov - LLM (Large Language Models) in GUI, nič več ne bo tako, kot je bilo! Zavedati se bomo morali, da nismo več edino inteligentno bitje v tem našem svetu. Zato je ključno, da razumemo temeljne principe delovanja tako človeške kot tudi umetne inteligence in le na podlagi tega razumevanja ustrezno kognitivno modeliramo učna okolja 21. stoletja. In prav razumevanju kognitivnih, miselnih procesov, je namenjena ta knjiga (Kordigel Aberšek, Aberšek, 2021).

2 Osnovna filozofska in kognitivna izhodišča

Namen tega poglavja je predstaviti nekaj idej s področja filozofije duha s posebnim poudarkom na naturalističnem stališču, domnevi, da je lahko duh »naturaliziran,« in to tako, da ga zapišemo z jezikom znanosti, fizike, nevroznanosti ali z jezikom drugih naravoslovnih znanosti.

Filozofija duha je veja sodobne analitične filozofije, ki proučuje naravo duha, odnos med mentalnimi in fizičnimi pojavi, ki ga imenujemo problem duha in telesa, mentalno vzročnost in problem zavesti. *Dualizem* in *monizem* sta dve glavni šoli duha, ki poskušata rešiti problem duha in telesa. *Dualizem* je stališče, ki trdi, da sta duh in telo na nek kategoričen način ločena drug od drugega. Zasedujemo ga lahko od Platona v zahodni filozofiji ter šole Shankhya in Joge v vzhodni, hinduistični filozofiji. Najbolj natančno je dualizem definiral René Descartes v 17. stoletju. Dualizem se nadalje deli na dve osnovni smeri, *dualizem substanc*, ki trdi, da je duh neodvisna materija, in *dualizem lastnosti*, ki meni, da je duh skup neodvisnih lastnosti, ki izhajajo iz možganov, vendar ga ne moremo reducirati nanje, niso pa tudi od njih distinktivna substanca.

Monizem izhaja iz prepričanja, da duh in telo nista ontološko različni vrsti entitete. To perspektivo so v zahodni filozofiji predlagali Parmenid, Leukip, Demokrit in Epikur v 5. stoletju pred našim štetjem in so jo kasneje posvojili tudi racionalisti in paralelisti, npr. Barucha Spinoza in na poseben način tudi Locke in Hume.

2.1 Zgodovinski razvoj filozofije duha

Že starogrški filozofi so si postavljali vprašanja kot: *Kako se učim, pomnim, zaznavam? Kakšna je vloga jezika? Ali duša preživi smrt telesa? Kakšen je odnos med duševnimi in telesnimi procesi? Ali smo svobodni v svojih odločitvah? Kako vem, da nekaj vem?* Pomembni antični filozofi Platon, Aristotel, Leukip in Demokrit so v svojih delih nanje odgovarjali na različne načine, delo pa so nadaljevali Descartes, Spinoza, Hume, Kant in mnogi drugi.

Najprej se bomo osredotočili le na Descartesova izhodišča, ki jih je zapisal takole:

»Tako vidimo na primer, kako se glave, malo potem, ko so bile odrezane, še gibljejo in grizejo zemljo, čeprav ni več duše v njih. Nadalje sem razložil, kakšne spremembe se morajo zgoditi v možganih, da povzročajo bedenje, spanje in sanje; kako morejo svetloba, vonji, okusi, toplota in vse druge lastnosti zunanjih predmetov vtisniti s posredovanjem čutov možganom različne ideje; kako mu morejo lakota, žeja in druga notranja občutja prav tako zbuditi posebne vtise. Dalje sem povedal, kaj je treba imeti za osrednje čutilo, v katero se vsi ti vtisi stekajo; kaj za spomin, ki hrani sledove teh vtisov; kaj za fantazijo, ki lahko vtise raznoliko spreminja in iz njih sestavlja nove ideje. Fantazija je tudi, ki more s primernim dovajanjem animaličnih hlapov v mišice premikati ude telesa na mnogo različnih načinov, bodi da trkajo na telesna čutila predmeti zunanjega sveta, bodi da ga vzbujajo notranja čustvovanja. Tako se morejo naši udi gibati pod učinkom fantazije, ne da bi jih vodila volja. To se nikakor ne bo zdelo čudno tistim, ki vedo, koliko različnih *avtomatov* ali gibljivih strojev more izdelovati človekova spretnost z uporabo prav majhnega števila kosov v primeri z velikim množtvom kosti, mišic, živcev, arterij, ven in vseh drugih delov, ki so v telesu sleherne živali: gledali bodo na telo, kakor na kak stroj, ki so ga naredile roke boga in ki je zaradi tega brezprimerno bolje urejen in premore mnogo čudovitejša gibanja kot katera koli priprava, ki jo zmore človeška iznajdljivost. Na tem mestu sem se še posebej pomudil in hotel prikazati, da bi nam, ko bi res obstajali taki stroji, ki bi po organih in zunanji obliki bili v vsem enaki opici ali kateri že drugi živali brez razuma, nikakor ne bilo moč spoznati, da se po svojem ustroju razlikujejo od teh živali. Ko pa bi eksistirali stroji, ki bi bili podobni našemu telesu in bi posnemali naša dejanja, kolikor je to pač praktično mogoče, bi nam se vedno ostali dve povsem zanesljivi poti do spoznanja, da zavoljo tega ti stroji se niso resnični ljudje. Prva teh poti je, da bi ti stroji nikoli ne imeli daru govora ne drugih znakov, ki jih uporabljamo ljudje za izražanje svojih misli drugim. Sicer si je lahko predstavljati kak stroj, ki je tako zgrajen, da izgovarja določene besede, in celo to, da izreka marsikaj, kar

povsem ustreza prisotnosti predmetov, ki delujejo na njegove zunanje organe. Če se na primer dotaknemo določenega mesta, bo stroj vprašal, kaj hočemo od njega, če drugega, bo zavpil, da mu prizadevamo bolečino, in podobno. Toda stroj ne more iz lastnega nagiba nikoli tako postavljati besed, da bi smiselno odgovarjal na vse, kar se izreče vpricho njega, kot zmorejo to kateri koli ljudje, pa naj bodo se taki topoglavci. Druga preizkusna pot pa je naslednja: čeprav bi mogli taki stroji opravljati mnoge reci prav tako dobro ali celo bolje kakor kdo izmed nas, bi nujno odpovedali v nekaterih drugih rečeh, po katerih bi bilo moč spoznati, da njih delovanje ne izvira iz razuma, marveč edinole iz namestitve njih organov. V tem ko je namreč razum univerzalni instrument, ki more rabiti ob sleherni priliki, morajo biti organi strojev za vsako posebno delovanje posebno razvrščeni. Zaradi tega je praktično nemogoče, da bi bilo v enem stroju dovolj raznovrstnih delov, ki bi mogli sprožiti njegovo delovanje v vseh življenjskih prilikah, kakor dela to za našo dejavnost razum. Prav po teh dveh preizkusnih sredstvih moremo tudi spoznati razliko med človekom in živaljo.» (Descartes, 1957: 81–82)

V današnji, z različnimi tehnologijami zaznamovani družbi, si le stežka predstavljamo, kako so bili Descartesovi sodobniki ogorčeni ob njegovi primerjavi človeškega organizma z avtomatom in koliko prevrata in mentalnega naboja je vsebovala njegova mehanistična koncepcija. Da bi bolje dojel Descartesovo prisposodobno primerjave človeškega organizma s strojem, navedimo naslednja, za ta način razmišljanja izredno značilna mesta:

»A kakor pač vstopajo ti *animalični hlapovi* v možganske prekate, pronicajo odondod v pore njegove substance, od tod pa v živce: tukaj pa, kakor pač vstopajo v ene bolj kakor v druge, ali celo samo skušajo vstopiti, imajo moč, da spremenijo obliko mišic in povzročajo tako gibanje vseh udov. Kakor ste morebiti videli v špiljah in fontanah naših kraljevih parkov, zadošča tista sila sama, po kateri teče voda, ko privre iz vira, da se morejo gibati razni stroji, da, celo zaigrati na neke orgle ali izgovarjati nekake besede, po različni razporeditvi cevi, po katerih teče voda. – In v resnici se dajo živci stroja, ki ga opisujem, najbolj primerjati cevem teh fontan; njegove mišice in kite raznim drugim pripravam in vzmetem, ki posredujejo gibanja; njegovi analitični hlapovi vodi, ki jih spravlja v pogon in katerim je srce izvirek, cerebralna votlina pa rezervoar. Nadalje dihanje in druga podobna delovanja telesa, ki so naravna in normalna in jih povzroča gibanje *animaličnih hlapov*, ne potekajo nič drugače kot v uri ali mlinu, ki ju lahko redni pretok vode poganja naprej in naprej. Zunanji predmeti, ki po svoji goli prisotnosti delujejo na čutne organe tega stroja in ga tako naravnavajo, da se giblje na različne načine po razporedbi posameznih delov njegovih možganov, ti zunanji predmeti, pravim, so kakor tuji gledalci, ki so prišli v nekatere duplje teh fontan, in povzročajo – ne da bi na to sploh mislili – gibanja, ki potekajo v njihovi prisotnosti: v te špilje ne morejo namreč vstopiti, ako ne hodijo po ploščah, razporejenih tako, da ima, denimo, njih bližanje kaki kopajoči se Diani ta učinek, da se Diana skrije med trsje; in če jo še dalje zasledujejo, jim stopi nasproti kak Neptun in jim zagrozi s svojim trizobnikom; ako pa se obrnejo drugam, jim skoči na pot kaka morska pošast in jim bruhne vodo v obraz ali kaj podobnega, kakor se je pač zdelo muhavi iznajdljivosti zgraditelja teh fontan. In ko bo naposled v tem stroju nameščena *racionalna duša*, bo imela svoj glavni sadež

v možganih in bo tam uravnavala, podobno mehaniku, ki ravna fontano in mora biti tam, kjer se stekajo vse cevi teh strojev, kadar hoče njih delovanje spraviti kakor koli v pogon, ali pa jih ustaviti ali spremeniti. – Da pa bo vse to lažje razumljivo, bom predvsem govoril o strukturi živcev in mišic in pokazal, kako imajo zgolj zato, ker skušajo animalični hlapovi v možganih stopiti v kak živec, moč, da pri tej priči vzgibljejo kak organ. Potem, ko bom dodal nekaj besed o dihanju in drugih takih preprostih in trdno se ponavljajočih gibanjih, bom povedal, kako delujejo predmeti zunanjega sveta na čutne organe. Po vsem tem bom natančno razložil, kaj se dogaja v možganskih prekatih in porah, kako se tu gibljejo animalični hlapovi in katere so tiste naše funkcije, ki jih mora ta stroj posnemati z njih pomočjo. Ko bi namreč začel pri možganih in bi po vrstnem redu sledil toku animaličnih hlapov, kakor sem to storil glede krvi, se mi zdi, da bi moja razprava ne mogla biti povsod tako jasna.» (Descartes, Traktat o človeku, A.T. XI, 130–132; navedeno po Descartes 1957: 273–275)

Če se še pomudimo pri Descartovem razmišljanju o avtomatu oz. stroju, podobnemu človeku, humanoidnem robotu:

»Ti ljudje [v tem imaginarnem svetu] bodo, kakor smo mi, sestavljeni iz duše in telesa. Potrebno pa je, da vam najprej in ločeno opišem telo, nato – zopet zase – dušo, na kraju pa, da vam pokažem, kako morata ti dve naravi biti povezani in združeni, da sestavljata nam podobne ljudi. – Predpostavljam, da [človeško] telo ni nič drugega nego prstén kip ali stroj, ki ga oblikuje bog ob prizadevanju, da bi ga naredil kar moči podobnega nam, tako da mu podeli na zunaj ne le barvo in obliko vseh naših udov, marveč da namesti tudi v notranjost telesa vse tiste dele, ki so potrebni, da stroj hodi, jé, diha, skratka, da posnema vse tiste naše funkcije, za katere si moremo misliti, da izvirajo iz materije in da so odvisne edinole od razporedbe organov – tako vidimo, da uram, umetnim fontanam, mlinom in drugim podobnim strojem, čeprav so izdelki človeških rok, ne manjka moči, da se gibljejo z lastno sposobnostjo na različne načine. In zdi se mi, da si v stroju, ki ga predpostavljam kot delo božjih rok, ne moremo misliti tolikšne različnosti gibanj in mu prisojati toliko umetnosti, da bi ne imel razloga za domnevo, da bi jih moglo biti v njem še več.« (Descartes, Traktat o človeku, A.T. XI, 119–120; navedeno po Descartes 1957: 275)

Descartes v svojih zgoraj opisanih izvajanjih vpelje termin »animalični hlapovi,« ki predstavljajo del »fantazije« in predstavljajo del človekove nezavednosti, ki usmerjajo različne človeške aktivnosti brez njegove volje. Iz te nezavednosti pa lahko izhajajo lastnosti pri mehanicistični interpretaciji življenja v živih in neživih sistemih. Ostra ločnica, ki jo začrta Descartes med dvema vrstama lastnosti dobi neko posebno obliko takrat, ko se ukvarja z vprašanjem življenja pri ljudeh. Vpelje namreč analogijo med telesom živega človeka ter truplom na eni strani in uro, ki teče, ter tisto, ki je razbita, na drugi strani:

» /.../ telo živega človeka se loči od tistega pri mrtvem človeku ravno tako, kot se loči ura ali drugi avtomat (to je stroj, ki se giblje sam), ko je navita in vsebuje v sebi telesni princip teh gibanj, za katere je bila ustvarjena, skupno z vsem potrebnim za njeno delovanje od te iste ure ali drugega stroja, ko je zlomljena in princip njenega gibanja preneha biti aktiven.« (Strasti duše, CSM I, 329–330, Descartes, 1985)

Običajno Descartesova pojmovanja prikazujejo kot prelom s predhodno tradicijo. To lahko na eni strani prikažemo z znanim primerom iz Aristotelovega dela O duši:

»Zato živo bitje ne zmore bivati, če mu je hrana odvzeta. Ker se tu nahajajo tri stvari: tisto, kar je hranjeno, in tisto, s čimer se hrani, in to, kar hrani, je to, kar hrani, *prvotna duša*, prehranjevano je telo, ki ima to dušo, s čimer pa se hrani, je hrana.« (Aristotel, 1993, II, 4, 416b.)

Za Aristotela je prehranjevanje kot posebna lastnost živih organizmov nekaj, kar izhaja iz psihičnega, duševnega principa, medtem ko je pri Descartesu to povsem fizični ali materialni princip. Na drugi strani imamo Akvinca, ki pravi, da so občutki neko vmesno stanje med materialnim in imaterialnim. Čuti so pasivne moči, ki so podvržene spremembam na osnovi zunanjih dražljajev. Pri tem pa:

»Sta dve vrsti spremembe, ena naravna in druga duhovna (*spiritualis*). Naravna sprememba je, ko je forma izvora spremembe sprejeta v spreminjajočem se objektu na naravni način bivanja, kot je dobljena toplota, ko se nekaj segreva. Duhovna sprememba je, ko je forma izvora sprejeta v spreminjajočem se objektu na duhovni način tako, kot je forma barve sprejeta v očesu, ne da bi oko s tem postalo obarvano. Čutila lahko delujejo le, če je tu duhovna sprememba, pri čemer je intenca čutne forme prišla v čutni organ. Drugače, če bi naravna sprememba zadostovala, bi vsa naravna telesa čutila vedno, kadar bi bila podvržena spremembi.« (ST 1, 78, 3c, Akvinski, 1989)

Pri Descartesu je čutenje v celoti prestavljeno v območje imaterialnega oz. duhovnega. Lastnosti, kot so barva, zvok, toplota, vonj, okus, niso modusi razsežnega, ampak obstajajo le v našem mišljenju:

»Toda barve, vonji, okusi in tako naprej so, sem opazil, zgolj določeni občutki, ki obstajajo v moji misli in so tako različni od teles, kot je bolečina različna od oblike in gibanja orožja, ki jo proizvede.« (Šesta zbirka odgovorov, CSM II, 297, Descartes, 1984)

Telo živega človeka in delujoča ura imata določen ponotranjen princip gibanja. Hkrati pa so njuni fizični deli organizirani tako, da omogočajo funkcioniranje. Torej sta na delu dve specifični komponenti Descartove metode pojasnjevanja življenja. Po eni strani je to funkcionalistični vidik, ki poudarja, da je življenje lastnost in ne substanca. V bistvu je to množica funkcij, ki je skupna vsem in ne le tistim organizmom, ki so živi. Na drugi strani pa je mehanicistična pojasnjevalna komponenta, ko pravi, da je možno za vsako vrsto žive stvari pojasniti, kako izvršuje svoje življenjske funkcije. To pomeni, da za Descartesa ne obstaja nikakršna metafizična razlika med živim strojem in neživim avtomatom. Oba imata notranji princip gibanja. Razlika je le v tem, da so deli živega stroja urejeni tako, da zagotavljajo uresničevanje življenjskih funkcij, medtem ko deli avtomata (še) niso tako urejeni.

Vidimo torej, da za Descartesa ne obstaja metafizična razlika med človekom in strojem. Kljub temu pa je poudarjal, da je naša sposobnost misliti to, kar je razlika med človekom in strojem, kar človeka ločuje od drugih živih bitij. Izrekel je: »*Cogito ergo sum* – *Mislim, torej sem*«. Danes pa se je ta stari latinski pregovor nekoliko preoblikoval in mu je na koncu, za besedo *sum*, dodana še beseda *machina*. *Mislim, torej sem stroj* (Aberšek, 2013).

Naj zaključim to razpravo s citatom Richarda L. Gregoryja:

»Imenovati žival stroj je namigovati, da njen način delovanja ni bistveno drugačen, kot je delovanje strojev, ki jih lahko konstruira in naredi človek ... S tem povemo, da so živi organizmi v bistvu sestavljeni tako, da jih v principu lahko razumemo tako, kot inženirji ali fiziki razumejo svoje sisteme«. (Gregory, 1961: 310)

2.2 Novejši filozofski pogledi na problem duha in telesa

Na podlagi *reduktivistično fizikalističnega* pogleda lahko zaključimo, da psihologija in duševna stanja potrebujejo obrambo, če jih želimo ohraniti kot samostojni vedi. Razlaga redukcionistov, kako bi psihologijo umestili v naravoslovje, je bila zavržena s strani filozofije znanosti, saj je v psihologiji takšnost pomemben element. Če želimo razumeti duh, način človeškega mišljenja in razmišljanja po kriterijih današnje

znanosti, se moramo osredotočiti na tak pogled na znanost, ki priznava in upošteva splošno *pojasnjevalno vrzel*⁹ v psihologiji. Kaj to pomeni za filozofijo duha?

Če obstaja vrzel, je redukcionistično upanje popolnoma nespametno in predvsem neutemeljeno. Še toliko bolj je, če vemo, da obstaja razlagalna vrzel celo znotraj fizike same ali med fiziko in kemijo, molekularno in klasično genetiko ipd., nepriznavanje takšne vrzeli znotraj psihologije dvomljivo. Resnično pa to pomeni, da so v psihologiji potrebne povezave med duhom in možgani, tako kot so znane splošne povezave med dvema področjema naravoslovnih znanosti.

Nasprotno pa je to na prvi pogled dobra novica za zagovornike pojasnjevalne vrzeli. Odkritje, da je razlagalna vrzel vsakdanja, je voda na mlin tako dualistom kot tudi nereduktivnim materialistom. Vendar pa, kar na eni strani dobimo, na drugi strani takoj izgubimo. Torej, če upoštevamo, da je naravni svet združljiv z razlagalno redukcijo, potem unikatna, univerzalna razlagalna vrzel v primeru duha predstavlja pomemben problem. Če je psihološka vrzel ena od mnogih razlagalnih vrzeli, potem ta problem ni več tako pomemben. In če zagovorniki vrzeli v kombinaciji z negativno EMC¹⁰ trdijo, da mentalni fenomen ne supervenira nad dejstvi fizike, se njihov položaj zaplete predvsem zaradi dejstva, da takšno odstopanje potem velja tudi za druge naravoslovne znanosti, npr. za kemijo in biologijo, tako da tudi ti dve znanosti ne supervenirata nad fiziko. Celotno nereduktivni fizikalisti (kot npr. Davidson) so mnogokrat predpostavljali, da je duh edinstven in da ga ne moremo poenostaviti. (Horst, 2007)

Ta problem pa je še posebej kočljiv za dualiste, ki se morajo spopadati z zelo neprijetno in bistveno dilemo. Če namreč vztrajno zagovarjajo negativni EMC, potem niso več dualisti, temveč postanejo pluralisti višjega reda. Če pa se odpovedo negativnemu EMC, se odpovedo glavnemu argumentu, ki ga tradicionalno uporabljajo pri podpori dualizma. S stališča znanstvenega pluralizma, dualizem z upoštevanjem negativnega EMC postane nekonsistenten; dualizem *brez* negativnega EMC je sicer konsistenten, vendar izgubi enega od najpomembnejših argumentov dualizma.

⁹ Pojasnjevalna vrzel je razkorak med tem, kar je objektivno dostopno, to se pravi dostopno z gledišča tretje osebe in s tem dostopno znanosti, ter med zavestnimi občutki, ki so dostopni subjektivno, torej z gledišča prve osebe.

¹⁰ EMC (angl. Epistemology – to – Metaphysics Connection Principal) – ločimo pozitivno in negativno EMC. Pozitivno EMC bi lahko ponazorili s primerom: Ali ima redukcija A na B za posledico, da A supervenira nad B? Pri negativni EMC pa se postavi vprašanje: Ali ima nereducibilnost A na B za posledico, da A ne supervenira nad B?

Nereduktivni fizikalizem se sooča s podobnimi problemi, saj večina nereduktivnih fizikalistov zavrača negativni EMC, zato jim ne grozi odklon od radikalnega ontološkega pluralizma. Prav tako nereduktivni fizikalisti niso ogroženi zaradi zavračanja interteoretične redukcije v znanosti, saj nereduktivni fizikalisti mnogokrat predpostavljajo, da se lahko nementalne lastnosti skladajo z osnovno fiziko. Vendar pa vse le ni tako samo po sebi umevno in enostavno. Zgodovinsko gledano je osnovni razlog za razmišljanje, da je lahko fizikalizem pravilen, to, da lahko fenomene posameznih znanosti fizično *razložimo*. Tako dolgo kot predpostavljamo, da to velja za naravoslovne znanosti, postaja tudi logična induktivna hipoteza, da fizikalna dejstva določajo vsa dejstva, tudi če so to popolnoma izolirani otočki sveta in četudi jih ne znamo razložiti. Vendar se mora v luči znanstvenega pluralizma nereduktivni pluralizem vrniti k osnovam in se na novo vprašati, katera je najboljša metafizična razlaga znanosti, upoštevajoč dejstva splošne pojasnjevalne vrzeli. S tega stališča pa ni jasno, ali obstaja kakšen razlog, da dajemo prednost fizikalistični razlagi pred drugimi alternativnimi razlagami. Tako kot dualizem brez negativnega EMC, je tudi nereduktivni fizikalizem sicer *konsistenten*, vendar pa ni dovolj jasno, zakaj bi verjeli, da je prav on pravilen.

Na kratko in zelo poenostavljeno bi lahko stališča do narave duševnosti prikazali z dvema sklopoma domnev. Po eni strani se nam zdi, da je duševnost povsem drugačna od vsega ostalega v svetu in mora zato imeti poseben ontološki status. Po drugi strani pa sprejemamo, da je človek biološko bitje in del narave ter da so procesi, ki se dogajajo v njem, vključno z duševnimi procesi, naravni procesi.

Prvi pristop lahko podpremo z naslednjimi domnevami:

- *Nerazsežnost mentalnega*: Zdi se nam, da naši občutki, želje, čustva in misli nimajo prostorskih značilnosti, medtem ko imajo stvari okoli nas prostorske značilnosti, kot so velikost, oblika in mesto v prostoru.
- *Neposreden pristop do mentalnih stanj*: Do spoznanja o naših duševnih stanjih in procesih prihajamo neposredno, s samoopazovanjem (introspekcijo). Kadar nas boli roka, je to že dovolj, da vemo, da smo v tem stanju. Če pa želimo spoznati fizične predmete v naši okolici, jih spoznavamo s pomočjo čutov, instrumentov, eksperimentov itd.
- *Problem intencionalnosti*: Kadar mislimo, verjamemo ali si želimo, so naše misli, prepričanja in želje vedno o nečem (npr. želimo, da bi šli na počitnice). Lahko

so tudi o nečem, kar ne obstaja (npr. o okroglem kvadratu). Težko pa verjamemo, da bi to veljalo tudi za zgolj fizične stvari.

- *Problem zavesti*: Ob branju zgodb nas preplavi žalost. Včasih čutimo neznosno bolečino v križu. Le kako bi lahko ta naš občutek analizirali s fizikalnimi znanstvenimi izrazi, ki so dostopni vsem?

Drugi pristop podpirajo naslednje domneve:

- če so procesi v človeku fizični procesi, potem je zelo verjetno, da so tudi duševni procesi naravni fizični procesi;
- iz zgodovine razvoja znanosti smo se naučili, da to, da nekega pojava še ne znamo razložiti z znanostjo – s sredstvi in metodami, ki so nam v tistem času na voljo – še ne pomeni, da je ta pojav nekaj nefizičnega, nematerialnega;
- iz vsakdanjih izkušenj lahko sklepamo na medsebojno povezanost duševnega in telesnega.

Pri domnevah prvega pristopa filozofi večinoma izhajajo iz svojih vsakdanjih izkušenj, domneve drugega pristopa pa so vezane predvsem na spoznanja iz znanosti. Zagovorniki naturalistične teorije (domneve drugega pristopa), nekakšni znanstveni optimisti, morajo poiskati razlago, kako so ti pojavi mogoči v naravi. Zagovorniki dualistične teorije (domneve prvega pristopa) pa menijo, da ne obstajajo prepričljivi dokazi za naturalizem mentalnega in zagovarjajo razlage, ki vključujejo nadnaravnost in posebnost duševnega stanja. Filozofska razmišljanja že dolga niso poglavitna metoda spoznanj o duševnih procesih. Ko se je na prelomu 19. in 20. stoletja po zgledu empiričnih naravoslovnih znanosti psihologija osamosvojila od filozofije, je vodilna nemška psihološka šola poudarjala introspekcijo kot temeljno znanstveno metodo psihologije. Njihova temeljna hipoteza je bila, da je psihologija raziskovanje »fenomenologije« človekove duševnosti, zato so si prizadevali podati popoln opis zemljevida duševnosti, kot se prikazuje subjektu. Problem tega pristopa je bil, da so bili zbrani podatki o »čutnih atomih« odvisni od občutljivosti in sposobnosti »introspekcionistov«, tako da so ti dobivali različne izide, kar je omajalo »znanstvenost« pristopa. No, k zatonu je ob teh »notranjih« vzrokih pripomogla predvsem sprememba v raziskovalni klimi v Evropi in ZDA. Prevladovati je začela pozitivistična naravnost tako v znanosti kot v humanistiki. Posledica tega je bila, da je bila zavest izrinjena iz psiholoških raziskav, prav tako pa tudi metoda

introspekcije, ki naj bi bila neznanstvena in preveč obtežena z vprašljivo metafiziko. Uveljavil se je behaviorizem z načelom, da mora psihologija kot znanstvena disciplina preučevati le tisto, kar je mogoče raziskovati z opazovanjem, torej opazovano vedenje. Behavioristično metodologijo si najlažje predstavljamo, če si zamislimo človeka kot »črno škatlo«, ki sprejema podatke iz okolja in se nanje odziva. Znanstvenika zanima, kakšna je povezava med dražljaji in vedenjem, ne spušča pa se v raziskovanje same »črne škatle«, saj je po njegovem vse, kar se dogaja znotraj nje, izven ravni znanstvene psihologije. Na ta način so se behavioristični psihologi pri razlagi izognili tako sklicevanju na nevrofiziološka kot tudi na mentalna stanja. Vendar se je pokazalo, da je projekt, po katerem bi lahko vse pojme z mentalno vsebino definirali s pomočjo operacijskih definicij (kot dejanske ali potencialne reakcije organizma na določene dražljaje), neuresničljiv. Psihologi so se zato obrnili k vsakdanji zdravorazumski psihologiji, ki za razlago vedenja uporablja duševna stanja (prepričanja, želje, strahove in namere), in je nudila okvir, ki ga je bilo treba napolniti z znanstveno teorijo. S tem se je odprl nov prostor raziskovanja kognitivnih procesov (zaznavanje, učenje, oblikovanje pojmov, reševanje problemov ...).

2.2.1 Vprašanje redukcije

Iz zgornje razprave je razvidno, da so stališča naturalističnih teorij, ki sprejemajo to, da so duševni procesi bioloških organizmov udeleženi v nevronske procesih, različna. V prvem obdobju se je filozofsko raziskovanje omejilo predvsem na tiste vidike duševnosti, kjer je funkcionalistični pristop obetal uspešne razlage (propozicionalne naravnosti, intencionalnost), zanemarili pa so kvalitativne vidike. Kasneje se je, tudi zaradi velikega razvoja nevroznanosti in problema duševne vzročnosti, ponovno odprlo vprašanje redukcije. Provokativen pogled je prispeval filozof znanosti in nevroznanstvenik John Bickle. Predlagal je drugačen pristop k reševanju problema telo – duh, kjer bi tradicionalni filozofski pristop nadomestili z iskanjem odgovora glede možnosti interteoretske redukcije (npr. psihologije na nevrobiologijo) za posamezne duševne pojave. (Bickle, 1998) Tako se po njegovem mnenju lahko izkaže, da bo empirična znanost pokazala, da nimajo vsi duševni pojavi enakega statusa, saj bo redukcija zavzemala cel spekter, od pojavov, za katere je redukcija možna, do takih, kjer to morda ni mogoče. Kot uspešen primer »redukcije v praksi« navaja primer nevrobiološke redukcije »spominskega obrata«, ki posreduje pri pretvorbi kratkoročnega v dolgoročni spomin. (Bickle, 2003) Nekateri pa v

možnost takšne redukcije dvomijo, saj ugotavljajo, da je to trenutno iluzija. (Bregant et al., 2010) Izpostavljajo, da v Bicklovem primeru molekularne redukcije, tj. redukcije dolgoročnega spomina na njegovo celično-molekularno bazo, izbrana molekularna pot – CREB pot, ni niti zadosten niti nujen pogoj za dolgoročni spomin. Zato je ideja redukcijskih zvez med našim duševnim življenjem in njegovo celično aktivnostjo v tem trenutku bolj iluzija kot realnost.

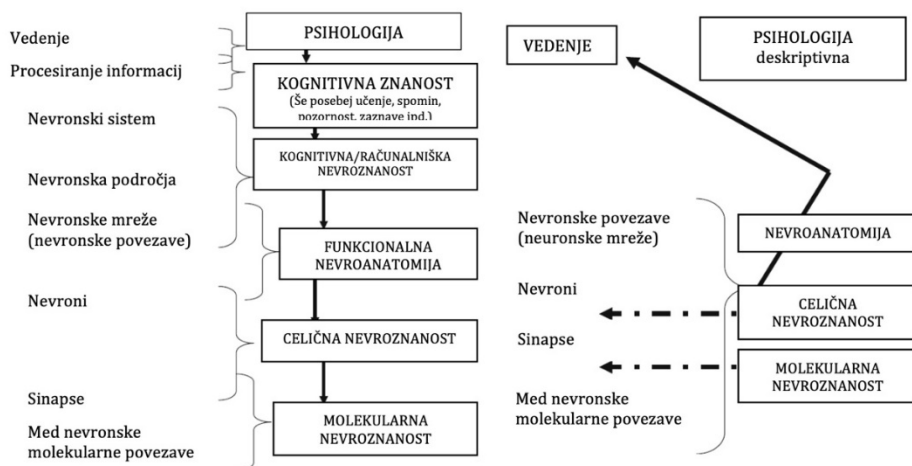
2.2.2 Trenutno stanje redukcije v kognitivni znanosti

Psihonevralna redukcija je v zadnjem času ponovno na udaru. Vendar tokrat s strani prejšnjih zaveznikov: kognitivnih nevroznanstvenikov. Vse bolj popularno postaja razmišljanje o možganih kot o *kompleksnem sistemu*, katerega teoretično pomembne lastnosti izhajajo iz dinamike in nelinearnih povezav med njegovimi sestavnimi deli. *Emergenca*¹¹ naj bi zamenjala redukcijo; razmišlja se, da je redukcija nezmožna razlagati možgane kot kompleksen sistem. Nekateri avtorji pa poudarjajo, da redukcija igra in bo tudi v prihodnosti igrala pomembno vlogo v filozofiji duha, tudi če bo nevroznanost v celoti sprejela ta »kompleksni sistemski« pogled, kar pa je malo verjetno. (Bickle, 2001) Pri natančnih raziskavah kompleksnih nevralnih sistemov bo na določena vprašanja še vedno možno odgovoriti le s pomikom po razlagalni skali od celote k posamičnim delom, skratka z redukcijo. Ta vloga redukcije lahko najde danes svoje mesto še vedno pri približnih, a dominantnih usmeritvah nevroznanosti, kot npr. pri razširjanju uporabe eksperimentalnih teorij in tehnik molekularne biologije in biokemije, kar bi s kakršnimikoli drugačnimi antiredukcionističnimi metodami znotraj kognitivne znanosti bilo skoraj nemogoče.

Ugotavljamo lahko, da je celo skrajni redukcionizem še vedno živ in uspešen pri »molekularni in celični« razlagi kognicije – področju raziskav, ki se nahaja znotraj celične in molekularne nevroznanosti, trenutno najbolj prevladujoče smeri nevroznanosti. (Bickle, 2006) Zaradi predvsem uspešnih in popolnih eksperimentalnih metod in vse bolj natančnih izidov sta vsaj dve trditvi »sodobnih filozofov« (nasprotnikov redukcionizma) in kognitivnih znanstvenikov napačni:

¹¹ Emergenca – v psihologiji, teoriji sistemov, znanosti in umetnosti je pojavnost, način, kako kompleksni sistemi in vzorci izvirajo iz mnogoterosti relativno enostavnih interakcij. Pojavnost je osrednja stvar teorij integrativnih nivojev in kompleksnih sistemov. Lep primer pojavnosti je nastanek snežink, ki ustvarjajo kompleksne simetrične vzorce v fizičnih sistemih ali ustvarjanje peščenih sipin.

- da o delovanju možganov ne vemo veliko;
- da z nižje nivojsko nevroznanostjo ne moremo neposredno razložiti kognicije in kompleksnih vedenj.



Slika A: Shematski prikaz trenutnih standardiziranih pogledov na organizacijske nivoje v nevroznanosti v povezavi s področji znanosti duha – možgani. Na sliki je »korak za korakom« prikazan pogled psihonevralne redukcije, ki jo prikazujejo puščice. Pri tem je redukcija uspešna le, če so značilnosti višjih nivojev organizacije (po odgovarjajočih znanstvenih teorijah) povezane z naslednjimi nižjimi značilnostmi.

Slika B: Shematski prikaz primerjav nevroznanstvenih in behaviorističnih razlag vedenja od molekularnega področja do psihologije. Črtkane puščice predstavljajo nivoje možnih izvajanj eksperimentov, polne puščice pa predstavljajo nivoje, na katerih so možne tudi meritve. Psihologija je veda, ki omogoča bolj deskriptivno generiranje podatkov in ne toliko razlagalno.

Slika 2.1: Redukcija duha na molekularno pot: tolmačenje redukcionizma, prisotnega v sedANJI celični in molekularni nevroznanosti.

Te eksperimentalne prakse raziskujejo neposredno molekularne komponente sestavnih delov celic in genske zapise v nevronih in pri tem neposredno merijo določena vedenja z uporabo splošno priznanih testov eksperimentalne psihologije za študij psiholoških fenomenov, kot so spomin, pozornost ali zaznave. Na spodnji sliki je na kratko prikazan primer redukcije v prevladujoči nevroznanosti. Slika A prikazuje, kako poskušajo znanstveniki s področja raziskovanja odnosa med duhom in telesom prikazati redukcijo korak za korakom. Takšen prikaz redukcije je pretežno sprejet tako s strani filozofije duha kot tudi s strani kognitivne znanosti. Nivoji po sliki 2.1A so zamenjani v razlagalni praksi celične in molekularne kognicije s termini na sliki 2.1B. Ugotovimo lahko, da redukcija, prikazana na tej sliki, ni več sistematična po principu korak za korakom od višjih k nižjim nivojem in prav tako

ne zahteva več razlagalnosti ustreznega nivoja, če zanj poznamo uspešne modele. Ti uspehi veljajo le za redukcije duha na molekularnih poteh in nevronih za njihov anatomske krog, in to je skrajni redukcionizem danes v trenutni nevroznanosti. (Bickle, 2006, Walmsley, 2010)

Zavestni kvalitativni subjektivni občutki (takšnosti, qualia) so vrzel (angl. explanatory gap). Vprašanja o duševnih pojavih so zato že v osnovi interdisciplinarna, ker imamo do duševnih pojavov tudi specifičen neznanstven dostop, tj. subjektivno izkušnjo. V tem je tudi bistvena razlika med kognitivnimi znanostmi in ostalimi znanstvenimi disciplinami. (Markič, 2008; Kordeš, 2004)

2.2.3 Vprašanje emergence

Beseda *emergenca* je izpeljana iz latinskega glagola *emergere*, ki v prehodni obliki pomeni »narediti, da se nekaj pojavi«, v neprehodni obliki pa »pojavljanje oz. nastanek«. *Emergenca* označuje v filozofiji in psihologiji fenomen, da določene lastnosti celote ne moremo razložiti z lastnostmi njegovih sestavnih delov. Tako kot na mnogih področjih je bil vodilni predhodnik pojavnih lastnosti nekega sistema Aristotel s svojo metafiziko:

» ... celota, kot bi se zdelo, ni le slučajna množica, temveč je nekaj več kot le vsota delov ... «
(Aristotle, *Knjižica H*, 1045: 8–10)

Prvi je uporabil pojem *emergenca* v današnjem pomenu George Henry Lewes v svoji knjigi iz leta 1875 z naslovom *Problemi življenja in duha* (Lewes, 1875). Uporabil jo je v zvezi z *emergenco* zavesti. Pomembno nadaljnjo, predvsem filozofsko težo konceptu *emergencizema*, pa so dali britanski *emergencisti*, kot so Samuel Alexander, Conwy Lloyd Morgan in Charlie Dunbar Broad v svoji teoriji o pojavnosti evolucije na začetku 20. stoletja.

Poskušajmo na grobo definirati njen pomen: *emergenca* je spontano izoblikovanje novih lastnosti ali struktur na makronivoju nekega sistema, in sicer kot posledica medsebojnega delovanja njegovih elementov. Pri tem *pojavnih lastnosti* sistema ni mogoče razložiti (jih izpeljati) iz posamičnih izoliranih lastnosti posamičnih elementov, ki sistem sestavljajo, npr. v filozofiji duha nekateri filozofi trdijo, da je zavest pojavna, torej *emergenta* lastnost možganov. Pojavne fenomene opisujejo

različne vede, npr. fizika, kemija, biologija, psihologija ali razne družbene vede. Skratka, pojavne bitnosti oz. entitete (lastnosti ali materije) »nastanejo« iz bolj temeljnih entitet, a so še vedno nove in se jih ne da reducirati nanje same. Če povzamemo z zelo popularnim izrekom: »Celota je mnogo več kot skupek njenih sestavnih delov.«

Dobro je znano, da intuitivna ideja pojavnih lastnosti izvira iz misli, da so čisti fizični sistemi sestavljeni iz delčkov materije in šele, ko ta doseže določeno stopnjo kompleksnosti v svoji organizacijski strukturi, lahko pričnejo izkazovati avtentično nove lastnosti, ki jih ne posedujejo njihovi sestavni delčki. Vprašanje pa se zastavi, ko skušamo narediti to idejo bolj natančno in jo spremeniti v nekaj, kar bo filozofsko in znanstveno uporabno. Za zagovornike pojavnosti je trditev, da je podana lastnost pojavna lastnost nekega sistema, dovolj pomembna in je zadostna razlaga tako o lastnostih tega sistema kot o tem sistemu samem. Vendar pa, kaj *natančno* pomeni, da je nekaj pojavna lastnost? Kaj pomeni reči, da je to »nova« lastnost? Kaj pojavne lastnosti *delajo* potem, ko so se pojavile? To so vprašanja, na katera bomo, če bomo želeli pojavnost obravnavati kot »resno« teorijo, morali odgovoriti.

Ponovno zanimanje za pojavnost danes je posledica razvoja in uporabe teorije kompleksnih sistemov in razprav o združljivosti mentalnih vzročnosti, intencionalnosti ali zavesti s fizikalizmom. V nadaljevanju si to oglejmo nekoliko podrobneje.

OSNOVNE IDEJE IN POSLEDICE

Okrog leta 1990 je ideja o pojavnosti (emergenci) doživela veliko vrnitev po desetletjih splošne zavrnitve in zaničevanja s strani vodilnih analitičnih filozofov. Kaže, da bo pojavnost ostala pomembna tudi v prihodnosti, kar bomo podrobneje obravnavali v naslednjem poglavju. Kaj je na pojavnosti takšnega, da jo dela tako zanimivo za najrazličnejše mislece, z najrazličnejšo splošno izobrazbo in iz različnih področij delovanja; od filozofov, prakticistov najrazličnejših strok, znanstvenih piscev. Ali je to le mikavno filozofsko vprašanje ali morda vprašanje za sociologijo znanosti? Videti je, da ima izraz *pojavnost* neko posebno privlačnost za mnoge mislece; v nasprotju z redukcijo, ki zveni zadrženo in gospodovalno, ima nek vzvišen, zaupljiv zven, saj vidimo izraz pojavnost enostavno, povezan z maloštevilnimi vidnimi vzroki o tem, ali je njena uporaba podprta s konsistentnim,

sprejemljivo poenotenim in porazdeljenim pomenom (in če je tako, kaj je to?). Trenutno je pojavnost predvsem le izraz filozofske stroke. Pod njim si skoraj vsak lahko predstavlja, kar si pač hoče, edini pogoj je, da ima posameznik sam jasno predstavo, kaj naj bi pomenil, in da je predstavljeni koncept zanimiv in teoretično uporaben. (Kim, 2006)

V nadaljevanju bomo poskušali izpostaviti dva pogoja, za katera menimo, tako kot tudi vrsta drugih strokovnjakov na tem področju, npr. Kim, da so potrebne komponente vsakega koncepta pojavnosti, izhajajoč iz zgodovinskih korenin zgodnjega dvajsetega stoletja. Ta pogoja sta *supervenienca*¹² in *nereducibilnost* (angl. irreducibility). Ob tem pa se bomo osredotočili tudi na vprašanje, kaj bi morali narediti, če bi želeli pojavnost pretvoriti v čvrst in metafizično koherenten koncept. Pri tem bo še posebej izpostavljena vzročnost od zgoraj navzdol. Lahko sicer ugotavljamo, da emergenca mentalnega ne more obstajati brez vzročnosti, vendar pa ji le-ta povzroča velike težave.

EMERGENCA IN SUPERVENIENCA

Da bi utrdili koncept pojavnosti, pričnimo z vprašanjem, ali lahko menimo, da so pojavne lastnosti supervenientne nad njihovimi osnovnimi pogoji – lastnostmi nižjega nivoja, iz katerih so se pojavile. Po Van Gulicku (2004) moramo ločiti tri stopnje pojavnosti:

- osnovno vrsto pojavnosti,
- preprosto vrsto pojavnosti in
- radikalno vrsto pojavnosti.

Prva stopnja ponuja relativno enostavno razlago, npr. celoten objekt ima maso 1 kg, pri čemer nobeden njegov del nima te mase, torej mase 1 kg. To seveda ni to, kar imajo v mislih vsi, ki se ukvarjajo s pojavnostjo. Drugo stopnjo bi po Van Gulicku lahko opisali: Celota ima značilnost, ki je drugačna od tistih, ki jih imajo posamezni deli. Tretjo stopnjo pa bi lahko označili z:

¹² Supervenienca – mentalno je odvisno od fizičnega ali določeno z njim tako, da se dve stvari, ki se ne razlikujeta v nobeni fizični lastnosti, ne razlikujeta niti v nobeni mentalni. (Kim, 1996: 10) Namesto besedne zveze biti odvisen od fizičnega ali biti določen z njim se uporablja besedna zveza supervenira nad fizičnim, načelo pa se imenuje supervenienca duha nad telesom. (Bregant, 2004: 48)

1. emergentna lastnost je drugačne vrste, kot ta, ki jo imajo njeni sestavni deli;
2. je oblike, katere narava in obstoj nista *nujno potrebna* pri določanju značilnosti njenih sestavnih delov.

Problem te oblike ni njeno teptanje doktrine mikrodeterminizma (ali mereološke supervenience). Na ta način se supervenienca in emergenca stapljata tako, da supervenienca postaja komponenta pojavnosti. (Kim, 2006)

EMERGENCA IN REDUKCIJA

Očitno je, da obstajajo določena notranja trenja pri mnogih avtorjih, ki se ne morejo enoznačno odločiti med redukcijo in pojavnostjo. Skratka, poskušali bomo predstaviti te notranje dileme, kako so nastale, in ob tem izpostaviti nekatere možnosti razvoja obeh področij v prihodnosti.

Oboji, tako zagovorniki pojavnosti kot redukcionisti se strinjajo, da je »vedenje kot celota« določena z lastnostmi (izrazom to je) in s strukturnimi povezavami, ki jih karakterizirajo komponente te strukture. Kjer so njihova stališča različna, po Broadu (1925) upoštevamo domnevo, logično konceptualno dedukcijo, da lastnostih celote lahko izpeljemo iz dejstev njihovih osnovnih ravni, kar redukcionizem zagovarja in pojavnost zanika. Ta možnost izpeljave je v tem kontekstu ključna. Vprašanje: »Ali je pojavno lastnost *M* *možno* razložiti iz osnovnih dejstev *F*? Ali je dejstvo, da je *M* neposredno moč izpeljati iz *F*?« Zaradi tega lahko ugotavljamo, da Nagelovski model redukcionizma¹³ (Nagel, 1961, Bregant, 2004), ki je prevladoval pri takšnih in podobnih vprašanjih o redukciji, v zadnjih desetletjih nepomemben za reduktivno napovedovanje in njeno razlago v predstavljenem kontekstu. Za emergantiste je resnično razlagalno vprašanje le, zakaj veljajo ti zakoni premostitve.

Obstaja pa še drugačen model redukcije – redukcija s pomočjo funkcionalizacije ali na kratko funkcionalistična redukcija, ki bolje ustreza temu namenu. (Kim, 1999, 2005, Bregant, 2004, 2010) Premostitveni zakoni so srce Nagelovske redukcije: v resnici lahko pokažemo, da je uporaba premostitvenih zakonov potrebna in zadostna le za Nagelovsko redukcijo. (Kim, 2006) Po Kimovem mnenju moramo

¹³ Nagelovska redukcija – Nagel redukcijo definira kot »pojasnitev ene teorije (ali sistema eksperimentalnih zakonov), ki velja na enem področju raziskovanja, z drugo teorijo, ki velja na istem ali na drugem področju raziskovanja.« (Nagel, 1982: 338)

zato, da bi naša redukcija resnično poenostavila ontološko sliko sveta, premostitvene zakone $M \leftrightarrow F$ spremeniti v identiteto $M = F$, to pa bi med drugim podalo tudi odgovor na vprašanje, kako ali zakaj sta med seboj povezana prav M in F .

V nasprotju s tem funkcionalističnim procesom redukcije s pomočjo »funkcionalizacije« lastnosti reduciramo v obliko lastnosti na osnovnem nivoju (Kim, 2005). Tako funkcionalistični redukcionizem daje poenoteno sodbo ideje pojavnosti, da se pojavnih lastnosti ne da reducirati na osnovni fenomen in tudi ne razložiti ali jih napovedovati na njihovi osnovi. Če to drži, potem funkcionalistični redukcionizem tudi dosega ontološko redukcijo.

Tako sta supervenienca in funkcionalna nereducibilnost dva potrebna pogoja za pojavnost. Ali pa sta skupaj tudi zadostna? Mnogi avtorji kot Aristotel, pa McGinn in Van Gulick verjamejo, da sta zadostna v konceptu, ki so ga zasnovali klasični emergentisti, kot so Alexander, Morgan in Broad. To pomeni, da je bil zapis pojavnosti britanskih emergentistov podkarakteriziran. Če upoštevamo zadnje predloge v povezavi s pojavnimi lastnostmi v kompleksnih sistemih, kot so ideje o kaosu, nelinearni dinamiki, potrebnosti simulacij (bolj kot preračunov) in tako naprej, jih lahko raziskujemo zato, da bi ugotovili, ali ustrezajo klasičnim konceptom pojavnosti, zajetih v teh dveh pogojih (supervenienca in funkcionalna nereducibilnost). Seveda sodba o tem, ali kateri od teh novih predlogov ne ustreza klasičnim konceptom, sama po sebi ne kaže, da ni zanimiv obetaven koncept. Vendar združitev supervenience in funkcionalne nereducibilnosti lahko služi kot uspešno merilo za presojo; vsaka izpeljava iz tega je odstop od klasičnega koncepta emergence. Nove predloge lahko analiziramo in primerjamo med seboj glede na to, kako močno in na kakšne načine odstopajo od klasičnega koncepta.

NOMOLOŠKA EMERGENCA

Ena od možnosti, kako obiti ta razkorak med razlagalno prakso dinamične kognitivne znanosti in intuicijo, je rekonstruirati definicijo emergence. V tem poglavju si bomo ogledali eno od možnosti takšne rekonstrukcije, ki bi lahko razložila, kako razumeti emergenco kot zvezo med zakoni samimi. Tako tudi če sklepamo, da je kognitivno vedenje samo izpeljano iz kognitivnih zakonov, ostaja še vedno odprta možnost, da teh kognitivnih zakonov ne moremo deduktivno izpeljati

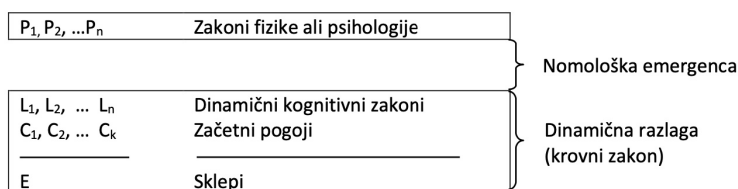
iz fizikalnih zakonov. Tako imamo še vedno živ koncept emergentizma (kakor nededuktivnosti).

To idejo je predlagal van Gelder (1991). Ko je obravnaval povezavo med enačbami dinamične kognitivne znanosti in ustreznimi zakoni fizike, je zapisal:

»V principu lahko predpostavimo, da so enačbe ultimativno izpeljane iz splošnih fizikalnih zakonov in možno je celo izvesti takšno izpeljavo. Vendar je pri tem bistveno to, da tako dolgo, dokler imamo enačbe sistema tako izpeljane, to nima praktične vrednosti za aktualno razlago vedenja sistema samega, in tudi če takšnih razlag nimamo, še vedno ne zgubimo nobenih razlagalnih sil.« (Van Gelder, 1991)

Van Gelder hoče s tem poudariti, da je dinamična razlaga še vedno dobra razlaga, tudi če dinamičnih kognitivnih zakonov ni mogoče izpeljati iz nižjerednih zakonov fizike. Tako tudi če potrebuje dinamična razlaga dedukcijo na nekem kognitivnem nivoju, dovoljuje nesklepanje med nivoji; še vedno so v nekem pomenu območja za emergenco, kljub premislekom o opisu krovnih zakonov in izpeljave redukcije na njihovi osnovi.

Ta pogled, ki ga Walmsley imenuje »nomološka emergenca,« upošteva, da lahko zakoni dinamične kognitivne znanosti emergirajo oz. vzniknejo iz ustreznih zakonov fizike in psihologije. Če se nomološka emergenca uporabi na kognitivnih sistemih, potem lahko neko kognitivno vedenje izpeljemo ali napovemo na osnovi višjerednih zakonov, vendar ti višjeredni zakoni sami ne morejo biti izpeljani ali napovedovani iz izjav (npr. nevropsihologije ali fizike). Še vedno bodo v nekem pogledu emergentno kompatibilni z dinamično kognitivno znanostjo, le da se emergenca ne bo dogodila, kot bi se, preden se ne bo pričela dinamična razlaga. Nomološko pojavnost bi lahko prikazali shematsko na sliki 2.2, kjer je podana razlaga modela krovnega zakona. (Walmsley, 2010)



Slika 2.2: Grafična predstavitev nomološke pojavnosti

Izberite svoj najljubši primer iz dinamične kognitivne znanosti. Zakon, ki ga boste v svojem primeru našli, bo napisan v obliki diferencialne enačbe, ki ustreza zapisu $L_1, L_2, \dots L_n$ na sliki 2.2, in emergenca bi bila potrebna zato, da dobimo iz zakonov fizike in psihologije, ki upravljajo obravnavane mehanizme, dinamične zakone, ki upravljajo kognitivna vedenja. Na primer v primeru Kelsovega modela opazovano ritmično gibanje prstov še vedno lahko razložimo z enačbami, vendar lahko rečemo, da so enačbe same nomološko pojavne s stališča zakonov fizike ali nevropsihologije, torej da vzniknejo iz fizikalnih zakonov.

Ena od prednosti takšnega prikazovanja emergence je, da ostaja nevtralna do številnih metafizičnih vprašanj, ki predstavljajo problem ostalim sodbam. Na primer nomološka emergenca je kompatibilna z obema, materialističnim in procesnim metafizičnim stališčem. Vse to pomeni za ta pogled na emergenco to, da o zakonih v emergentni zvezi ni treba sklepati, kot da izhajajo eden iz drugega. Ni pomembno, ali se zakoni razprostirajo preko materije ali procesa, nomološka emergenca je koncept emergence, ki je lahko primeren za oba metafizična pogleda.

Kot je bilo že omenjeno, je koncept nomološke pojavnosti *kompatibilen* mnogim drugim razlagam dinamičnih kognitivnih znanstvenikov. Kot tak predstavlja eno od možnih poti ohranjanja šibke verzije intuicij dinamike in pojavnosti, s katero je povezana ali z njo kompatibilna. Ni pa razvidno, ali lahko ta koncept pojavnosti podpira močnejšo intuicijo, da lahko na ta način dinamična kognitivna znanost *razloži* emergentizem o duhu.

Sliki 2.2 in 2.7 sta posebej uporabni pri razlagi tega problema. V predstavljenem kontekstu bomo nomološko pojavnost obravnavali prvenstveno le v povezavi z dinamično razlago. Dinamična kognitivna znanost morda ne more razložiti oz. osvetliti nomološke pojavnosti o duhu – ne more nam povedati, zakaj je tako, da ne moremo v osnovi sklepati o zakonih kognicije na osnovi zakonov fizike oz. psihologije – zato ker pač deduktivne razlage kot take v principu niso del kognitivne znanosti.

Ostaja vprašanje, ali je lahko ta oblika nomološke emergence tista, ki bi si jo dinamicisti želeli. Predpostavljamo lahko, da mnogi ne. Ta koncept pojavnosti ima verjetno vrsto temeljnih omejitev o tem, kako lahko razumemo svet, verjetno s

posledicami, ki jemljejo pogum vsem nadaljnjim potencialnim interdisciplinarnim znanstvenim raziskavam. To trditev je zapisal tudi C. D. Broad:

»Ob teoriji emergence se moramo sprijazniti z veliko manj enotnosti v zunanjem svetu in veliko manj zaupljivih povezav med različnimi znanstveniki.« (Broad, 1925: 77)

Nadalje lahko navedemo še »nerazložljivo golo dejstvo,« ki spominja na znamenito frazo Samuela Alexandra:

»Obstoj emergentnih lastnosti v takšni meri, kot je opisano, je nekaj, kar je znano pod pritiskom surovih empiričnih dejstev, ali, kot jaz raje rečem z manj ostrimi izrazi, sprejeto z »naravno pieteto« raziskovalcev. To ne dopušča nobene razlage.« (Alexander, 1920: 46)

Če to povzamemo, lahko ugotovimo, da v tem trenutku ni opaziti, da bi predstavljena pojavnost (emergenca) prinašala kako revolucionarno prednost pred t. i. »misterianizmom,« ki ga je razvijal npr. Colin McGinn (1999). Njegova izhodiščna trditev je, da je problem duh–telo uganka, ki načelno ni in ne more biti rešljiva. McGinnovi kritiki lahko kot protiargument njegovi teoriji dobro uporabijo kar njegov zamegljen pogled na mysticizem, obravnavan v njegovi nomološki emergenci. Iz tega zaključijo, da vse tiste, ki izjavljajo, da se je treba dinamični pojavi intuiciji izogniti, lahko zavrnejo kar z njihovimi lastnimi argumenti.

FUNKCIONALISTIČNA REDUKCIJA

Mnogi dinamični kognitivni znanstveniki bi se želeli izogniti temeljni neslogi, vendar je enotnost med različnimi poddisciplinami kognitivne znanosti ali med različnimi oblikami debat v mnogih primerih pomanjkljiva, kar seveda željo po slogi dela skoraj nemogočo. Če citiramo samo Randall Beera:

» /.../ dinamična kognitivna znanost ima cilj, da ustvari 'skupen jezik za kognicijo,' za nevropsihološke procese, ki jo podpirajo, za nekognitivna človeška vedenja in za adaptivna vedenja enostavnih živali. /.../ Kot taka je 'poenoten teoretični okvir za kognitivno znanost.'« (Beer, 2000: 97)

Vendar ima ta poskus iskanja soglasja pravzaprav veliko skupnega z Nagelovsko redukcijo. Še posebej poskus poenotenja različnih nivojev (v našem primeru kognitivnega in nevralnega) z njihovo karakterizacijo ob uporabi poenotnega

besednjaka, se lahko spreobrne v to, kar je Michael Silberstern imenoval »semantični poskus« redukcije. S tega stališča bi lahko:

»/.../ redukcijo razumeli kot neke vrste 'izomorfizem'¹⁴ ali 'ekspresivno ekvivalenco' med modeli.« (Silberstern, 2002: 88)

Ideja poenotenja, kot vse kaže, leži izven dosega Kelsovih dinamičnih študij. Njegov pogled je, da lahko *enako* vrsto teoretičnih orodij uporabimo za razlago tako nevralne kot tudi kognitivne dinamike; enake enačbe, ki karakterizirajo vzorce sposobnosti gibanja prstov, bodo prav tako lahko okarakterizirale nevralne osnove te sposobnosti. V epilogu svoje knjige je leta 1995 zapisal:

»Moj cilj je združiti na eni strani nevralne procese in mentalne oz. kognitivne procese na drugi strani v nek skupni jezik. To je jezik dinamičnih vzorcev, (ki omogoča) navzkrižne povezave med nivoji nevralnih in kognitivnih funkcij.« (Kelso, 1995: 289)

Kelsova ideja je zelo podobna »poenoteni znanosti,« ki so jo zagovarjali redukcionisti v prvi polovici 20. stoletja. Šel je še celo naprej in besednjak teorije dinamičnih sistemov (TDS) povezal s tradicionalnim besednjakom s področja duh – telo:

» /.../ razvrstitveni parameter izomorfni povezav med duhom in telesom, voljo in možgani, mentalnimi in nevralnimi dogodki. Duh sam je prostorsko časoven vzorec, ki povezuje metastabilne dinamične vzorce možganov. Dualizem duha in telesa je tako zamenjan s posamičnim izomorfizmom, katerega srž so semantični pomenski parametri vzorca.« (Kelso, 1995: 288–289)

To lahko osmislimo, če uporabimo ustrezno terminologijo, ki jo uporablja TDS sama. Razložimo razliko, ki se pogosto opisuje med »realnimi« oz. »konkretnimi« dinamičnimi sistemi, kot so npr. solarni sistemi ali Wattov centrifugalni regulator in »matematičnimi«, tudi »abstraktnimi« dinamičnimi sistemi, kot so sistem enačb, ki jih uporabimo za opisovanje realnih sistemov. (Giunti, 1997) Zanimive značilnosti teh razlik so, da dovoljujejo, da lahko več kot en realni sistem opišemo z istim matematičnim dinamičnim sistemom. *Iste* enačbe lahko tako opisujejo vrsto *različnih* konkretnih sistemov.

¹⁴ V abstraktni algebri je izomorfizem (iz grščine ἴσος (isos) enak in μορφή (morfe) oblika) oblika preslikav med objekti, ki izkazujejo zveze med dvema lastnostnima ali strukturama. Kadar med dvema strukturama obstoja izomorfizem, pravimo, da sta strukturi izomorfni. V določenem smislu sta skratka strukturi identični, če zanemarimo manj signifikantne razlike.

Tako lahko Kelsovo obliko redukcionizma razumemo kot zahtevo, da je to, kar združuje duh in možgane, dejstvo, da sta oba opisana z diferencialnimi enačbami enake oblike. Ta zahteva ni povezana toliko s tem, da bi morali kognicijo reducirati na nevropsihologijo, kot s tem, da bi obe lahko »reducirali« na enako porazdeljeno osnovno dinamiko.

Kljub vsem predstavljenim argumentom obstaja resen ugovor, povezan predvsem s poskusom poenotenja dinamične razlage enakih fenomenov na različnih nivojih – še posebej na vedenjskem in nevralnem nivoju. Ta ugovor sta Thelen in Smith (1994) zapisali takole:

»Mi smo še posebej poudarili, da sta vedenje in razvoj dinamična na mnogo nivojih razlaganja, še posebej, da so fenomeni, opisani na nivoju vedenja skladni s tem, kar je znanega o možganih in njihovem delovanju.« (Thelen, Smith, 1994: 14)

Thelen in Smith sta v povezavi s tem poudarili tudi analogijo z zvezami med kvantno mehaniko in klasično fiziko:

»Kvantna mehanika *ne razlaga* delovanja objektov na makronivoju. Vendar pa kvantna mehanika razlaga prehajanja in spremembe v objektih. Interakcije in dinamika kvanta pa kljub temu posredno razlagata, kako se objekt *spreminja* na makronivoju. /.../ Moč razlage je v dinamiki procesa, v pogledu na raziskovanje od spodaj navzgor. Razlagalna moč je v skupnem ocenjevanju mikro in makro nivoja. /... /To ni tradicionalni redukcionizem.« (Thelen, Smith, 1994: 39)

Očitno je, da je zgornja razlaga redukcije v celoti v nasprotju s »tradicionalnim (nomološkim) redukcionizmom« nagelovskega tipa. Poskusimo to trditev nekoliko utemeljiti.

Nagel je redukcijo definiral kot pojasnitev ene teorije, ki velja na enem področju raziskovanja, z drugo teorijo, ki velja na istem ali na drugem področju raziskovanja. (Nagel, 1982: 338) Skratka, pri pojasnjevanju neke teorije, ki velja npr. v biologiji, z neko teorijo, ki velja v kemiji, bi bili področji raziskovanja obeh teorij sicer različni, pri pojasnjevanju neke teorije, ki velja v fiziki, z neko drugo teorijo, ki prav tako velja v fiziki, pa bi bili področji raziskovanja obeh teorij isti (Bregant, 2004: 59). Prvo teorijo, s katero pojasnjujemo drugo, imenujemo primarna, širša teorija, drugo

teorijo, ki jo pojasnjujemo s prvo, pa sekundarna, ožja teorija. Nagel je postavil tudi dva formalna pogoja za redukcijo:

1. kadar sekundarna teorija T_2 vsebuje določene izraze, ki jih primarna teorija T_1 ne vsebuje, govorimo o heterogeni redukciji, in da bi prišli do heterogene redukcije T_2 na T_1 , moramo izraze iz sekundarne teorije prevesti v jezik primarne. To imenujemo *pogoj povežljivosti*;
2. ko so izrazi, ki jih vsebuje samo sekundarna teorija, zamenjani z izrazi iz primarne teorije, s katerimi so povezani (ko je izpolnjen povezovalni pogoj), lahko sekundarno teorijo logično izpeljemo iz primarne teorije. To imenujemo *pogoj izpeljivosti*. (Nagel, 1982: 346-347)

Takšen model obravnava redukcijo kot relacijo med dvema vrstama znanstvenih teorij, pri čemer se teorija razume kot skupek izjav, ki predstavljajo njene zakone. Tako lahko sekundarno teorijo reduciramo na primarno samo, če lahko zakone, ki nastopajo v sekundarni, reduciramo na zakone, ki nastopajo v primarni. Iz tega izhaja tudi ime, ki ga takšna redukcija ima, namreč *nomološka redukcija*. (Bregant, 2004: 60)

»V nasprotju s tem mora *funkcionalistična redukcija*, če želi poenostaviti ontološko sliko sveta, premostitvene zakone, kot je $M \leftrightarrow N_b$, spremeniti v identiteto $M = N_b$, to pa bi med drugim dalo tudi odgovor na vprašanje, kako ali zakaj sta med seboj povezana ravno M in N_b . Vprašanje je, ali to lahko storimo.« (Bregant, 2004: 210)

Prvi korak do te formulacije M -ja kot funkcionalne lastnosti je v pojmih njegovih vzročno-nomoloških relacij do drugih lastnosti, kar je v popolnem nasprotju z Nagelovimi osnovnimi definicijami redukcije. Iz tega izhaja tudi zgoraj omenjena razlika med obema redukcijama.

Kljub temu lahko tudi trdimo, da če obstaja možnost, da razlagamo spremembe na makronivojih (primarnih) z mehanizmi mikronivojev (sekundarnih), potem zagotovo moramo imeti *neke vrste* redukcionizem. V primeru E. Thelen in L. B. Smith pa imamo okoliščino, pri kateri bolj, kot da bi makronivoje reducirali na mikronivoje, oba nivoja razlag poenotimo z njunimi skupnimi poudarki časovnih sprememb in matematičnih formalizmov, ki te razlage opisujejo.

Ponovno se zastavi vprašanje, kako razumeti to novo obliko »nenagelovskega« redukcionizma. (Walmsley, 2010) Za detajlno simulacijo kortikalnega modula¹⁵ je potrebnih 10.000 nevronov. Za simulacijo možganov podgane bi model z enako natančnostjo moral imeti 100 milijonov nevronov, kar bi zahtevalo 20.000 krat več spomina. Simulacija človeških možganov bi zahtevala dodatno 1000-kratno povečanje tako spomina kot tudi računalniške moči. Menimo, da bi lahko našli podobnost med to razlago in modelom »funkcionalističnega redukcionizma«, ki ga je predlagal Kim (1997, 1998). Kimov model je eksplicitno zasnovan kot alternativni model Nagelovskega redukcionizma. (Kim, 2005, Markič, Bregant, 2007)

Če povzamemo, Kimov funkcionalistični redukcionizem je osredotočen na iskanje odgovora na vprašanje:

»Kakšen fizikalizem in koliko tega fizikalizma si lahko privoščimo?« (Bregant, 2009: 1)

Iskanje rešitev je sestavljeno iz treh korakov: *najprej* je treba formulirati višje nivojske lastnosti, ki jih je treba reducirati v izraze njihovih vzročnih vlog, *potem* je treba te lastnosti opazovati na elementarnem nivoju izvajanja vzročnih nalog, t. i. realizatorjev, in *nazadnje* je treba ustvariti teorijo, ki bo razložila, kako to narediti.

Eden od primerov, ki jih je Kim podal, je primer genov. Namesto, da bi gene obravnavali v odvisnosti od njihovih notranjih, resničnih lastnosti, Kim dokazuje, da moramo specifikirati gene vzročno-posledične povezave, da lahko rečemo npr. nekaj podobnega kot: »/.../ gen je tisti mehanizem v bioloških organizmih, ki je vzročno odgovoren za prenos dednih karakteristik od staršev na potomce.« (Kim, 1998: 25) Oblike enačb, s katerimi zapišemo standardne dinamične modele v kognitivni znanosti, bi lahko obravnavali podobno, kot določajo vzročno-posledične povezave med pomembnimi spremenljivkami v kognitivnem sistemu.

Drugi del, ki ga moramo sprejeti kot fizikalisti, je povezan z zavestjo, z vprašanjem kako lahko stvar, kot je zavest, obstaja v takšnem fizičnem svetu, povezan z iskanjem lastnosti ali mehanizma, mnogokrat na mikronivoju, ki bi zadovoljil te vzročno-posledične podrobne opise in s tem zadovoljil predpisane vzročne zahteve. Vzemimo primer gena in DNK.

¹⁵ Kortikalni modul, poimenovan tudi kortikalni steber ali hiper steber, je skupina nevronov v možganski skorji.

»Gen je tisti biološki dejavnik ali mehanizem, ki skladno z obstoječimi zakoni opravi vzročno v smislu prenosa notranjih in zunanjih lastnosti z generacije na generacijo, molekula DNK pa tisto, ki to vzročno vlogo v resnici odigra, t. j. tisto, kar vzročno delo gena opravi v našem svetu in v vseh ostalih, ki so mu podobni.« (Bregant, 2004: 211)

Skratka molekula DNK zadovolji vzročno zahtevo po odgovornosti za prenos dednih karakteristik od staršev na potomce. V primeru dinamične kognitivne znanosti, bi ta oblika funkcionalističnega redukcionizma zahtevala, da najdemo možganski sistem, ki se pokori enakim enačbam, kot tisti, ki ga usmerjajo višja nivojska vedenja (in to bi nam dovoljevalo, da upoštevamo Kelsovo in Silbermanovo uporabo termina *izomorfizem* dobesedno. Enačbe, ki označujejo oboje, kognitivno in nevrnalno dinamiko, bi imele enako obliko). Če bi to bilo mogoče, bi lahko rekli, da se da dinamika duha reducirati na dinamiko možganov, s čimer se vsi ne strinjajo.

Oblika tristopenjskega procesa funkcionalnega redukcionizma bi bila to, na kar so Kelso, Thelen in Smith namigovali. Skratka, če bi lahko našli diferencialne enačbe, ki bi vodile višje nivojske vedenjske in nižje nivojske nevrnalne vidike pri zadani nalogi, bi lahko ugotovili, da sta si (dobesedno) izomorfni. Tako bi primarna naloga dinamičnih kognitivnih znanstvenikov morala biti, da poiščejo enake fenomene na različnih nivojih razlag zato, da bi uspešno razložili to vrsto združitve. Ni potrebno posebej poudariti, da je tovrstni material, o katerem sanjajo kognitivni znanstveniki, že bil narejen. Poudarimo lahko, da če usvojimo perspektivo funkcionalističnega redukcionizma, lahko uspešno opustimo obe verziji emergentne intuicije. To je razlog, da naj funkcionalistični redukcionizem ne bi bil všečna opcija za tiste, ki želijo nereduktivno obravnavati duh in kognicijo – in to kljub dejstvu, da ga tudi nekateri dinamični kognitivni znanstveniki eksplicitno predlagajo.

V nadaljevanju poglavja se bomo osredotočili na izhodiščno problematiko kognitivnega modeliranja, na inteligenco in zavest.

2.3 Zgodovinski oris modeliranja v kognitivni znanosti

V uvodu je bilo prikazano (Bechtel, Abrahamsen, 2002), da ima konekcionizem svoje korenine v štiridesetih letih dvajsetega stoletja, ko so sinergijsko povezali nevronske sisteme, preračunavanja in inteligenco, ter kratko razvojno pot od enostavnih, formalnih nevronov in perceptronov. To pot so v šestdesetih letih

dvajsetega stoletja zasenčili simbolni modeli umetne inteligence. S krepitvijo konekcionizma v osemdesetih letih dvajsetega stoletja pa nevronske mreže ponovno postanejo obetavne. V naslednjih dveh desetletjih dvajsetega stoletja postaneta oba modela dominantna na področju kognitivne znanosti.

V devetdesetih letih dvajsetega stoletja se je zgodila nekakšna kontrarevolucija kognitivni revoluciji, ki ju predvodita teorija dinamičnih sistemov (angl. DST – dynamical system theory) (van Gelder, 1991) in »utelešena kognicija« (angl. embedded/embodyed cognition) (Clark, 1997, Bechtel, Abrahamsen, 2002). Prične se govoriti o *dinamičnih sistemih*, katerih temelje je postavil Newton s svojimi tremi zakoni gibanja in razvojem metod izračuna gibanja nebesnih teles, napovedovanja posebnih planetarnih konfiguracij. Poincarejevi geometrijski približki dinamike pa so v poznih letih 19. stoletja postali osnova za teorijo dinamičnih sistemov dvajsetega stoletja.

Ideja, da je kognicija dinamična in bi jo najbolje razumeli z uporabo orodij DST se pojavi v devetdesetih letih dvajsetega stoletja, vendar je kljub temu še vedno nekako na obrobju kognitivne znanosti. Robert Pott in Timoty van Gelder želita prepričati strokovno javnost, da ima DST revolucionaren vpliv na kognitivno znanost. Pri teh svojih prizadevanjih poudarjata predvsem:

- da sta dinamični in računski sistem v osnovi popolnoma različni *obliki* sistemov, zato sta modela kognicije na osnovi obeh oblik do najglobljih temeljev drugačna;
- da kognitivni sistem ni računalniški sistem (statični, formalno definirani in zapisani sistem), temveč je dinamični sistem. Dinamični sistemi niso le možgani, temveč tudi celoten živčni sistem, telo in okolica, ki vpliva na človeka, in
- nenazadnje, kognitivni sistem ni *diskreten*¹⁶ sistem temveč je *zvezčen*.¹⁷

¹⁶ Diskretni sistemi so sistemi s končnim številom stanj, zaradi česar ga lahko natančno opišemo z ustreznim matematičnim modelom. Mnogokrat diskretne sisteme poimenujemo tudi digitalni sistemi. Kot predstavnik diskretnih sistemov po navadi navajamo digitalni računalnik (kjer vse procese opisujemo v binarni obliki).

¹⁷ Zvezni sistemi, poimenujemo jih tudi analogni sistemi, so sistemi z neskončnim številom stanj vrednosti neke spremenljivke. Predstavnik zveznih sistemov je analogni računalnik.

Tako je Bechtel¹⁸ (Bechtel, Abrahamsen, 2002) predstavil osnove dinamičnih konceptov in orodja, uporabo enostavnih konceptov na primerih raziskav mrež, kako uporabiti model kaosa pri mrežnih modelih, filozofske razprave in argumentiranja Van Gelderjevih in Portovih trditev ter bolj kontroverzne filozofske diskusije Terenca Horgana in Johna Tienzona (1996).

2.4 Diskretni sistemi nasproti zveznim

Ena od glavnih intuitivnih prednosti klasičnih računalniških pristopov h kogniciji je bil na videz jasen tradicionalni zapis kognitivnih prikazovanj. Odkar razumemo kognicijo kot funkcioniranje nečesa, npr. *LISP* programa,¹⁹ je predstava konstruirana iz delov, ki sestavljajo LISP atome v celoten izraz. Predstava je natančna podatkovna struktura, ki ima semantični kontekst (z upoštevanjem znotraj ali zunaj kognitivnega sistema). Lahko jo spreminjamo in prilagajamo po trenutni potrebi. Takšni simboli imajo nesporne podobnosti v dejanskem svetu in frazah naravnega jezika. (Fodor, 1975)

Razvoj *konekcionističnih modelov* v osemdesetih letih, mrež, sestavljenih iz med seboj povezanih vozlišč, je omogočil prvo alternativo predstavitev kot stalnih in povezanih manipulativnih znakov. V konekcionističnih modelih so izidi procesa identifikacije (recimo, prepoznavanje pisave ali potez človeškega obraza) le začasni vzorec aktivnosti skozi poseben niz vozlišč in ne nekaj, kar je podobno medsebojno neodvisnemu objektu. Možnost predstavitve v tej bolj prožni obliki nas privede do zapisa *distribuiranih predstav*, kjer ni možno najti nobenega navideznega »objekta«, ki bi ustvaril predstavo, ampak imamo le partikularno distribuirane vzorce preko enakega niza vozlišč, ki se uporabljajo tudi za mnoge druge vzorce. Konekcionisti poudarjajo, da takšna predstavitev ne deluje kot dober primer za *simbol* kot konvencionalno predstavo računalniški tradiciji, vendar še vedno lahko deluje kot predstavitev v mnogih podobnih primerih.

¹⁸ Bolj podrobno bomo o tem govorili v poglavju 7.

¹⁹ LISP je družina računalniških višjerednih programskih jezikov, ki je izvorno narejen kot praktičen matematičen zapis računalniških programov, predvsem na področju umetne inteligence. Ime programskega jezika izhaja iz angleških besed - List Processing.

Razvoj dinamičnih modelov zaznav in motoričnih nalog nas vodi k nadaljnjim razširitvam teh zapisov predstavitvene funkcije, ki vsebuje časovno spremenljive trajektorije, omejene cikle, povezane limitne cikle in atraktorje, h katerim lahko stanje sistema teži. Z dinamičnega gledišča statične računalniške predstave bodo imele zelo omejeno vlogo v kogniciji.²⁰ (Port, 2002)

2.4.1 Emergenca in redukcija v dinamični kognitivni znanosti

Čeprav segajo korenine obravnavanja dinamike v kognitivni znanosti v čas kibernetike²¹ in ekološke psihologije,²² je ta pristop doživel razcvet šele v sredini devetdesetih let 20. stoletja, predvsem zaradi nekaj zelo odmevnih in pomembnih objav s področja filozofije, psihologije, nevroznanosti in umetne inteligence. Kljub tem začetnim evforičnim poskusom do danes niso natančno definirana ne področja dela in tudi razvoj dinamične kognitivne znanosti ne. Kot je omenjeno v projektu Modri možgani, se predvideva intenziven razvoj na tem področju vse tja do leta 2020, verjetno pa še mnogo dlje v prihodnost. Čeprav so bili mnogi elementi dinamičnih modelov (npr. pomen časa, pomen časa v povezavi s telesom in okoljem) v preteklosti že podrobno obdelani, kljub vsemu še niso zgrajeni trdni temelji dinamične kognitivne znanosti in njena *filozofska* izhodišča in posledice. V nadaljevanju si bomo zato nekoliko podrobneje ogledali, kako lahko dinamični modeli v kognitivni znanosti osvetlijo problem duh–telo. (Walmsley, 2008, 2010)

Že Newton je pokazal, da enaki zakoni narave upravljajo oboje, zemljo in božja gibanja, pa vseeno ni odkril zakonov misli. Skoraj 300 let po njegovi smrti je *dinamika*, področje matematike, ki ga je populariziral prav on, postala ena od popularnih človekovih načinov raziskovanja samega sebe. V zadnjih 15 letih ugotavljamo povečan interes po vrsti konceptov, ki jih poimenujemo s skupnim imenom »dinamični modeli kognicije«.

²⁰ Bolj podrobno se bomo tem problemom posvetili v poglavjih 6 in 7.

²¹ Kibernetika je interdisciplinarna študija struktur reguliranih struktur. Kibernetika je ozko povezana s teorijo kontrole in sistemsko teorijo, vsaj z njeno obliko prvega reda. Kibernetika se uporablja tako v fiziki kot tudi v družbenih sistemih, temelječih na jeziku. Beseda kibernetika je bila prvič uporabljena v kontekstu "študija samonadzora" v Platonovem delu Zakoni, da bi označili nadzor nad ljudmi. Termin, besedo 'cybernétique,' je uporabil leta 1834 fizik André-Marie Ampère, da je z njo označeval znanost nadzora v njegovem sistemu klasifikacije človeških znanj. Sodobna kibernetika postane interdisciplinaren študij, ki povezuje področja kontrole sistemov, npr. v teoriji električnih omrežij, strojništvu, modeliranju logike, razvojni biologiji in nevroznanosti v 40. letih. Omenimo le še kibernetično didaktiko, o kateri bomo povedali nekoliko več v nadaljevanju.

²² Ekološka psihologija je termin, ki so ga pričele uporabljati mnoge šole psihologije. Najpomembnejši avtorji so James J. Gibson Roger G. Barker Herb Wright in ostali z Univerze v Kansasu in Lawrence.

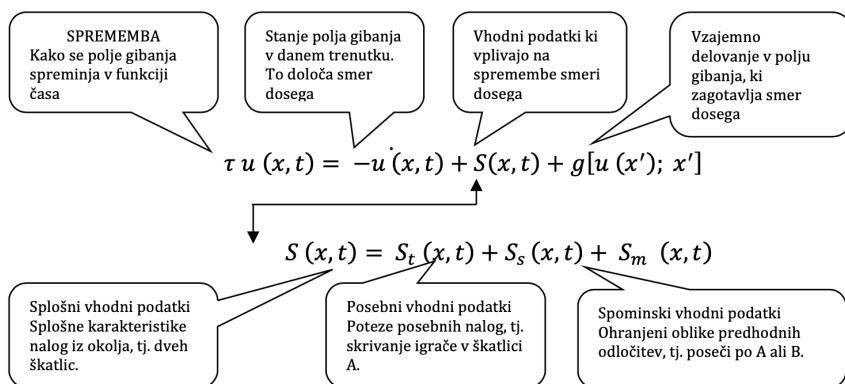
Dinamična kognitivna znanost v bistvu skuša pokazati, da duh vodijo enake vrste zakonov, kot so zakoni, ki vodijo gibanje planetov. Kot je zapisal Kelso (2003):

»Dinamika ne pronica samo v naravoslovne znanosti, temveč tudi v družboslovje, to je psihologijo in nevroznanost. Kognitivna znanost /.../ katere koncept in paradigme so zgodovinsko prisotne v jeziku digitalnih računalnikov, bi bile lahko zadnje od *Les Science Humaines*, ki bi padle pod urok dinamike.« (Kelso, 2003: 48)

V uvodni diskusiji se bomo osredotočili le na dobro znana primera dinamičnih sistemov v kognitivni znanosti, to sta model Kelsa (1995), ki je povezan z dinamiko telesa, in model Telen in Smith (1994), povezan z delovanjem duha in mentalnimi procesi. Oba bomo skušali le filozofsko interpretirati. Poglejmo na kratko oba.

MODEL A – NE – B

Prvi takšen, s filozofskega stališča zanimiv model, je model, ki rešuje problem, ki ga je obravnaval že Piaget (1970). Kot dinamični model pa sta ga obravnavali Linda B. Smith in Esther Thelen (Thelen, Smith, 1994, Smith, Thelen, 2003). Poimenovali sta ga kar na kratko: A – ne – B model. Izhodiščni Piagetov problem se glasi nekako takole: če dojenčku med 8. in 10. mesecem skriješ igračo (ali slaščico) pred njegovimi očmi pod eno od dveh škatlic, bo segel, da bi dobil igračo, po pravilni škatlici. Če pa, potem, ko skrijemo objekt nekajkrat pod eno od obeh škatlic (npr. škatlico A), zamenjamo skrivališče in objekt »skrijemo«, tako da bo to videl, pod drugo škatlico (škatlico B), bo mladostnik naredil »A – ne – B napako« in posegel po škatlici A, čeprav bo videl, da smo objekt skrili pod škatlico B. Piaget je izvirno trdil, da je napaka nekaj, kar je povezano z mladostnikovim razumevanjem stalnosti objekta. Seveda je tekom let nastalo mnogo drugih tolmačenj, kar pa ni predmet našega problema. Skratka, psihologinja Ester Thelen je skupaj s svojimi kolegi ugotovila, da so zaradi tega, ker se vse nastale hipoteze osredotočajo na mladostnikove notranje mehanizme, načine, kako razmišlja o svetu, takšni pogledi na napako *kognitivni*. Napaka predstavlja mentalno predstavo o objektu. Zato so predlagali, da bi se pri snovanju modela osredotočili le na razlago napake in ne tudi na mladostnikovo aktivnost *poseganja po objektu*, pri kateri so pomembni dejavniki gledanje, načrtovanje, poseganje in pomnjenje. Na osnovi tega so razvili dinamični model o tem, kako mladostnik pride do tega, da poseže v določeno smer s kvantificiranjem vplivnih dejavnikov ter njihovo povezavo v »veliki enačbi«, kar je prikazano na spodnji sliki.



Slika 2.3: Velika enačba avtoric Thelen in Smith.

Enačba je odvisna od vrste dejavnikov, vključno z obstoječim stanjem načrtovanega polja gibanja, splošnih in posebnih vidikov nalog (kot npr. prisotnost dveh škatlic in predmeta, skritega pod eno od njih), vplivni spomini predhodnih rešitev (kar je grobi ekvivalent navadam) in interakcija med lokacijami na načrtovanem polju gibov. Določa, kako se mladoletnik nagiba k temu, da bi dosegel določene spremembe smeri v času kot izid sprememb parametrov. Thelen et al. so pokazali, kako tukaj opisani dinamični sistem izvaja klasično A – ne – B napako, vključujoč mnoge od miselno povezanih domislic, ki jih predhodne hipoteze niso bile zmožne obvladovati. Nadalje model izkazuje nekaj pomembnih zaključkov, ki povezujejo naravo in potrebo po »predstavi« v duhu in s povezavami med dinamičnim in »nameščen« ali »utelešen« presojo o kogniciji, ki je s tem v zvezi. Z ozirom na to diskusijo je narava razlage, ki jo ponuja ta model, še posebej zanimiva.

RAZLAGA MODELA

Prikazani dinamični model Thelen in Smithove zelo dobro ustreza modelu krovnege zakona razlage Hempela in Oppenheima. (Hempel, 1965)

Začetni pogoji, ki morajo biti pred pričetkom znani, so:

- trenutno stanje v gibalnem polju,
- splošni in posebni »spominski« vnosi v sistem,

c. »funkcija«, s katero preračunavamo vnose.

Otrokovo vedenje (ki se izraža v spremembah teh parametrov) sledi kot matematična (tj. deduktivna) posledica te enačbe ob upoštevanju predpisanih začetnih pogojev.

V tem primeru ima razlaga partikularnega dogodka (poseganja po škatlici) potem, ko se dogodi, enako logično obliko, kot jo ima napoved dogodka v prihodnosti. Podobno lahko s to enačbo tudi napovedujemo, kaj se bo zgodilo v spremenjenih razmerah, ki bi jih želeli raziskati; skratka, model omogoča napovedovanja onstran znanih podatkov (napovedi, ki bodo morda potrjene v prihodnosti). Thelen in Smithova sta zapisali formalni karakter modela kot:

»Ponujamo formalni dinamični teoretični model, ki temeljni na kognitivnem utelešanju in ki simulira dobro znan $A - ne - B$ učinek, ob tem pa omogoča tudi nova predvidevanja in napovedi, ki bodo ustrezala novim eksperimentalnim izidom.« (Thelen et al., 2001: 1)

PRIMER

Jean Piaget je k razumevanju učenja prispeval *razvojno perspektivo*, ko je v seriji individualnih poskusov proučeval, po katerih mehanizmih otroci različnih starosti spoznavajo svet okoli sebe in kako se učijo. Razložimo v nadaljevanju ozadje dinamičnega problema na izhodiščnem konkretnem Piagetovem primeru, skratka kako uporabiti dinamični koncept mnogovzročnosti in s tem povezanega časa. (Smith, Thelen, 2003) Vprašanje, ki ga je zastavil že Piaget (1985), se je glasilo: »Kdaj mladostniki dosežejo *razumevanje stalnosti objekta*?« Kot smo opisali, si je izmislil enostavno nalogo skrivanja objekta (npr. igrače), ki jo je za njim uporabljalo več generacij raziskovalcev (to je postal tudi standardiziran test). 8 – 10 mesecev stari otroci delajo pri odkrivanju objekta resno »napako«, saj, če preteče nekaj časa od trenutka, ko so videli, kje je skrita igrača, do trenutka, ko so jo morali poiskati, niso posegli po dejanskem mestu, npr. po škatlici B, kjer so pred tem videli skrito igračo, temveč po škatlici A, kjer so igračo pred tem vedno našli. Ta $A - ne - B$ napaka je še posebej zanimiva zato, ker je tesno povezana s kasnejšimi razvojnimi stopnjami: otroci starejši od 12 mesecev te napake niso več počeli. Kaj pogojuje to dramatično spremembo? Ali 12 mesečni otroci vedo nekaj, česar 10 mesečni še ne vedo? Piaget je ugotovil, da otroci starejši od 12 mesecev vedo, da objekt (igrača) lahko obstaja

neodvisno od njihovih lastnih aktivnosti. Iz vseh teh raziskav je bilo ugotovljeno, da med tema dvema mesecema otroci spremenijo svojo predstavo o prostoru, spremeni se delovanje njihove *prefrontalne možganske skorje*, naučijo se zadržati svoje odzive in spremenijo svoje razumevanje naloge ali povečajo trdnost svoje predstave. To so izhodišča in izidi poskusa, vendar nihče zagotovo ne zna razložiti popolnega vzorca eksperimentalnih izidov. (Smith, 1999) Ponujena je formalna teorija, ki s pomočjo dinamičnega modela polja (Erlhagen, Schoener, 2002) razlaga, kako se je napaka A – ne – B pojavila kot produkt mnogovzročnosti v kontekstu časovne skale. Razlaga se začenja z analizo opazovanja (zaznavanja), načrtovanja, poseganja in pomnjenja, ki so ključnega pomena za izpolnitev zastavljene naloge. (Smith, Thelen, 2003)

S predstavljenim dinamičnim matematičnim modelom A – ne – B so simulirani samo kognitivni odzivi otrok in niso upoštevani vsi ostali robni pogoji, kot so npr. aktivnost *poseganja po objektu*, pri katerih so pomembni dejavniki gledanja, načrtovanja, poseganja in pomnjenja.

Naloga dinamike

Predstavljeni dinamični model simulira otrokovo odločitev poseči po škatlici A ali B in s časovno integracijo upošteva različne vplive na to odločitev. Prikazani model je sicer nevrolško navdihnjen, vendar je abstrakten, ni anatomsko določen, skratka, ne upošteva načina poseganja po objektu. Model upošteva enodimenzionalno polje aktivnosti z definiranjem parametra prostora potencialnih stanj poseganja (v prikazanem primeru le lokaciji cilja A ali B). Vhodni podatki so podani s položajem cilja in njegovim vplivom na polje.

Preden je otrok videl, kje je skrit objekt (npr. igrača), sta obstajali dve identični lokaciji A in B. Ko je eksperiment usmeril otrokovo pozornost na lokacijo A, tako da je otrok videl, da je tam skrita igrača, je s tem povzročil visoko prehodno aktivacijo (pozornost) na lokacijo A. Nato se je ta pozornost razvijala vse do trenutka, ko je presegla potreben prag, zaradi česar je otrok posegel po lokaciji A. Najpomembnejše pri tem je, da enkrat, ko je otrok že posegel, spomin na to postane vhodni podatek za naslednji poskus. Tako se pri naslednjem A poskusu pojavi povečana aktivnost (pričakovanje) po mestu A, prav zaradi predhodne izkušnje. Prav zato otrok ponovno poseže po A-ju. Pri večkratni ponovitvi tako pride do intenzivnih spominov na predhodne aktivnosti in vsaka naslednja aktivnost je še

močneje zaznamovana s predhodno. Sledi sprememba lokacije objekta iz A v B, kar prikazuje slika 2.6b. Kljub označeni lokaciji B (otrok vidi, da smo skrili igračo pod škatlico B), otrok po krajšem času, (zaradi predhodnih izkušenj) še vedno poseže po objektu A, naredi napako A – ne – B. Model napoveduje časovno pogojenost: pri poskusu obstaja zelo kratek čas, neposredno po tem, ko vidi, kam je predmet skrit, ko izbere pravo lokacijo (Wellman, et al., 1986), če pa se ta čas nekoliko podaljša, se napaka izvrši.

Ta nevrološki dinamični model je lahko splošno veljaven. Z njim lahko izvajamo različne podobne eksperimente in študiramo različne napake. To dosežemo s spreminjanjem zakasnitve, spreminjanjem dejanskih lastnosti skrivanja in s povečanjem ali z zmanjšanjem števila ponovitev poseganja po npr. lokaciji A. (Smith, 1999, Smith, Thelen, 2003) Skratka, s predstavljenim modelom je mogoče simulirati tako realni dogodek ali pa eksperiment. In ne samo to, ta nevrološki dinamični model, ki kreira opisano napako pri otrocih, je v svojem bistvu procesni model, ki upošteva ciljno usmerjene aktivnosti pri različno starih subjektih. To pomeni, da bi lahko s tem modelom simulirali različne naloge podobnega tipa za različne starostne skupine, kar je potrjeno tudi v Spencerjevih raziskavah. (Spencer, et al., 2001)

Implikacije modela

Teorija dinamičnih sistemov razvoja živih bitij pomaga reševati različna teoretična nasprotja. Na zelo elementarnem nivoju lahko trdimo, da npr. človekove omejitve zaradi biološke dediščine ter podobnost razvojnih okolij povzročijo ustrezne izide tega razvoja. Vsi zdravi otroci se naučijo hoditi, se razvijejo od stopnje ko delajo napako A – ne – B, pa do stopnje, ko je ne delajo več, naučijo se govoriti v maternem jeziku in si oblikujejo različne socialne povezave. Vendar, če bi podrobno raziskali ta razvoj, postanejo slike tega razvoja neprimerno manj deterministične, saj se npr. otroci iz iste družine (ista biološka dediščina in isto okolje) lahko razvijajo v popolnoma različnih smereh. Kompleksni sistemi določenih nivojev in z upoštevanjem določenih časovnih skal lahko imajo prav tako te lastnosti. Po eni strani se lahko samoorganizirajo (so avtonomni – sposobni so se učiti) zato, da ustvarijo kohezivne vzorce. Po drugi strani so lahko visoko nelinearni, včasih to poimenujemo tudi »občutljivi na začetne pogoje«. To pomeni, da že majhne spremembe ene ali več komponent dinamičnega sistema lahko vodijo do reorganizacije sistema in s tem do velikih sprememb v vedenju sistema samega.

Takšna nelinearnost se lahko kaže v razvoju sprememb, podobnih faznim premikom periodičnih funkcij (npr. funkcije $\sin$ ob faznem premiku za $\pi/2$), kar se lahko kaže v našem primeru tudi v bistveni razliki med 10 in 12 mesecev starimi otroci pri standardnemu testu A – ne – B. Vendar, če ta razvoj izhaja iz dogodkov v realnem času, potem ta nelinearnost lahko prav tako ustvarja tudi individualne razlike. Z dinamične perspektive je pomembno razumeti proces, s katerim vsakodnevne aktivnosti ustvarjajo razvoj otrokovih sprememb, tako splošne kot tudi posamične poti teh sprememb. In če razumemo ta proces, potem je možnost kreiranja dinamičnih matematičnih modelov, ki odražajo realne razmere, zelo velika, v nasprotnem primeru sicer lahko nek proces simuliramo, a nikoli ne moremo zaupati dobljenim izidom.

Pri človeškem razvoju vsak nevralni dogodek, vsaka reakcija, vsak nasmeh ali vsaka socialna interakcija ustvarja pogoje za naslednjo spremembo, in s tem vzročne sile v realnem času, ki sledijo tej spremembi. Če je tako, potem bomo prišli do globljega razumevanja tega razvoja le s študijo multivzročnosti, upoštevaje časovno zaporedje in odgovarjajočo samoorganiziranost.

MODEL HKB

Drugi model, ki ga bomo podrobneje predstavili v poglavju 7, je povezan z dinamiko teles. To je model ritmičnega gibanja prstov, ki je poimenovan po avtorjih Haken, Kelso in Bunz in ga na kratko označujemo tudi HKB model. (Haken et al., 1985)

Kar smo povedali za prvi model, bi lahko ponovili tudi za HKB model, ki določa, kako se relativne faze gibanja prstov v dvokoordinatnem prostoru kažejo v spremembah frekvence in oscilacije. Te spremembe lahko zadovoljivo opišemo z enačbo 2.1. Poznamo vrednosti ϕ ; v večini primerov upoštevamo le eno od dveh vrednosti, ali 0 ali $\pm 180^\circ$, ki ustrezata stabilnima stanjema, fazi ali antifazi gibanja. Nadalje, je razmerje a/b sorazmerno frekvenci oscilacij, ki jo prav tako poznamo iz eksperimentalnih raziskav. Dve stabilni stanji dobimo pri $d\phi/dt = 0$ (kjer je razmerje sprememb relativnih faz nič). Bifurkacijska točka – kritično področje, kjer se preklapi iz antifazne v fazo gibanja, zavzame prostor, ko frekvenca narašča – točka, kjer je $d\phi/dt$ maksimalen (t. j. kjer je razmerje sprememb relativnih faz maksimalno). Odvisno od naših interesov, lahko nato vstavimo znane vrednosti v enačbo in jo rešujemo tako, da dobimo razne, še neznane vrednosti. Z odkritjem vrednosti za a ,

b in ϕ , ko je $d\phi/dt = 0$, lahko utemeljimo razlago o tem, *zakaj* $d\phi/dt$ zavzame vrednosti, ki jih, ali napovemo, ali bo sistem stabilen pri teh vrednostih a , b in ϕ .

Ta primera sta dokaj enostavna primera v teoriji dinamičnih sistemov – zato ju poimenujemo tudi »klasična dinamična sistema« – v nasprotju z drugimi, kompleksnimi, ki uporabljajo teorijo kaosa, entropijo, in neravnotežno dinamiko, ki zahtevajo neprimerno bolj kompleksne metode razlag in filozofskih pristopov in se jih bomo le dotaknili v zaključku.

Vlada prepričanje, da tudi ta vrsta klasičnih dinamičnih modelov pomembno utrjuje nereduktivni fizikalistični pogled na duha in kognicijo, saj mnogi izhajajo iz predpostavke, da morajo biti dinamični kognitivni sistemi povezani z emergenco. Še bolj, obstaja vrsta zapisov zahtev, da mora biti dinamična kognitivna znanost pomembno povezana z doktrino emergence. Tim van Gelder, npr. meni:

»Širše dinamične perspektive je kognicija videti kot pojavi izhod iz potekajočih interakcij vrste povezanih kvantitativnih spremenljivk bolj kot sekvenčna diskretna transformacija, ki izhaja iz ene podatkovne strukture v drugo.« (Van Gelder, 1999: 12)

Drugi avtorji, kot npr. Jeff Elman, pa so podali nekoliko bolj močno zahtevo, da dinamični modeli lahko razložijo emergenco:

»Modeli dinamičnih sistemov zadevajo tudi interakcije in pojavnost; bolj splošno jih lahko upoštevamo kot *matematični okvir za razumevanje* vrste pojavnosti in višje rednih interakcij, ki jih najdemo v obeh, tako v konekcionizmu kot tudi v modelih umetnega življenja.« (Elman, 1998: 504)

Navkljub vsemu se zdi, kot da se skriva delček redukcionizma tudi v literaturi o dinamiki. Še posebej, kot smo že omenili, mnogi dinamiki vidijo matematiko, teorijo dinamičnih sistemov kot jezik za opisovanje tako psiholoških kot tudi fizioloških vidikov kognicije. Iz tega lahko predpostavimo poenotenje različnih nivojev opisov, zato lahko menimo, da teorija dinamičnih sistemov prevzema nekakšno vrsto redukcionizma, podobno tisti, ki jo najdemo v t. i. »poenotenih znanosti«, gibanju, ki je bilo popularno v prvi polovici dvajsetega stoletja.

DINAMIČNA RAZLAGA

Pričnimo z zapleti za emergentizem, ki so posledica vrste razlag v dinamični kognitivni znanosti. Razlagalni cilji dinamicizma so relativno znani. Dinamiciisti želijo najti diferencialne enačbe, ki bodo opisale želen kognitivni fenomen, saj na ta

način lažje in bolje opišejo naravo kognicije. Ta cilj je mnogokrat v nasprotju s klasično kognitivno znanostjo, ki poskuša to doseči podobno kot računalniški programi – z vrstami algoritmičnih kod, ki opisujejo diskretne prehode med statičnimi stanji.

Splošno so dinamični sistemi »sistemi, ki se v času razvijajo z iteriranjem aplikacije ob uporabi dinamičnih pravil«. (Jost, 2005) Če uporabimo to definicijo za *kognitivne sisteme*, dobimo zapis, ki sta ga zapisala Van Gelder in Port, dva najpomembnejša zagovornika dinamičnih sistemov:

»Dinamični modeli določajo, kako doseči spremembe pri spremenljivkah stanja pri katerikoli odvisnosti trenutne vrednosti te spremenljivke same ali drugih parametrov. Rešitve vodilne enačbe povedo, kakšno bo stanje sistema v katerikoli časovni točki tako dolgo, dokler je znano začetno stanje in čas, ki je pretekkel.« (Van Gelder, Port, 1995: 19)

To na kratko ponazorimo na modelu HKB. (Haken, et al., 1985) Za izhodišče so uporabljene gibalne diferencialne enačbe, s katerimi lahko ponazarjamo gibanje posamičnih prstov in študiramo povezave med njimi. S pomočjo te enačbe lahko izračunamo za predpisane začetne pogoje, lego prstov (stanje sistema), v katerikoli časovni točki za znano časovno stanje in pretekli čas.

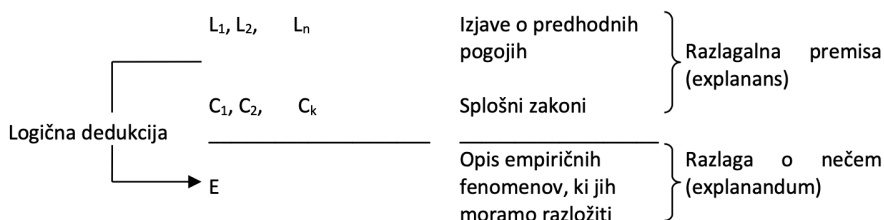
Podobno je zapisal tudi Andy Clark:

»Matematika specificira dinamične zakone, ki določajo, kako se vrednosti niza spremenljivk stanja razvijajo skozi čas. (Takšni zakoni so lahko npr. sestavljen sistem diferencialnih enačb). Če podamo začetno stanje, potem trenutne sekvence stanja, ki so določene z zakoni dinamike, ustvarijo trajektorijo skozi prostor.« (Clark, 1997: 100)

Tako je dinamični koncept kognicije predstavljen z zakoni psihologije, izražene z različnimi diferencialnimi enačbami, psihološka razlaga pa vsebuje prikaz, kako se razgrne ustrezen fenomen v odvisnosti od tega. Če vpeljemo za presojanje kriterij uspešnosti nekega modela, potem lahko ugotovimo, da so nekateri dinamični modeli že izpolnili ta pogoj, npr. Kelsov model ritmičnega gibanja prsta, ki predstavlja, kako lahko kompleksnost opazovanega fenomena elegantno pretvorimo v matematični zapis. Na temelju teorije dinamičnih sistemov je bistveni del modela diferencialna enačba (katere detajle bomo načelno obravnavali v naslednjih poglavjih), ki omogoča izpeljavo opazovanega vedenja in na ta način podaja znanje o začetnih vrednostih spremenljivk, podaja začetno stanje opazovanega sistema. V odvisnosti od našega zanimanja lahko sedaj znanje s pomočjo tega modela nadgrajujemo, v model

vstavimo vrednosti tako dobljenih spremenljivk in izpeljemo nove vrednosti, vrednosti spremenljivk, ki jih želimo napovedati ali razložiti. Tako z napovedovanjem in razlaganjem ritmičnih gibov prstov z uporabo Kelsovega modela dobimo razlago *krovnih zakonov* (angl. covering laws). Model krovnih zakonov obravnava razlago problema kot dedukcijo iz zakonov v konjunkciji s stališči o predhodnih pogojih. Hempel to definira kot:

»[Krovno zakonska] razlaga odgovarja na vprašanje: »Zakaj so se razlagalni fenomeni sploh zgodili?« s tem, da pokaže, da fenomeni izhajajo iz določenih partikularnih okoliščin $C_1, C_2, \dots, C_k$, v odvisnosti od določenih zakonov $L_1, L_2, \dots, L_r$. Z izpostavljanjem tega argument sam kaže, da ko postavimo pod vprašaj določene okoliščine in določene zakone, postane pojav tega fenomena *pričakovan* in tako v tem pomenu ta razlaga omogoča, da *razumemo, zakaj* se je obravnavani fenomen sploh zgodil.« (Hempel, 1965: 337)



Slika 2.4: Oblika Hemplove in Oppenheimove razlage krovnega zakona

Na osnovi splošno znanega stališča, ki sta ga izrazila Hempel in Oppenheim, ima razlaga strukturo veljavnih deduktivnih argumentov, v katerih je razlaga povezava dveh premis – izjave o začetnih pogojih in izjave o splošnem zakonu – in pojasnilo je zaključek, ki vsebuje opis empiričnega fenomena, ki smo ga želeli razložiti, kar grafično prikazuje slika 2.4. (Walmsley, 2008)

Primer (Hempel, 1965):

Živosrebrni termometer uporabljamo za merjenje temperature vročih teles, predvsem kapljev. Če termometer npr. potopiš v vročo vodo, to povzroči, da se v prvem trenutku živosrebrni stolpec v termometru zniža, čemur nato sledi hiter dvig. Kako bi lahko razložili ta fenomen?

Porast temperature na začetku vpliva le na stekleno cevko (kapilaro), v kateri se nahaja živo srebro. Ta se pod vplivom toplote razširi, poveča se volumen prostora, v katerem se nahaja živo srebro, s čimer se posledično zniža nivo živega srebra v tej cevki. Takoj nato, ko zaradi prevajanja toplote porast temperature doseže živo srebro, se njegov volumen prav tako prične povečevati. Ker je temperaturni koeficient prostorskega raztezka (termični razteznostni koeficient) mnogo večji kot pri steklu, to vpliva na dvig nivoja živega srebra v stekleni cevki. Ta razlaga je sestavljena iz dveh vrst izjav. Prva vrsta izjav opisuje določene pogoje, ki so izpolnjeni *v odvisnosti od oz. v istem trenutku kot*, ko bo fenomen razložen. Na kratko bi lahko to vrsto izjav poimenovali *predhodni pogoji*. V našem primeru ta predhodni pogoj vsebuje med ostalim tudi dejstvo, da je termometer sestavljen iz steklene cevke, ki je delno napolnjena z živim srebrom, ter da termometer potopimo v vročo vodo.

Izjava druge vrste pa izraža določene splošne (krovne) zakone; v našem primeru ti vsebujejo zakone termičnega raztezanja tako živega srebra kot tudi stekla, ter izjavo o majhni toplotni prevodnosti stekla. Ti dve skupini izjav, če ju adekvatno in celovito formuliramo, razložita obravnavani fenomen, kar posledično pomeni, da se bo nivo živega srebra v začetku nekoliko znižal, nato pa bo skokovito narastel.

Obravnavani primer lahko tako razložimo z obravnavo splošnih (krovnih) zakonov s tem, da pokažemo, da se dogodek sklada s temi zakoni ob upoštevanju določenih posebnih, začetnih pogojev.

Model krovnega zakona razlaganja je bil v ozki povezavi z redukcijo, ki jo je razvil Nagel (1961), saj je v obeh deduktivna relacija bistvena. Po Nagelovskem modelu redukcije, če lahko sklepamo, da zakon teorije *M* lahko izpeljemo iz zakona teorije *K*, potem lahko rečemo, da lahko *M* reduciramo v *K* – redukcija je izvedljiva. (Walmsley, 2010) Včasih bi lahko celo rekli, da so primeri Nagelovske redukcije primeri, kjer je reducirana teorija razložena s samo teorijo redukcije. Tako je Silberstein zapisal, da Nagelovski model redukcije

»/.../ smatra interteoretsko redukcijo kod deduktivno, in v posebnem primeru (krovni zakon) tudi kot pojasnjevalno.« (Silberstein, 2002: 85)

Ta tesna povezava med razlago krovnega zakona in Nagelovsko redukcijo postavlja zanimiv problem pred dinamične kognitivne znanstvenike; razlaga po krovnem zakonu zahteva vrsto deduktivnih zvez med razlagalno premiso (tisto, s čimer razlagamo – lat. *explanans*) in samo razlago o nečem (skratka tisto kar razlagamo – lat. *explanandum*), kjer pa večine teorij o emergenci zahtevajo odsotnost takšnih dedukcij, saj je emergenca vedno neredukcionistična.

Vzemimo pod drobnogled upoštevanje takšnih razlagalnih modelov. Po eni strani nič več ni možno govoriti o vedenju dinamičnih sistemov kot pojavnosti v tradicionalnem tehničnem pomenu, kar bi neposredno zelo verjetno povežalo te modele z Nagelovskimi modeli redukcije. Če lahko sklepamo, da je opis vedenja gibanja prstov sestavljen iz zakona (Kelsovega modela) in začetnih stanj (tj. specificiranja začetnih vrednosti spremenljivk), potem je ustrezneje, da rečemo, da je gibanje prstov reducibilno v Nagelovem pomenu na omenjeno teorijo dinamike. Upošteva ta razkorak si oglejmo enega od možnih odzivov.

2.5 Dinamični sistemi in »utelešena kognicija«

Zagovorniki dinamičnih sistemov v kognitivni znanosti, nekateri govorijo kar o »nekognitivistični« kognitivni znanosti, menijo, da bi se kognitivni znanstveniki lahko izognili mnogim težavam, če bi opustili reprezentacijsko teorijo in se namesto tega naslonili na nereprezentacijske mehanizme, ki delujejo v interakciji z okoljem, v katerem je organizem. Dober primer takega pristopa je robotika. (Brooks, 1999) Na primer, preprosti roboti za gibanje po okolju nimajo zapletenih kognitivnih zemljevidov, ampak uporabljajo le preproste notranje ure in mehanizme, ki sledijo svetlobi.

Osnovne predpostavke kognitivistov, ki se pri razlagah sklicujejo na strukture v glavi (prepričanja, pravila, pojmi in sheme), po mnenju zagovornikov »utelešene kognicije« (angl. EEC – Embodied Embedded Cognition) niso ustrezne. Namesto pojmov reprezentacije in računanja so v središču teorije dinamičnih sistemov novi pojmi, kot so nelinearnost, procesiranje v realnem času, prostor stanj, *atraktor*,²³

²³ Atraktorji so končna stanja mreže, ki so lahko točkasta (kar pomeni, da mreža obstane v tem stanju) ali periodična (mreža kroži po periodičnem zaporedju stanj - ciklu).

samoorganizacija in povratna zanka. V slednjem so dediči kibernetikov, ki so kot glavno teoretsko orodje uporabljali prav kontrolni sistem povratne zanke.

Dinamične hipoteze predpostavljajo, da so kognitivni agenti dinamični sistemi v nasprotju s prevladujočo računalniško hipotezo, ki trdi, da so kognitivni agenti digitalni računalniki. Skratka, dinamična hipoteza postavlja empirično alternativo računalniški hipotezi. (Van Gelder, 1999)

David Hume je sanjal o znanstveni psihologiji, v kateri bi matematični zakoni urejali mentalno sfero, prav tako kot Newtonovi zakoni urejajo materialno sfero. (Hume, 2000) Univerzalno silo gravitacije, kjer telesa učinkujejo sorazmerno s svojo maso, bi lahko nadomestili z univerzalno silo asociacij, kjer bi ideje učinkovale sorazmerno svoji podobnosti. Dinamika duha bi se postavila vzporedno z dinamiko materije.

Humove sanje pa niso prva takšna vizija duha, navdihnjena z odkritji znanosti. Sodobni fiziki so odkrili izjemno enostavne in elegantne matematične zakone, katerih matematično reševanje je izjemno zapleteno. Eden zgodnjih začetnikov te teorije je bil Thomas Hobbes, ki je razvil svoj model mehanizma mentalnih operacij in pri tem izhajal iz predpostavke, da so misli simbolni preračun, s pravili usmerjana manipulacija simbolov znotraj glave. (Hobbes, 1651/1994)

Ugibanja 17. stoletja postanejo v 20. stoletju znanost. Hobbsova ideja se razvije v *računsko hipotezo* (angl. computational hypothesis – CH) in kognitivni agenti postanejo v osnovi digitalni računalniki.²⁴ Morda najbolj znana razlaga tega postane Newellova in Simonova (1972) doktrina, da: »Fizikalni simbolni sistem izpolnjuje potreben in zadosten pogoj za posplošene aktivnosti inteligence.« (Newell, Simon, 1972: 116) To hipotezo sta predlagala kot *zakon kvalitativne strukture*, primerljivo s celično doktrino v biologiji ali tektonskimi ploščami v geologiji. Izraža osrednji vpogled v raziskovalne paradigme, ki so bile dominantne v kognitivni znanosti skoraj 40 let.

²⁴ Digitalni - analogni računalnik: Analogni računalniki so oblika računalnika, ki uporablja za reševanje problema zvezni način spreminjanja fizikalnih fenomenov, kot so električne, mehanske ali hidravlične vrednosti. Verjetno je bil eden prvih mehanskih analognih »računalnikov« Antikythera mehanizem, konstruiran zato, da bi preračunaval astronomske lege. Predvideva se, da je bil izdelan okrog 100 let p. n. š. Nasprotno od tega digitalni računalniki predstavljajo spreminjanje spremenljivk inkrementalno in to v binarni obliki (samo z 0 in 1).

V zadnjih letih pa je Humova alternativa nenadoma ponovno deležna pozornosti. Ena od najpomembnejših razvojnih prelomnic je ponoven vznik konekcionizma, ki modelira kognicijo kot dinamične sisteme. (Smolensky, 1995)

Podobno pomemben je bil tudi pojav kognitivne nevroznanosti in znotraj nje prevlada dinamičnih teorij. Dinamične oblike postajajo del splošnega okvirja pri mnogih delih s področja psihofizike, zaznav, razvojne psihologije, kognitivne psihologije, robotike in raziskav avtonomnih agentov, umetne inteligence in socialne psihologije. So osrednji del številnih splošnih modelov, kot so *ekološka psihologija*, *sinergija*²⁵ in *morfodinamika*.²⁶ (van Gelder, 1999)

Teorijo dinamičnih sistemov uporabljajo za raziskave na področju nevroznanosti in kognitivnega razvoja, še posebej v Piagetovih teorijah kognitivnega razvoja. Zastopa prepričanje, da se da kognitivni razvoj bolje predstaviti s fizikalnimi teorijami kot s teorijami, ki temeljijo na sintaksi in umetni inteligenci. Prav tako zastopa prepričanje, da so diferencialne enačbe najprimernejše orodje za modeliranje človeškega vedenja. Te enačbe so interpretirane tako, da predstavljajo agentovo kognitivno trajektorijo skozi prostor stanja. Če to povemo z drugimi besedami oz. dinamiki dokazujejo, da bi morala biti psihologija opis (z diferencialnimi enačbami) kognicije in vedenja nekega agenta v določenem okolju in pod notranjimi pritiski. Za to se pogosto uporablja tudi jezik kaosa.

2.6 Utelešena kognicija

Utelešena kognicija (UK) je filozofsko teoretična pozicija v kognitivni znanosti, ozko povezana s situacijsko kognicijo, poosebljeno kognicijo, poosebljeno kognitivno znanostjo in teorijo dinamičnih sistemov. Teorija navaja, da se inteligentno obnašanje pojavlja izven vzajemnega delovanja med možgani, telesom in svetom. Svet ni le »igrišče«, na katerem igrajo možgani. Natančneje, možgani, telo in svet so enako pomembni dejavniki pri razlaganju tega, kako se posebno, inteligentno vedenje v praksi pojavi.

²⁵ Sinergija (iz grške besede synergos, συνεργός, ki pomeni delati skupaj) se nanaša na pojav, ko dva ali več povzročiteljev oziroma vplivov, ki delujejo skupaj, ustvari učinek, ki je večji od učinka, ki bi ga napovedali, če poznamo samo učinke posameznih povzročiteljev ali vplivov.

²⁶ Morfodinamika se nanaša na dinamične spremembe v morfologiji, vedi o zgradbi in obliki organizmov (oblikoslovje).

Poosebljanje se definira kot ideja, da notranji telesni ustroj močno vpliva na višje kognitivne procese v možganih, še posebej preko emocionalnega sistema. (Damasio, 1996, 1999)

Utelesenost se označuje kot ideja, da fizične interakcije med telesom in svetom močno omejujejo možno vedenje organizmov. Fizična interakcija vpliva na vedenje organizmov, kar vpliva na kognitivne procese.

Ta teorija je neposredna reakcija na trenutne dominantne kognitivne simbolne paradigme, ki trdijo, da je za kognitiven sistem ključen računalniško predstavitveni sistem (kot npr. računalniški software), ki procesira vhodne informacije in generira izhodne (vedenja) na osnovi notranjega informacijskega procesa. V tradicionalnem kognitivizmu vzročni koren vedenja leži v »virtualnih« procesih, ki jih vodi software, in se izvaja v naših možganih. Možgani so le strojna oprema (hardware, na katerih je programska oprema implementirana. Telo (senzorji in aktuatorji) so samo vhodno-izhodne naprave, ki so v službi možganov. Svet je zgolj igrišče (objekt), na katerem kognitivni agent deluje.

V nasprotju s to simbolno teorijo utelešena kognicija zagovarja tezo, da aktualni fizični procesi v telesu in interakcije med telesom in svetom delno ustvarjajo nekaj, kar lahko poimenujemo »kognitivni sistem kot celota«. Telo, svet in možgani ustvarjajo sistem. Skupaj ti deli sistema povzročajo inteligentno vedenje, ki je na naraven način povezano s teoretičnim konceptom EEC. Vse lepo in prav, vendar imamo pri tej teoriji na koncu mnogo več vprašanj kot odgovorov. Če naštejemo le nekatera:

- Je UK resna samostojna (pozitivna) teorija, ali je samo šopek negotovanj o tem, kaj vse je narobe v kognitivizmu?
- Je UK deskriptivna, namesto tega, da bi v resnici kaj razlagala o kogniciji?
- Kako lahko EEC razloži lingvistične procese in procese eksplicitnega kognitivnega razmišljanja?
- Katere bi bile najbolj informativne empirične hipoteze z EEC perspektive?
- Ali lahko uporabljamo tradicionalne metode (npr. stimulus – odziv paradigme) eksperimentalne psihologije za preverjanje EEC hipotez?

2.7 Situacijska kognicija in konekcionizem

Situacijska kognicija (angl. *situated cognition*) meni, da je znanje neločljivo od dejanja s tem, da prikazuje, da je vse znanje aktivno povezano s socialno, kulturno in fizično sfero. S to predpostavko, ki zahteva epistemološki premik od empirizma, spoznavna teorija predlaga model znanja in učenja ki zahteva bežno razmišljanje bolj kot pomnjenje in priklic konceptualnega znanja. V bistvu kognicija ne more biti ločena od konteksta. Zato učenje smatramo kot posameznikovo povečevanje njegovih lastnih sposobnosti skozi različne okoliščine (izkušnje) bolj kot kopičenje znanja. S perspektive spoznavne kognicije lahko svet, telo in kognitivne funkcije možganov analiziramo tako, da uporabljamo enaka konceptualna orodja kot so kompleksne spominske strukture ali mentalni modeli, npr. inštrukcije, ki lahko pomagajo učečemu asimilirati in akomodirati nove informacije v že obstoječe informacije ali obstoječe kognitivne strukture. To je še posebej pomembno zato, ker v veliki meri poenostavlja naše razumevanje preslikav med temi področji in se z lahkoto interpretira kot biološka adaptacija telesa in možganov na okolje.

Konekcionistični model je diskretni dinamični sistem in takšni so tudi algoritmi učenja v njem. Vendar so temeljni kamen resničnega dinamičnega pristopa h kogniciji fenomeni, ki se dogajajo v zveznem času, nekaj, česar vsi konekcionistični modeli ne zmorejo. Seveda se nevronske mreže pogosto uporabljajo za študijo časovno spreminjajočih se fenomenov, vendar obstajajo tudi druge dinamične metode, ki ne uporabljajo konekcionističnih mrež. Razvoj konekcionističnega modeliranja od leta 1980 je zagotovo pripomogel k napredku razvoju raziskav v smeri dinamičnega razmišljanja, vendar pa konekcionistični modeli ne prikazujejo vedno najboljše dinamičnih hipotez o kogniciji.

2.7.1 Prednosti in slabosti dinamičnih modelov

Dinamično modeliranje ponuja vrsto pomembnih prednosti pred tradicionalnimi računalniškimi metodami modeliranja kognicije. (Port, 2002) Če navedemo le najpomembnejše:

1. Prvič, biološka verjetnost digitalnih, časovno diskretnih modelov še vedno ostaja velik problem. Ne ve se ne kako in ne kje naj bo v možganih naprava (predel), ki bi deloval podobno kot računalniški čip. Najpogostejši odgovor v

preteklosti je bil: »Ne vemo zares, kako pravzaprav možgani delujejo, zato je ta hipoteza prav tako verjetna, kot katerakoli druga«. (Chomsky, 1956) Takšni argumenti izpred trideset, štirideset let niso več sprejemljivi danes, ko lahko z nevropsihološkimi eksperimentalnimi metodami prikazujemo različne oblike diskretnosti. Vendar to še vedno ne potrjuje enostavne predpostavke, da bi bila kakršnakoli oblika enot in operacij zveznih sistemov uporabna za digitalni model kognicije.

2. Drugič, končno lahko časovne podatke na osnovi tega neposredno vključimo v kognitivne modele. Če želimo napovedovati kritični razpon dogodka v času, lahko sedaj povežemo fenomene, kot so:
 - procesiranje časa (npr. reakcijski čas, čas prepoznavne, odzivni čas itd.);
 - časovne strukture v motoričnih vedenjih (npr. sposobnost, govor, lokomotorika, ples),
 - časovne strukture in stimuliranje (npr. za govorno in glasbeno zaznavo, interpersonalno koordinacijo, ko gledamo teniško igro itd.).

Jezik dinamičnih sistemov zagotavlja konceptualni slovar, ki omogoča poenotenje kognitivnih procesov v možganih tako s psihološkimi procesi v naši telesni periferiji kot z dogodki okolja izven organizma. Poenotenje procesov skozi te mehke (fuzzy) in delno umetne meje omogoča resnično posebej oz. utelešeno umeščanje razumevanja človeškega vedenja vseh tipov. Diskretno časovno modeliranje tradicionalnih približkov zahteva, da postavljamo meje vedno nekje izven diskretnega časa digitalnih vidikov kognicije zvezno-časovne psihologije (kompetence nasproti performansam). (Chomsky, 1956)

3. Tretjič, kognitivni razvoj in časovno zaporedno procesiranje lahko sedaj integriramo, saj učenje zaznavnih in motoričnih vedenj usmerjajo podobni procesi, tudi če imamo različne časovne skale. Simbolni ali računalniški modeli nasproti temu zahtevajo, da obravnavamo učenje in razvoj obeh kot popolnoma različna procesa, popolnoma nepovezana z motoričnimi in zaznavnimi aktivnostmi.

Končno, upoštevanje zgornje prednosti dinamičnega modeliranja je dejstvo, da dinamični modeli vsebujejo diskretni čas in digitalne modele kot poseben primer. Uporabimo jih tam, kjer ostale poti niso možne (vzorčenje zveznih dogodkov dovoljuje diskretno modeliranje zveznih funkcij, vendar ostanejo simulacije same po sebi diskretne in so samo modeli zvezne funkcije do polovice vzorčne mere). (Port, Cummins, McAuey, 1995) Tako je tudi vsak digitalni računalnik dinamični sistem z realnimi vrednostmi napetosti v zveznem času, ki so diskretizirane z neodvisno notranjo uro, ki jo ima vsak digitalni računalnik in ki uravnava njegovo delovanje. Ta moč dinamičnega modeliranja je zelo pomembna za naše razumevanje človeške (ali pa tudi živalske) kognicije.

Obstaja tudi vrsta slabosti dinamičnega modeliranja.

1. Prva je, da je matematika dinamičnih modelov neprimerno manj raziskana in manj razvita, kot je matematika digitalnih sistemov. Zato je očitno težje razviti ustrezne modele, razen morda za previdno omejene in enostavne (pohlevne) primere.
2. Druga slabost je, da med nekaterimi kognitivnimi fenomeni (npr. izvajanje dolgih preračunov, razvijanje algoritmov, procesi o uporabi jezika) ljudje izvajajo operacije z diskretnimi simboli.

Vendar pa je potrebno izpostaviti, da so dinamični sistemi sposobni izkazovati tudi digitalno obnašanje (npr. nevronske verjetnostni modeli) in je le vprašanje časa, kdaj bodo razviti do mere splošne uporabnosti. Očitno je, da so računski (simbolni) modeli trenutno še vedno enostavnejši in bolj neposredni, čeprav imajo vrsto pomanjkljivosti. Kratkoročno bodo verjetno prevladovali, predvsem v obliki različnih hibridnih sistemov, dolgoročno pa bo prihodnost skoraj zagotovo »dinamična«.

2.8 Inteligenca

Če izključimo zavest, lahko razmišljanje prevedemo v matematično predvidljive aktivnosti, katerih osnova so vnaprej znane informacije.

(Mortimer Taube)

Inteligenca²⁷ izhaja iz latinske besede *intelligere*, ki izhaja iz besed *inter-legere*, kar pomeni izbrati, razločiti, razumeti ali opaziti, spoznati. Oblika besede *intelectu* je v srednjem veku postala tehnični termin za razumevanje in prevod za grški filozofski termin *nous*, ki je ozko povezan z metafiziko in teorijo neumrljive duše in konceptom aktivnega intelekta (angl. active intellect, tudi active intelligence). (Neisser, et al. 1996) Če bi hoteli inteligenco definirati s sodobnim jezikom, bi zelo težko prišli vsaj do približno enoznačnega odgovora. Če povzamemo nekatere najbolj razširjene:

1. Inteligenca je zelo splošna sposobnost, ki, med ostalim, vsebuje sposobnost razmišljanja, načrtovanja, reševanja problemov, abstraktnega razmišljanja, tvorjenja kompleksnih idej, hitrega učenja in učenja iz izkušenj. Ne samo učenje iz knjig, ozka akademska veščina ali reševanje inteligenčnih testov. Bolj kot to, odraža globljo in širšo sposobnost vključevati in prilagajati se okolju, delati stvari smiselne ali ugotavljati, kaj narediti. (Gottfredson, 1997)
2. Posamezniki se razlikujejo med seboj po razumevanju kompleksnih idej, prilagajanju okolju, učenju iz izkušenj, po uporabi različnih oblik sklepanja, premagovanju ovir z razmišljanjem. Koncept "inteligence" je poskus razložiti ves ta kompleksen niz fenomenov. Na nekaterih področjih smo že dosegli dokaj natančne definicije, medtem ko na ostalih še vedno vlada velik nered in nerazumevanje. Če bi danes 12 najbolj eminentnih strokovnjakov poprosili, da povedo definicijo inteligence, bi dobili dvanajst na nek način različnih odgovorov. (Perloff, et al., 1996)

Učenje, znanje in inteligenca so ozko povezani. V sami definicijah se pojavi tudi termin prilagajanje (adaptiranje) oz. ozko z njim povezano – *učenje*. Zato, da lahko rešujemo probleme, posamezniki potrebujemo:

1. znanje,
2. sposobnost uporabe teh znanj

VRSTE INTELIGENTNOSTI

Sistemov (živih ali neživih) ne moremo strogo razporejati po količini inteligence, saj moramo pri tem upoštevati, da obstajajo različne vrste inteligence. Začetnik in eden najbolj znanih avtorjev na tem področju je Howard Gardner (2011), ki je

²⁷ V nadaljevanju ne bomo delali posebne razlike med terminoma inteligenca in inteligentnost.

intelligentnost opredelil kot zmožnost reševanja problemov ali ustvarjanja izdelkov, ki so cenjeni v enem ali več kulturnih okoljih. Predhodne opredelitve intelligentnosti so se mu zdele pomanjkljive, saj so vključevale predvsem jezikovno, logično-matematično in delno prostorsko intelligentnost. Gardner je tako razvil teorijo o sedmih intelligentnostih, ki jih lahko vsi razvijemo do določene mere kompetentnosti. Te inteligence po Gardnerju so:

- *Jezikovna intelligentnost* je sposobnost učinkovitega besednega izražanja in razumevanja kompleksnih besednih sporočil. Jezikovna intelligentnost vsebuje občutljivost za pomen besed in za različne funkcije jezika. Uporabljamo jo na primer pri branju, tvorjenju in razumevanju besedil.
- *Logično-matematična intelligentnost* je sposobnost abstraktnega mišljenja in logičnega razmišljanja ter sklepanja. Uporabljamo jo pri reševanju abstraktnih matematičnih problemov, pri deduktivnem in induktivnem mišljenju (včasih govorimo tudi o dveh ločenih inteligencah, namreč induktivni in deduktivni inteligenci). Logično-matematična inteligenca je še posebej pomembna pri znanstvenem raziskovanju.
- *Prostorska intelligentnost* temelji na tvorbi vidnih predstav o poteku ali videzu stvari (barva, oblika, postavitev itd.). Ta intelligentnost obsega tudi dobro orientacijo, to je sposobnost, znajti se v prostoru in presoditi razmerja med predmeti v prostoru. Uporabljamo jo na primer pri gibanju po neznanem terenu, risanju ipd.
- *Glasbena intelligentnost* je sposobnost ustvarjanja in poustvarjanja glasbe, prepoznavanja melodij. Ljudje, ki imajo visoko razvito glasbeno intelligentnost so izjemno občutljivi tudi na barvo in ritem. Glasbena intelligentnost se izraža pri igranju instrumentov, petju, skladanju.
- *Telesno gibalna intelligentnost* temelji na spretni uporabi telesa, nadzorovanju in koordinaciji telesnih gibov. Uporabljamo jo pri usklajenem gibanju (športu, plesu), pa tudi pri spretnem ravnanju s predmeti, kamor spadajo tudi ročne spretnosti.
- *Medosebna intelligentnost* (ponekod jo imenujemo tudi *socialna inteligenca*) se kaže v sposobnosti razumevanja drugih ljudi in uravnavanja medčloveških odnosov. Odraža se v spretnosti pri navezovanju stikov in komunikaciji z drugimi ljudmi, kot sposobnost razumevanja njihovih težav, motivov in čustev ter v učinkovitem odzivanju na vedenje drugih.

- *Znotrajosebna inteligentnost, čustvena inteligenca* ali avtorefleksivna inteligentnost je sposobnost zavedanja in razumevanja lastnega doživljanja in vedenja. Ljudje, ki imajo visoko izraženo avtorefleksivno inteligentnost, razumejo vzroke in posledice lastnega doživljanja (na primer čustvovanja in vedenja) ter imajo dobro razvito samokontrolo.

Čustva in socialne veščine igrajo pomembno vlogo tudi pri inteligentnih računalniško vodenih sistemih. Pri čustveni in socialni inteligenci pri računalniško vodenih inteligentnih sistemih je potrebno predvideti predvsem odzive drugih, za kar je potrebno razumevanje motivov in čustvenih stanj (da nam to v računalništvu lahko uspe, uporabljamo npr. teorijo iger, teorijo odločanja, kot tudi sposobnost modeliranja človekovih čustev in veščin ter zaznave teh čustev). Za dobro interakcijo med človekom in računalnikom mora biti inteligentni stroj sposoben *izkazovati* čustva. Poleg vsega tega pa mora biti pri tej interakciji do človeka uglasen in občutljiv. Najbolje bi bilo, da bi imel tudi sam čustva. Mnogokrat se za to uporablja teorijo emocionalnega računalništva (angl. affective computing), ki je interdisciplinarna veda in združuje tako računalničarje kot tudi psihologe in kognitivne znanstvenike ter predstavlja razvoj sistemov in naprav, ki lahko razpoznavajo, interpretirajo, procesirajo in simulirajo človeške odzive. Začetnik te znanosti je bila Rosalinda Picard, ki je leta 1997 objavila svoj članek na temo emocionalnega računalništva. V svojem delu je izpostavila motive, zakaj meni, da je pomembno simulirati empatijo, in trdi, da naj bi bili inteligentni stroji sposobni interpretirati čustvena stanja ljudi in se jim s svojim vedenjem in odzivi prilagajati.

Nekateri avtorji opisujejo tudi več kot sto vrst človeške inteligence. Sistem (človek ali stroj) je lahko boljši v nekem tipu inteligence in slabši v drugem. Očitno je, da potrebujemo integriranje vseh različnih oblik inteligence v poenoten sistem (nekakšen nadzorni sistem), tako da lahko med reševanjem problemov preklapljamo med različnimi oblikami inteligence. Pri tem večina špekulacij o umetni inteligenci sedaj še ne upošteva naslednjega nivoja – zavesti (ki bi lahko bila dober kandidat za nadzorni sistem, kar bomo poskušali prikazati v naslednjem poglavju).

2.8.1 Inteligentnost, ustvarjalnost in nadarjenost

Večina raziskav, ki je skušala povezati velikost možganov z inteligentnostjo, poroča o dokaj nizkih korelacijah, ki se gibljejo okoli vrednosti 0.35. Podobni so izidi, ki temeljijo na preučevanju vojaških veteranov z možganskimi poškodbami. Tudi

najhujše poškodbe možganov so vodile v zmanjšanje IQ za največ 10 točk. Če bi se osredotočili na te ugotovitve, bi lahko rekli, da inteligentnost nima veliko skupnega z velikostjo naših možganov.

Raziskovalce, ki so se ukvarjali z odnosom med ustvarjalnostjo in možgansko aktivnostjo, je najbolj pritegnila hemisferičnost – dejstvo, da je vsaka od obeh hemisfer – možganskih polovic, specializirana za določeno vrsto in način obdelave podatkov. Z gotovostjo lahko rečemo, da govor nadzoruje leva možganska polovica ali hemisfera in da s prostorskim razmišljanjem ali prepoznavanjem obrazov upravlja desna možganska hemisfera. Leva hemisfera je izrazito razumska, besedna, produktivna, medtem ko je desna hemisfera izrazito čustvena, vizualna, nazorna, intuitivna, meditativna. Ko govorimo o zmožnosti razsojanja, obe možganski hemisferi delujeta skladno. Vsak dejavnik spoznanja ima svoje sestavne dele tako v desni kot v levi polovici možganov. Da ne bi prišlo do tekmovanja med levo in desno hemisfero, med njima obstaja delitev »dela«.

Številne publikacije že od sedemdesetih let prejšnjega stoletja poskušajo razložiti skrivnosti ustvarjalnosti, ki izhajajo iz desne možganske hemisfere. Filozof Karl Popper (Poper, Eccles, 1978) in nevrokirurg, nobelovec John Eccles (Eccles, 1985), opisujeta desno hemisfero kot »podrejen, manjvreden« del možganov. Leva polovica možganov je dolgo veljala za prevladujočo sfero, ki je odgovorna za človekovo edinstveno zmožnost govora in s tem tudi za samozavedanje. Skladno s tem so obstajale razprave o tem, da je večji del ljudi desničarjev prav zato, ker je za govor zadolžena prevladujoča leva hemisfera, kar v primeru levičarjev kaže, da center za govor ni nujno v levi možganski polovici. S teoretičnega vidika je takšno sklepanje razumljivo. Veliko avtorjev je ustvarjalnost opisalo kot združevanje nasprotij. Koestler (1967) je predlagal termin bisociacije, ki opisuje ustvarjalnost kot združevanje nepovezanih struktur, ločenih zamisli, dejstev in zaznav v enovitih možganih. Prav zaradi tega sta bili dve hemisferi, ki sta predstavljali dva načina kognitivnega procesiranja, idealna nevropsihološka razlaga ustvarjalnosti. V prvi fazi je le malo ali nič interakcije med obema hemisferama. Vsaka hemisfera neodvisno od druge razvija svoje procese. V drugi fazi se interakcija poveča. Slikovitost, ki se je razvila v desni hemisferi, se realizira preko bolj logične leve hemisfere.

Še ena značilnost ustvarjalnega procesa – fazi inkubacije in iluminacije, v katerih se pozornost loči od problema in se rešitve spontano pojavljajo – je potrdila raziskovalce v domnevi, da je ustvarjalnost desnohemisferni proces. Nadaljnjo osnovo za te domneve najdemo tudi v psihoanalitičnih razlagah ustvarjalnosti. Za Freuda (1925) je ustvarjalnost izraz potlačenih osebnih izkušenj, ki so sestavina osebne podzvesti. Z druge strani Jung (1933) razume ustvarjalnost kot proizvod arhetipov, ki so srž kolektivne podzvesti.

ALI OBSTAJA RAZLIKA MED USTVARJALNOSTJO IN INTELIGENTNOSTJO?

Za boljše razumevanje sposobnosti je pomembno, da vemo, katere procese ljudje uporabljajo, kadar rešujejo različne vrste problemov. V psihologiji obstajajo številne študije, ki so se in se še zmeraj ukvarjajo s tem vprašanjem. Pri tem uporabljajo različne metode, celo različna imena za enake ali podobne psihološke fenomene. Na primer, faktorske teorije so se ukvarjale z razlikami med inteligentnostjo in ustvarjalnostjo, medtem ko so kognitivno usmerjene šole govorile o razlikah pri reševanju dobro in slabo opredeljenih problemov.

Osnovna strategija, ki jo ljudje uporabljamo za iskanje poti skozi dobro opredeljen problemski prostor, je analiza sredstvo – cilj. (Ernst, Newell, 1969) Za Andersona (2007) je to osnovni miselni proces sploh. Ta proces določata dve sestavini, zmanjševanje razlik in zastavljanje vmesnih stanj, ki jih imenujemo vmesne cilje. Zmanjševanje razlik pomeni, da izberemo takšne operacije, ki bodo proizvedle stanja, ki so bližja, bolj podobna temu, kar želimo doseči. Vmesna stanja v tem procesu krčenja imenujemo vmesni cilji. Simon (1979) domneva, da na ta način ljudje rešujemo vse probleme, tudi slabo definirane. Pri slabo definiranih si izberemo nek manjši problem (npr. pišemo esej o ustvarjalnosti in se lotimo najprej le enega odstavka ali poglavja), ki ga rešujemo, kot da bi bil dobro definiran. Iz te razlage lahko sklepamo, da je reševanje dobro definiranih problemov osnovno in je ustvarjalnost le poseben primer. Povsem nasprotno stališče zastopajo drugi raziskovalci, ki menijo, da ima ustvarjalno reševanje širšo možnost uporabe kot strategija sredstvo–cilj, ki je po njihovem uporabna le za skupino dobro opredeljenih problemov z natančno določenim ciljem.

Pričujoči pregled ugotovitev o odnosu med možgani in sposobnostjo je prav gotovo sprožil številna vprašanja. Vprašanja, ki morda nikoli ne bodo dobila odgovora. Vsekakor pa je pregled raziskav posredoval tudi nekaj ugotovitev, ki nam omogočajo boljše razumevanje inteligentnosti in ustvarjalnosti. Naj omenimo le nekatere. Ugotovitev, da nadarjeni posamezniki uporabljajo le za nalogo relevantna področja možganov, nakazuje, da je njihov način razmišljanja »manj kaotičen« in »bolj enostaven«. Takoj se zastavi vprašanje, ali to velja tudi za ustvarjalnost. Lahko bi rekli, da vsaj začetne faze reševanja ustvarjalnih problemov zahtevajo bolj kompleksno aktivacijo nevrnske mase.

Prezgodaj je še, da bi lahko trdili, da nevropsihološke ugotovitve podpirajo to ali ono teorijo inteligentnosti.

2.8.2 Umetna inteligenca

Dolgoročni cilj raziskav s področja strojnega učenja, ki danes postaja vse bolj popularno in uporabno, je ustvariti nek umeten sistem, ki bo s samostojnim učenjem dosegel ali celo presegel človeško inteligenco. Širše področje raziskav s tem istim pa na kratko poimenujemo *umetna inteligenca UI* (angl. *artificial intelligence*). Omenimo še enkrat njeno splošno definicijo povzeto po Marvinu Mynskyemu:

»Umetna inteligenca je *znanost o izdelavi strojev*, ki so sposobni narediti stvari, za katere je po naših merilih potreben um.« (Marvin Minsky, 1968, ustanovitelj laboratorija za UI na MIT-ju, v Copeland, 2007: 1)

oz. njeno simbolno predstavitev po sliki 2.5



Slika 2.5: UI in njene notranje relacije

Če govorimo o umetni inteligenci, ne pričakujemo od inteligentnega sistema (v tem primeru računalnika, stroja, robota), da je ekstremno inteligenčen (izjemno sposoben) samo v enem vidiku inteligence. Pričakujemo, da je kompleksen, inteligenčen na vseh področjih, ki jih zahteva človeška inteligenca pri reševanju problemov. Raziskave s področja UI se ubadajo z razvojem sistema, ki deluje bolj ali manj inteligentno in je sposoben reševanja relativno zahtevnih problemov. Te metode mnogokrat temeljijo na osnovi posnemanja človeškega reševanja problemov. UI področja, razen strojnega učenja, so povezana s predstavljanjem znanja, razumevanja govora, avtomatskega sklepanja in dokazovanja teoremov, logičnega programiranja, kvalitativnega modeliranja, ekspertnih sistemov, igranja iger, hevrističnih reševanj problemov, umetnih zaznav, robotov in kognitivnega modeliranja. O kognitivnem modeliranju bomo povedali več v naslednjem poglavju. Tu se osredotočimo le na splošne probleme, povezane z UI.

V vseh UI področjih igrajo pomembno vlogo algoritmi strojnega učenja. Praktično mora biti povsod prisotno učenje. Z uporabo tehnik učenja se lahko sistem uči in izboljšuje svoje zaznave (postaja adaptiven), razumevanje govora, sklepanje ipd. Področje logičnega programiranja je prav tako v tesni zvezi z induktivnim logičnim programiranjem, ki se uporablja pri razvoju logičnih programov npr. za določanje ciljev (npr. GPS naprave). Za razvoj ekspertnih sistemov lahko uporabljamo strojno učenje za ustvarjanje podatkovnih baz iz primerov usposabljanj reševanja problemov. Inteligentni roboti morajo neizogibno izboljšati svoj postopek za reševanje problemov s pomočjo učenja. Končno, tudi kognitivno modeliranje se praktično ne more izvajati brez upoštevanja učnih algoritmov.

MEJE SIMBOLNEGA RAČUNANJA

Teorija simbolnega računanja razkriva, da samo neznamen del vseh problemov, ki jih lahko formalno opišemo, lahko tudi algoritmčno rešimo. Danes v znanosti uporabljamo naslednje formalne simbolne jezike za opisovanje (modeliranje) realnosti:

- *Matematično logiko*, ki je matematična disciplina, ki preučuje formalne sisteme v povezavi z načinom, kako opišejo intuitivna koncepta dokaza in računanja kot dela temeljev matematike. Čeprav bi si laik lahko mislil, da je matematična logika matematike, je ta v resnici bližje *matematski logiki*. Sestavljajo jo tisti deli

logike, ki jih lahko matematično modeliramo. Nekdaj se je tej disciplini reklo *simbolna logika* (za razliko od filozofske logike) ali metamatematika (ta izraz je dandanes omejen na nekatere poglede teorije dokaza). (Boolos, et al., 2002) Matematična logika izhaja iz sredine 19. stoletja in je podpodročje matematike, ki se je osvobodilo tradicionalnega študija logike. (Ferreirós, 2001: 443) Pred tem pojavom so logiko študirali skupaj z retoriko in filozofijo.

- *Programski jezik* je stroju berljiv umetni jezik, ki je bil razvit, da izraža izračune, ki jih lahko izvaja stroj oziroma računalnik. Računalniška znanost je veda, ki raziskuje in se ukvarja s teorijo programskih jezikov.

Računalniške programske jezike kategoriziramo v *pet generacij*:

- *prva generacija* je strojni jezik, ki ga razume centralna procesna enota;
- *druga generacija* je zbirni programski jezik, katerega skladenjska pravila so opredeljena z mnemoniki in operandi oziroma parametri;
- *tretja generacija* so visoko nivojski programski jeziki (angl. *high-level programming languages* – HLL);
- *četrt* zelo visoko nivojski programski jeziki (angl. *very high-level programming languages* – VHLL);
- *peta generacija* (angl. *domain-specific languages* – DSL), ki opredeljuje programske jezike, ki služijo reševanju specifičnih problemov ozkega področja.

V računalniški stroki so programski jeziki razdeljeni v generacije glede na *abstrakcijski nivo*. Jeziki nižjih nivojev so bližje zapisu, ki ga razumejo stroji, medtem ko so jeziki višjih nivojev razumljivejši človeku in so skladenjsko bolj podobni naravnim jezikom.

Prvi visoko nivojski programski jezik je bil Planckalkül, ki ga je razvil nemški računalniški pionir Konrad Zuse med letoma 1943 in 1945. Leta 1950 so se razvili in uveljavili še programski jeziki LISP (LISt Processor), FORTRAN (FORmula TRANslator) in COBOL (COMmon Business Oriented Language). Danes pa je v uporabi vrsta jezikov, kot so C, C++, Java ipd., ki so osnova različnih računalniških aplikacij.

- *Rekurzivne funkcije* so funkcije, določene kot končno zaporedje kombinatoričnih korakov. Kleene je skupaj s Churchem, Kurtom Gödlom, Alanom Turingom in drugimi razvil področje rekurzivne teorije, ki je omogočila dokaz, ali so določeni razredi matematičnih problemov rešljivi ali nerešljivi. Beseda *rekurzivno* (latinsko *recurrere*, kar pomeni *teči nazaj*) pomeni nanašajoče na samega sebe. Rekurzivna teorija je vodila k teoriji računalniško izračunljivih funkcij, ki je raziskovala tiste funkcije, ki se dajo izračunati z digitalnim računalnikom. Rekurzija v matematiki in računalništvu pomeni podajanje funkcije na tak način, da se v definiciji sklicujemo na to isto funkcijo (vendar pri drugačnem argumentu). Tak način podajanja imenujemo *rekurzivno podajanje* ali *rekurzivna formula* (tudi *rekurzivna definicija*). Predstavlja občo metodo poenostavljanja, pri kateri problem razdelimo na podprobleme enakega tipa. V računalniškem programiranju to imenujemo *razdeli in zmagaj* in je ključ do konstruiranja mnogih pomembnih algoritmov. Nasprotna metoda je *dinamično programiranje*, pri katerem problem rešujemo z reševanjem vedno večjih primerov tako dolgo, da dosežemo želeno velikost. (Causey, 2001)
- *Formalna gramatika* se uporablja v računalništvu in jezikoslovju, mnogokrat jo poimenujemo preprosto *gramatika*. To je natančen opis formalnega jezika, to je skupine nizov znakov. Obstajata dve glavni kategoriji formalnih gramatik, in to:
 - *generativne gramatike*, ki predstavljajo skup pravil za ustvarjanje nizov znakov jezika;
 - *analitične gramatike*, ki predstavljajo skup pravil za analizo pripadnosti niza znakov jeziku. Analitična gramatika skratka opisuje, kako prepoznati niz znakov v skupini.

Ko je Noam Chomsky prvi predstavil formalizem generativnih gramatik daljnega leta 1956, jih je klasificiral v tipe, kar danes poznamo kot Chomskyjeva hierarhija. Razlika med temi posameznimi tipi je predvsem v tem, da imajo višji tipi vedno bolj stroga produkcijska pravila in s tem lahko izražajo vedno manj formalnih jezikov. Npr. gramatika neomejenih produkcij (Tip 0) obravnava katerikoli jezik, ki ga sprejema Turingov stroj, algoritemski sistem, miselni stroj. Turingov stroj si je zamislil angleški matematik Alain Mathison Turing leta 1936. Delo Turingovega stroja le oponaša človeško delo in računa po strogih predpisih. (Turing, 1950, Bregant, 2010)

Vsi ti formalizmi imajo enakovredno izpovedno moč in vsi imajo enake omejitve: lahko le delno opišejo fenomene v diskretnem svetu (diskretnih funkcijah) in praktično le nezaten del zveznega sveta (zvezne funkcije). Zato, če upoštevamo, da je svet zvezen, je najbolj verjetno, da ga ni moč opisati s katerimkoli formalizmom, ki smo ga sposobni ustvariti s svojim (racionalnim) razumom, ali pa se motimo? To pa bo impliciralo, da katerikoli znanje, ki ga znanost pozna, ne more biti ultimativno, je lahko le aproksimacija neke realnosti.

Odkar so bili odkriti elektronski računalniki ne beležimo nikakršnega pomembnega napredka v smeri *ultimativnega cilja* – *ustvariti inteligenten stroj* na osnovi algoritmov strojnega učenja. Omenimo lahko nekatere pomembne korake na tej poti:

- Lenatsov avtomatični matematik – zanimiv sistem za odkrivanje novih konceptov v matematiki;
- velik uspeh računalnikov na področju kompleksnih iger, kot so dama, šah, backgammon;
- umetne nevronske mreže za modeliranje kognitivnih procesov v možganih in na podlagi tega;
- Andersonov ACT-R modulni kognitivni model delovanja človeških možganov, ipd.

Osnovna omejitev vseh programskih jezikov in ostalih formalizmov, opisanih zgoraj, katerih osnova je teorija računalništva, velja tudi za katerikoli matematični logični algoritem, neodvisno od tega, kako napreden in kompleksen je. Zelo nedvoumne omejitve veljajo tudi za teorijo učenja. Izpeljano iz računalniške teorije mora biti strojno učenje v obliki algoritma. Tako kot je pričakovati, vse omejitve, ki veljajo za računanje, veljajo tudi za učljivost.

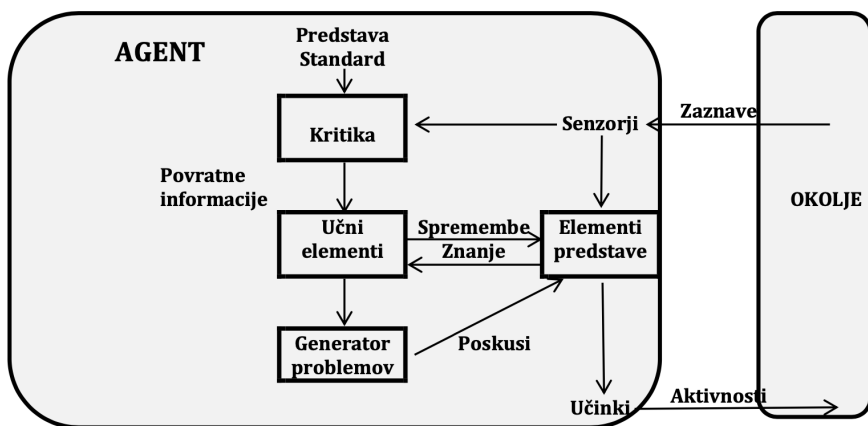
INTELIGENTNI AGENTI

Področje *inteligentnih agentov* je razmeroma mlado, raziskave in razvoj na tem področju pa potekajo v treh smereh:

- v smeri razvijanja teorije o agentih,
- v smeri arhitekture agentov in

- v smeri jezikov, orientiranih na agente (tistih, v katerih so napisani, kakor tudi onih, ki služijo »sporazumevanju med agenti«).

In kaj pravzaprav je agent? Rečeno splošno je agent entiteta, ki za nekoga nekaj počne. Enoznačno definicijo je težko podati, saj je pomen je preprosto preobširen. Definicij je mnogo, vse pa odražajo različnost razumevanja pojma agenta. Pojem agenta v najširšem pomenu prikazujeta dva diametralna koncepta. Eden je sposobnost agenta, da deluje avtonomno, neodvisno od uporabnika, in drugi, da o določenem interesnem področju dela sklepe. Agent je entiteta, ki svojo okolico spremlja s pomočjo senzorjev in nanjo deluje z aktuatorji (efektorji). Avtonomni agenti so sklopi, ki sprejemajo kompleksno dinamično okolico, jo »čutijo« in v njej delujejo avtonomno in tako uresničijo nabor ciljev, katerim so namenjeni. So stalne programske entitete, namenjene za posebno uporabo. Niso podprogrami in se od njih razlikujejo: imajo lastne ideje o izvrševanju nalog, lasten raspored. Od več funkcionalnih aplikacij pa se razlikujejo predvsem po tem, da so mnogo manjši. Delo agentov spremljamo po njihovem delovanju v določeni okolici. Rezultat tega delovanja pa je, da se stanje okolice spreminja. Agente opisujejo in zaznamujejo naslednje lastnosti: učenje, sodelovanje, personalizacija, samostojnost, družabnost, reaktivnost, proaktivnost, mobilnost, racionalnost, dobronamernost, prilagodljivost ipd.



Slika 2.6: Inteligentni agent

Za poimenovanje agentov se lahko uporabljajo tudi druga imena: spletni pajki, plazilci, roboti itd. V tem delu bomo torej uporabljali pojem agent in pod tem pojmom obravnavali objekt, ki samostojno deluje v okolju, se mu je sposoben prilagajati, ima sposobnost zaznavanja stanja, tam kjer je, z delovanjem menja stanja okolice in, kar je zelo pomembno, sposoben se je učiti. Simbolično je inteligentni agent prikazan na sliki 2.6.

(NE)ZMOŽNOSTI UMETNE INTELLIGENCE

Praktično vse raziskovalne metode s področja umetne intelligence skušajo razviti sistem, ki bi se obnašal inteligentno in bi bil sposoben reševati relativno težke probleme. Razvojne metode imajo mnogokrat osnovo v človeškem načinu reševanja problemov. *Dolgoročni cilj, ki si ga je tehnološki svet zastavil, je, da bi računalniška inteligenca (njene sposobnosti) dosegla ali celo presegla človeško inteligenco.* Pomemben vidik razumevanja sposobnosti umetne intelligence je vpliv učenja na inteligenco, hitrost reševanja problemov, osnovne omejitve algoritmov in posnemanje inteligentnega obnašanja:

- *Vpliv učenja na inteligenco:* Z učenjem se sposobnosti sistema povečujejo, zato tudi inteligenca narašča. Človeška inteligenca je dinamična in se v življenju neprestano spreminja, po navadi se večja. Seveda pa moramo pri tem upoštevati tudi različnost intelligence.
- *Hitreje je bolj inteligentno:* Prilagajanje okolju in reševanje problemov sta boljša (bolj učinkovita), če sta hitrejša. Zato je inteligenca znatno povezana z hitrostjo in časom. Vsi testi intelligence so časovno omejeni, tako kot tudi vsi izpiti. Tako lahko zaključimo, da čim hitrejši je računalnik, bolj je inteligenten, vzporedno (paralelno) procesiranje je bolj inteligentno od zaporednega (serijskega) itd.
- *Omejitve intelligence:* Če hipotetično človeka naredimo (degradiramo) ekvivalentnega računalniškemu algoritmu, potem vse omejitve računalniške teorije veljajo tudi za človeka in sposobnosti njegove intelligence. Če predpostavimo, da je človek sposobnejši »stroj« kot (digitalni) računalnik (na primer zvezni in ne diskretni stroj), potem so človeške aktivnosti neopisljive, saj so med drugim povezane tudi s človekovo zunanjo okolico. Posledično iz te predpostavke torej izhaja, da je nemogoče algoritmično izpeljati umetno inteligenten sistem, ki bo v celoti posnemal človeško obnašanje.

- *Posnemanje inteligentnega obnašanja*: Danes so sodobne tehnologije kot filmi, multimedija, računalniki, roboti in virtualna realnost izjemno prepričljive in sugerirajo, da je možno posnemati prav vse in s tem doseči senzacijo realnosti. Zato so stroji, če seveda izključimo zavest, v principu dovolj inteligentni, da ustvarijo občutek umetne inteligence. (Pomislimo samo na velike kapacitete spomina, ki vsebujejo rešitve za vse možne situacije). Če dodamo še izjemne sposobnosti procesiranja (super paralelnost s super hitrimi procesorji) algoritmov za učinkovito iskanje ogromnega števila informacij in algoritmov za strojno učenje sposobnih spletnih izboljšav in to povežemo z ustrezno hevristiko, potem lahko takšne stroje v resnici imenujemo »inteligentne«, saj lahko presegajo ljudi v mnogih, če ne v vseh »praktičnih« nalogah. Vendar moramo ponovno poudariti, da takšni stroji verjetno še vedno nimajo zavesti, torej niso primerljivi z živimi bitji.²⁸

V principu smo sposobni določiti (zaznati ali objektivno izmeriti) katerokoli lastnost sistema, ki ima neko sposobnost učenja in določeno stopnjo inteligence. V nasprotju z učljivostjo in inteligenco je zavest nekaj popolnoma drugega. Potrebno jo je povezovati z osebno izkušnjo in katerikoli objektivni opazovalec je ne more preprosto verificirati.

Možno je objektivno določiti sposobnost učenja, količino usvojenega znanja, sposobnost (inteligenco) prilagajanja okolju in reševanja problemov. Z različnimi testi lahko merimo določene tipe inteligence, dobljeni izidi pa so le bolj ali manj zanesljivi. V nasprotju s tem v principu ni možno preverjati zavesti sistema. Ali je (biološki ali umetni) sistem zavesten ali ne, ve samo sistem sam, seveda če je zavesten. Zunanji opazovalec nima nikakršne možnosti ugotavljati prisotnosti ali odsotnosti zavesti. Posameznik lahko govori o zavesti, če ima sam zavest in če predpostavlja, da ima sistem, ki mu je podoben in o katerem želi govoriti, prav tako zavest. Vsak zavestni sistem lahko posnemamo z nezavestnim sistemom (npr. Turingov test umetne inteligence), da bi iskali (vedno nepopolno) podobnost, zato je vsak objektivni opazovalec zelo lahko ukanjen. (Bregant, 2010, Abramsen, 2008)

²⁸ Bolj podrobno bomo o zavesti govorili v poglavju 8.

2.9 Sklep

V tem poglavju smo se osredotočili predvsem na osnovna filozofska izhodišča filozofije o duhu in se pri tem na temelju zgodovinskega razvoja posvetili novejšim filozofskim pogledom na duha ter se dotaknili vprašanj redukcije in emergenge. Temu je sledil oris modeliranja s poudarkom na kognitivne procese, pri čemer je bilo izpostavljeno, da na področju modeliranja duha obstajajo trije prevladujoči pristopi – modeli, diskretni, zvezni in dinamični oz. povedano z drugimi besedami, simbolni, konekcionistični in dinamični modeli. V poglavju smo tako predstavili nekaj idej s področja filozofije duha, s posebnim poudarkom na naturalističnem stališču, ki ga v nalogi zagovarjamo, in premisi, da je duh lahko »naturaliziran« in to tako, da ga zapišemo z jezikom znanosti, fizike, nevroznanosti ali z jezikom drugih naravoslovnih znanosti. Pri tem od zgoraj omenjenih treh zagovarjamo predvsem dva možna pristopa, hibridni, skratka kombinacijo simbolnih in konekcionističnih pristopov, katerih najpomembnejši predstavnik je ACT-R, in še posebej dinamični pristop, ki se intenzivno razvija v zadnjih letih in predstavlja rdečo nit te naloge.

V poglavju sta izpostavljeni dve diametralni stališči do narave duševnosti, dva sklopa domnev. Po eni strani se nam zdi, da je duševnost povsem drugačna od vsega ostalega v svetu in mora zato imeti poseben ontološki status. Po drugi strani pa sprejemamo, da je človek biološko bitje in je del narave, torej so tudi procesi, ki se dogajajo v njem, vključno z duševnimi procesi, naravni procesi. Pri domnevah prvega pristopa filozofi večinoma izhajajo iz svojih vsakdanjih izkušenj, domneve drugega pristopa pa so vezane predvsem na spoznanja znanosti. Zagovorniki naturalistične teorije (domneve drugega pristopa), med katere sodimo tudi sami, nekakšni znanstveni optimisti, moramo poiskati razlago, kako so ti pojavi mogoči v naravi. Zagovorniki dualistične teorije (domneve prvega pristopa) pa menijo, da ne obstajajo prepričljivi dokazi za naturalizem mentalnega, in zagovarjajo razlage, ki vključujejo nadnaravnost in posebnost duševnega stanja.

Če se osredotočimo na drugi pristop, ki ga v delu tudi zagovarjamo, so lahko razlage emergentne ali redukcionistične. Emergenca označuje v filozofiji in psihologiji fenomen, da določene lastnosti celote ne moremo razložiti z lastnostmi njegovih sestavnih delov. Ideja o emergenci je atraktivna in mikavna na različne načine, zato ni težko razumeti, zakaj je tako popularna. Vendar pa ni enostavno narediti idejo natančno in ji dati resnično vsebino. Dve pomembni zadevi ostajata v emergenci

nerešeni. Prva je, dati emergenci čvrst pozitiven karakter, da lahko vključi supervenienco in nereducibilnost v funkcionalističnem smislu. Druga, prav tako pomembna, pa je soočiti se s problemom vzročnosti od zgoraj navzdol. Nekako morajo emergentisti poiskati razlago, kako lahko imajo pojavne lastnosti različno vzročno moč same po sebi, še posebej, moč vplivati na dogodke in procese na osnovnem nivoju. (Kim, 2006) Pred nami je skratka dilema v obliki odprtega vprašanja – ali imaš emergenco (in s tem mysticizem) in je problem filozofsko zanimiv ali pa imaš neke vrste redukcijo in zato nimaš več filozofskega interesa. Če dinamični kognitivni znanstveniki žele obdržati povezavo z emergenco in filozofijo duha, bodo morali ta pogled združiti tudi z nujno sprejetja trdnih dejstev, povezanih z McGinnovim misterianizmom (McGinn, 1999), kar pa je malo verjetno. Njegova izhodiščna trditev, da je problem duh–telo uganka, ki načelno ni in ne more biti rešljiva, se nam ob današnjem znanstvenem napredku pač ne zdi verjetna. Tak pogled bi ustvaril zagotovo pomembne pregrade na poti interdisciplinarnih raziskav v kognitivni znanosti.

Na drugi strani dinamični kognitivni znanstveniki lahko zavrnejo zvezo z emergenco (v kakršnemkoli pomenu), s čimer dosežejo konsistentnost z obliko interdisciplinarnosti svojega raziskovalnega dela. Seveda je rešitev te dileme – emergenco sprejeti ali zavrniti – odvisna v veliki meri od vrste prihodnjih raziskav v dinamični kognitivni znanosti, saj bo morda prav uporaba »neklasične« dinamike v kognitivni znanosti (npr. upoštevanje teorije kaosa, dinamike, ki je daleč od stabilnosti) obrnila popolnoma novo stran. Jost (2005) na primer poudarja, da v tem primeru:

»Osnovna pravila neklasične dinamike, čeprav so morda enostavna, pa vendar iterirajoča aplikacija, še vedno lahko ustvarjajo asimptotična vedenja, takrat ko s časom limitiramo proti neskončnosti. Tega pa ni tako enostavno napovedati na podlagi dinamičnih pravil samih.« (Jost, 2005: 1)

V tej situaciji, predvsem zato, ker obstajajo pomembni razkoraki, sta dinamični sistem z upoštevanjem nomološke emergence in mysticizma duha in telesa na isti nereduktivni poti. To, kar je Jost v tem primeru želel poudariti je, da

»/.../ ne obstajajo preproste poti za doseg ali napovedovanje končnih izidov, razen da pustimo, da dinamični sistemi to naredijo sami.« (Jost, 2005: 1)

Ta pogled je podoben pogledu Beera (2000), ki meni, da je makrostanje sistema »šibko pojavno«, če ga lahko izpeljemo samo s simulacijo na osnovi kombinacije zakonov sistema in ob upoštevanju ustreznih robnih in začetnih pogojev. Ni pa dorečeno, ali ta princip emergence »izvedi (simuliraj) in vidi« lahko razumemo kot edino epistemološko gledišče. Vendar, odkar se zdi, da bodo neklasični dinamični modeli morda le ustrezali tej definiciji, še obstajajo področja dinamičnih sistemov, ki bodo tudi v prihodnosti ohranila emergentizem v dinamičnih sistemih.

Kot smo že omenili, moramo počakati, da bo razvoj discipline dinamičnih sistemov prišel do stopnje, ko bo lahko sam odgovoril na zastavljeno dilemo, ali pa, da bo razvil in odkril neko popolnoma novo razlago. Delček tega razvoja pa je tudi v tem delu. V sedanjem trenutku bi lahko rekli le to, da smo na razpotju, na bifurkacijski točki, ko smeri razvoja še niso (popolnoma) dorečene. V delu pa bomo poskušali zagovarjati prednosti dinamičnih sistemov in nakazati razvojne možnosti le-teh v prihodnosti.

3 Kognitivna znanost

Bistveni premik v razmišljanju o svetu je doprinesel razvoj znanosti v zadnjih 500 letih, če se pri tem osredotočimo predvsem na zahodno civilizacijo. Človek je odkrival veselje, razmišljal o nastanku sveta in se vse bolj obračal k sebi in vase. Tako v zadnjem času razvoj predvsem kognitivne nevroznanosti napoveduje na tem področju velike spremembe. V nadaljevanju se bomo osredotočili predvsem na to, kako razvoj znanosti – s poudarkom na razvoju kognitivne znanosti – vpliva na posamezne filozofske discipline, predvsem na filozofijo duha.

Kako mislimo, kako se učimo, pomnimo, sanjamo, kako nastane užitek, kje se skrivajo čustva, kako je s svobodno voljo, kako se odločamo, kaj je zavest? Na tovrstna vprašanja skuša odgovoriti kognitivna znanost. Področja človekove duševnosti se loteva interdisciplinarno – s povezovanjem spoznanj vseh disciplin, ki lahko kaj povedo o kognitivnih pojavih. Kognitivna znanost tako združuje nevroznanost, psihologijo, filozofijo, jezikoslovje, umetno inteligenco in družbene vede. Duševne procese poskuša obravnavati celostno in na tak način priti do globljega razumevanja področja, ki nam je izkustveno najbližje. (Markič, Bregant, 2007, Markič, 2010)

V zadnjih desetletjih doživlja kognitivna znanost skokovit razvoj. Še posebej zahvaljujoč napredku nevroznanosti (Morris, Filenz, 2003) počasi začenjamo slutiti, da se je mogoče znanstveno lotiti raziskovanja duševnih pojavov in celo zavesti – področja, ki je bilo še do nedavnega rezervirano samo za misterianiste. (Dennet,

2005, McGinn, 1999). Svoja spoznanja skušajo kognitivni znanstveniki prenesti tudi v prakso – še posebej na področju učenja in poučevanja, procesov sodelovanja in na področju računalniškega strojnega učenja in odločanja.

Kognitivna znanost je zrasla iz gibanja kibernetike v petdesetih letih 20. stoletja in je od takrat doživela številne paradigmatске spremembe. Študije kognitivne znanosti, ki so v zadnjih dvajsetih letih zacvetele po vseh pomembnih svetovnih univerzah, so večkrat obarvane s katero od konstitutivnih disciplin (na primer kognitivna lingvistika, kognitivna nevroznanost, kognitivna antropologija ipd.), vse bolj pa postaja znano, da edino enakopravna obravnava vseh področij zagotavlja celostno obravnavo duševnih procesov. Univerza v Berkeleyu predstavlja strukturo kognitivne znanosti kot kaže slika 1.1.

3.1 Razvoj kognitivne znanost

Kognitivna znanost je, kot rečeno, interdisciplinarno področje raziskovanja duševnih procesov, v katero so kot osrednje discipline vključene filozofija, psihologija, računalništvo, nevroznanost, lingvistika in družbene vede. Vprašanja, na katera skušamo odgovoriti kognitivni znanstveniki, niso nova, saj so si jih že prej zastavljali filozofi in znanstveniki posameznih disciplin. Izbrani kognitivni pojav so raziskovali na ravni, ki je ustrezala njihovemu znanstvenemu področju, in z uporabo svoji disciplini lastnih metod. Ker so ostajali zaprti v svoji disciplini, se je prepogosto dogajalo, da niso znali koristno vključiti znanja iz drugih disciplin. Kognitivna znanost skuša z razvijanjem interdisciplinarnega pristopa premostiti take težave in priti do bolj celostnega vpogleda. V nadaljevanju bomo pokazali, kako je na začetku odločilno vlogo odigrala enotna teorija, ki je temeljila na hipotezi o informacijskem sistemu in računalniški metafori, danes pa je raziskovanje teoretsko mnogo bolj pluralno in zato zahteva nove premisleke o temeljih uspešnega sodelovanja.

V tem poglavju bomo na kratko predstavili razmišljanja in znanstvene pristope, ki so prispevali k nastanku kognitivne znanosti, osnovna teoretska izhodišča in vprašanja, ki jih prinaša interdisciplinarno povezovanje. O tem, kako se vključujejo posamezne discipline in kakšni so njihovi dosežki, pa lahko preberete v nadaljnjih poglavjih.

3.1.1 Rojstvo kognitivne znanosti

Po Georgeu Millerju je rojstvo kognitivne znanosti povezano z 11. septembrom 1956, drugim dnevom simpozija, ki ga je organizirala »Posebna interesna skupina teorije informacij« na Massachusetts Institute of Technology (MIT). V tistem trenutku sicer ni nihče opazil, da se je zgodilo kaj posebnega, zato to »gibanje« tudi ni takoj dobilo svojega imena. To se je zgodilo mnogo kasneje. Tako kot o različnih drugih odkritjih tudi o tem strokovna javnost ni enotna. Tako Stevan Pinker trdi, da je Harvard sedež tega gibanja, sedež kognitivne revolucije, strinja pa se z letnico 1956. Po Howardu Gardnerju pa bi se kognitivna revolucija lahko pričela že leta 1948, ko je bil zgodovinski Hixonov simpozij o cerebralnih mehanizmi vedenja na California Institute of Technology, na katerem so udeleženci že izpostavili teme, ki bodo značilne za kognitivno znanost. Tako je na primer matematik J. von Neumann govoril o primerjavi med elektronskimi računalniki in možgani, W. Mc Culloch je imel predavanje o tem, kako možgani obdelujejo informacije, in postavil vzporednico med živčnim sistemom in logičnimi napravami, psiholog K. Lashley pa je v svojem prispevku izzival behaviorizem. (Jeffress, 1951, Weidman, 2006) Institucionalno se je kognitivna znanost začela uveljavljati konec sedemdesetih let 20. stoletja.

Kljub vsemu za »rojstno leto« kognitivne znanosti večina priznava leto 1956, ko sta na področju psihologije izšli deli *Study of Thinking* (Bruner, Goodnow Austin, 1956) in *The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Information Processing* (Miller, 1956), na področju lingvistike *Three Models of Language* (Chomsky, 1956) in na področju računalništva članek *Logical Theory Machine: A Complex Information Processing System* (Newell, Simon, 1956), v katerem sta Allen Newell in Herbert Simon opisala kompleksen sistem za procesiranje informacij, ki sta ga poimenovala »logični teoretik«. Ta stroj je bil sposoben dokazovati izreke teoremov simbolne logike, kar je pomenilo prve popolne dokaze teorema, ki jih je kdajkoli izvedel računalnik. Ta sistem je bil v nasprotju s klasičnimi sistemskimi algoritmi, ki jih najpogosteje uporabljamo v računalništvu, zasnovan na hevrističnih metodah. Program je bil zapisan v formalnem jeziku, v nekakšnem naravnem psevdokodu, ki je primeren za kodiranje digitalnih računalnikov. V članku je navedena le specifikacija sistema in ne tudi njena aplikacija, konkretna računalniška realizacija. Logična teorija strojev je bila le del obširnega raziskovalnega programa, katerega primarni cilj je bil razumeti kompleksnost informacijskih procesnih

sistemov. Izidi informacijskih izsledkov so objavljeni v njunih kasnejših delih (npr. Newell, Simon, 1972). Ta članek je bil predstavljen na že omenjenem simpoziju o informacijski teoriji na Institute of Radio Engineers, MIT leta 1956.

Prav gotovo je zelo vplivalo na razvoj kognitivne znanosti delo britanskega matematika in logika Alana Turinga o formalnih logičnih sistemih. Za ponazoritev svojih izsledkov je razvil pojem preprostega stroja, ki ga danes imenujemo Turingov stroj, in pokazal, da lahko s takim preprostim strojem izvedemo vsako nalogo, za katero lahko natančno navedemo korake, ki so potrebni za izpolnitev naloge. (Turing, 1936) Turingove ideje so porodile misel o stroju, ki bi samo s svojo formalno strukturo posnemal delovanje duha. Razvoj elektronskih digitalnih računalnikov, univerzalnih Turingovih strojev, ki so veliko hitrejši in jih lažje programiramo, je kasneje omogočil izdelavo kognitivnih modelov in njihovo eksperimentalno preverjanje.

K razvoju kognitivnega modeliranja sta pomembno prispevala tudi kibernetika McCulloch in Pitts, ki sta leta 1943 pokazala, da lahko (idealizirane) nevrone in njihove povezave modeliramo s pojmi logike in da je možno vse, kar lahko izčrpno in nedvoumno opišemo, realizirati z ustrezno končno nevronske mreže. (McCulloch, Pitts, 1943) Njuno delo predstavlja izhodišče za dva pristopa, ki sta kasneje zaznamovala kognitivno znanost – eden vodi k nevronske mrežam, sestavljenim iz velikega števila med seboj povezanih nevronov, drugi pa k simbolnim modelom klasične kognitivne znanosti.

Porast kognitivizma v psihologiji, ki je bil do leta 1970 samostojni del behaviorizma, je bil imenovan *Kuhnova revolucija*. (Baars, 1997) Izhajajoč iz Kuhnove (1978) terminologije ponuja kognitivizem naslednje:

- svojstvene raziskovalne strategije,
- svojstvene pristope za konstruiranje psiholoških fenomenov.

Kognitivizem izhaja iz behaviorizma, spremembe pa so bile del širše kognitivne revolucije, ki ni le spremenila vrste disciplin, kot npr. kognitivno in razvojno psihologijo in nevroznanost, temveč je tudi vodila k meddisciplinarnim

raziskovalnim skupinam, znanim pod skupnim imenom – *kognitivna znanost*. (Bechtel, Abrahamsen, 2002) Področja, so naslednja:

- sklepanje,
- pomnjenje,
- jezik.

Ta področja pa so razširjena z zaznavami in motorično kontrolo. Z razvojem kognitivne paradigme postaja ideja, da kognicija vsebuje tudi manipulacijo simbolov, vse bolj osrednja. Simboli, ki označujejo zunanje fenomene, imajo seveda svoje pomene, *semantiko*. To so trajne entitete, ki jih lahko shranimo in ponovno prikličemo iz spomina ter jih lahko spreminjamo po določenih pravilih. Pravila, ki določajo, kako simbole sestavljamo, *sintaksa*, in kako naj jih transformiramo, so tista, ki usmerjajo kognitivne lastnosti. Leta 1980 se pojavi alternativni način razumevanja kognicije, nova Kuhnova paradigma – *konekcionizem*, *paralelno procesiranje* (PDP) ali *nevronske mrežne modeli*. Temeljni deli tega procesa sta dve knjigi avtorjev Rumelharta in McClellanda: *Parallel Distributed Processing – del 1 in 2* (1986).

Ta kognitivna revolucija (Gardner, 2011) je nov pristop v proučevanju duševnosti in ne bi bila možna brez nekaterih teoretskih prispevkov. Ključnega pomena so bili dosežki matematike in logike (Turing), razvoj informacijske teorije (Shannon), kibernetika (Wiener), odkritje nevronov (Ramón y Cajal) in njihovo modeliranje (McCulloch in Pitts) ter raziskovanje kognitivnih motenj kot posledic poškodb možganov. (Luria, 1987)

Temeljna opredelitev kognitivne znanosti – raziskovanje o tem, kako človek procesira informacije, je tako široka, da dopušča različna tolmačenja in pristope h kognitivni arhitekturi. Ob računalniški metafori, ki je odprla pot simbolnemu modeliranju in je zaznamovala začetek kognitivne znanosti (zato tudi naziv klasična kognitivna znanost), so danes prisotni še drugi pristopi, ki jih lahko pojmujeemo kot alternative in/ali dopolnitve.

3.2 Kognitivna teorija učenja

Kognitivni psihologi so iskali meje behaviorizma na primeru opazovanja vedenja. Pri tem so vključili mentalno strukturo in proces v teorijo učenja. Notranje predstave ljudi se lahko odražajo na zunanji realnosti, ki predpostavlja stanje objektivizma, to je, da sta duh in telo ločena, skratka, da duh deluje neodvisno od telesa. Ta princip lahko sedaj prenašamo od iz realnosti zunaj duha nazaj v sam duh. To pozicijo sta dokaj dobro obdelala Wilson in Meyer. (Wilson, et al., 2000)

Osnovno izhodišče, ki zanima kognitivne psihologe, je notranji mehanizem človeškega razmišljanja in procesa pridobivanja znanja. Želijo poiskati odgovore na to, kaj in kako človek pomni, ter najti mentalni proces, povezan s tem, kako delujeta integracija in priklic informacij (Kordigel Aberšek, Aberšek, 2021).

Začetek psihologije kot znanosti sega v leto 1879 in bi ga lahko pripisali Wilhelmu Wundtu, ki je ustanovil prvi psihološki laboratorij v Leipzigu. Razvil in uporabljal je metodo imenovano *introspekcija*,²⁹ ki je kognitivna metoda za raziskovanje tega, kako deluje duh. Osnovne ideje in predpostavke kognitivne psihologije izhajajo iz zgodnjih desetletij 20. stoletja in jih predstavljajo gestalt psihologija, Edward Tolmansovo kognitivno učenje in Jean Piagetova kognitivna razvojna teorija. J. Anderson (1885) je definiral tri osnovne vplive na razvoj sodobne kognitivne psihologije:

1. model procesiranja informacij;
2. umetna inteligenca; Allen Newell in Herbert Simon sta v kognitivno psihologijo vpeljala koncept računalniških znanosti, in s tem pripomogla k razvoju teorije psihologije (Newell, Simon, 1972);
3. lingvistika; Noam Chomsky trdi, da mora učenje jezikov vsebovati notranji konstrukt. Teorija, ki upošteva le opazovanje dražljaja pri jezikovnem učenju, ni dovolj. (Chomsky, 1956)

²⁹ Introspekcija je opazovanje lastnih čustev, hotenj ali misli, samoopazovanje; skratka opazovanje lastnega notranjega oz. psihičnega sveta.

3.2.1 Model procesiranja informacij

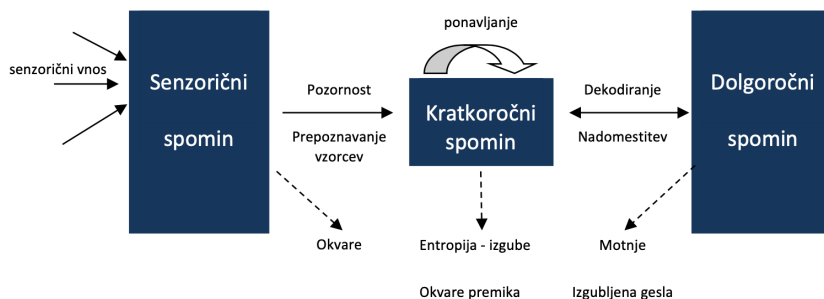
Osredotočimo se na prvi vpliv, to je model procesiranja informacij, ki je za razumevanje raziskovalnega področja ključen. V nadaljevanju se bomo osredotočili na umetno inteligenco. Prednost današnjih sodobnih računalnikov je, da omogočajo bogato teoretično metaforo za snovanje teorij o informacijskih procesih, predvsem pri ljudeh kot najbolj razvitih bitjih. Današnja arhitektura informacijskih procesov v računalnikih močno uokvirja zgodnja razmišljanja sodobnih kognitivnih psihologov, ki so vložili ogromno truda v razvoj teorij o nadzoru procesa informacij. (Barsalou, 1992) Raziskave kognitivnih teoretikov o tem, kakšen je človekov informacijski proces, kako zbira informacije iz svoje okolice, kako shranjuje te informacije v svoj spomin, kako jih ponovno prikliče iz spomina in jih na koncu vrača v okolico, še vedno potekajo izredno intenzivno. Tako kot računalnik tudi človek pridobiva informacije iz svoje okolice. Oboji, ljudje in računalniki, informacije priključijo iz spomina, ko jih v konkretnih okoliščinah ob reševanju konkretnih nalog potrebujejo. Oboji so omejeni s količino, ki jo lahko procesirajo v določenem času, oboji informacije obdelujejo in predelujejo tako, da lahko ustvarijo nove informacije, in ne nazadnje, oboji te informacije vračajo v okolje. (Broadbent, 1958, Newell, Simon, 1972)

Ozrimo se k začetkom. Broadbent je že leta 1958 predlagal splošen model človekovega sistema procesiranja informacij, prikazanega na sliki 3.1. Ta model predstavlja naslednje osnovne mehanizme:

- tri glavne spominske shrambe, senzorični spomin, kratkoročni spomin (začasni ali delovni spomin) in dolgoročni spomin, kjer se izvajajo operacije z informacijami;
- mehanizem procesa transformiranja informacij od vnosa do iznosa informacij v/iz vsakega dela spomina in med vnosom in iznosom informacij celotnega sistema, t. j. pozornost/prepoznavanje vzorca, dekodiranje in priključ v spomin;
- reprezentacija in shranjevanje informacij, t. j. mrežni model (Collins, Quillion, 1972), modeli primerjave lastnosti (angl. Feature Comparison Model) (Smith, Shoben, Rips, 1974), propozicijski modeli³⁰ (Anderson, 1983), modeli za

³⁰ Propozicijski modeli so računalniški psiholingvistični modeli za prepoznavanje pisne oblike jezika.

vzporedno obdelavo³¹ (Rumelhard, McClelland, in PDP raziskovalna skupina, 1986), konkurenčno kodirani modeli.³² (Pavivio, 1986)



Slika 3.1: Model informacijskega procesa

Broadbent je predlagal model, kjer je procesiranje informacij fiksno, periodičen prenos informacij iz enega dela spomina v drugega in kjer kontrola sistema predstavlja informacijsko povratno zanko od naslednje procesne stopnje k predhodnim procesnim stopnjam. Ta model so kasnejši avtorji dopolnjevali in spreminjali, vendar je v izhodišču še vedno pravičen.

Najširša sprejeta teorija danes je teorija predstav (angl. stage theory), ki temelji na delih Atkinsona in Shrifina iz leta 1986. (Atkinson, 1998) Ta model predpostavlja, da so možgani utelešenje živčnega sistema, ki procesira informacije od trenutka pridobitve (vnosa) do trenutka, ko jih shranimo v dolgoročni spomin. Sistem sestoji iz treh glavnih predstav, ki vsebujejo različne psihološke lastnosti:

- *senzorični spomin*, ki predstavlja zaznavno enoto, torej čutila,
- *kratkoročni spomin* in
- *dolgoročni spomin*.

³¹ Modeli za vzporedno obdelavo so modeli vzporedne obdelave (angl. parallel processing), ki je računalniška tehnologija, ki omogoča sočasno izvajanje več izračunov. Vzporedna obdelava, ki izračun razbije na manjše dele in namesto enega samega drugega za drugim izvede tisoče teh manjših izračunov hkrati, obeta ogromen napredek pri delovnih hitrostih izvajanja določenih ponavljajočih se opravil.

³² Konkurenčno kodirani modeli (angl. Dual Coding models) Teorija konkurenčnega kodiranja je kognitivna teorija, ki predvideva, da se vizualne in verbalne informacije procesirajo različno, po različnih kanalih in v človeškem duhu ustvarjajo različne predstave o informacijah. Na osnovi teh vhodnih informacij ustvarjamo znanje, ki ga lahko sprejemamo in pomnimo.

Zaznavne enote zgoščeno shranijo predstave o zunanjih dražljajih iz okolja vse do takrat, ko te informacije posredujemo naprej. Za vsak občutek obstajajo različne zaznavne enote, različna čutila. V teh enotah se shranjujejo informacije zelo kratek čas, predpostavlja se, da se izgubijo, takoj ko jih shranimo v kratkoročni spomin.

Kratkoročni spomin lahko razumemo kot zavestni spomin tudi zato, ker ne samo, da ohranja informacije, temveč ker tudi omogoča manipuliranje z informacijami, njihovo interpretacijo in transformacijo. V odvisnosti od nadaljnjih procesov, se lahko nove informacije v kratkoročnem spominu pretvorijo in tvorijo del dolgoročnega spomina.

Dolgoročni spomin je relativno neomejena in trajna shramba informacij. Informacije se v njem shranjujejo za kasnejšo uporabo.

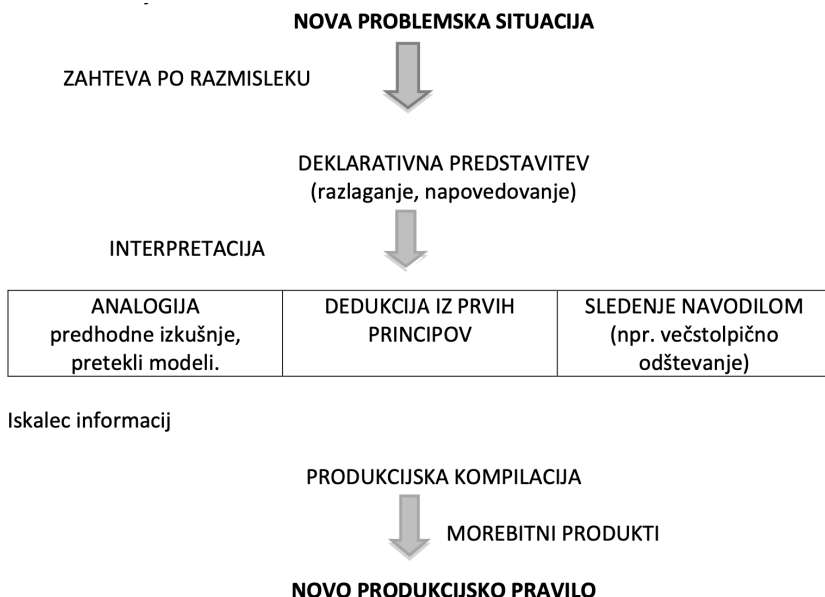
Model procesiranja informacij poudarja osnovne mehanizme in oblike predstav in procesov ter predstavlja model hranjenja informacij.

3.2.2 Izvajanje kognitivnega procesa

Izvajanje kognitivnega procesa, predvsem pri učenju, si lahko predstavljamo v obliki različnih delov možganov, ki jih po Andersonu (2007) imenujmo *moduli*. *Deklarativni modul* priključuje kritične (predvidoma potrebne) informacije in *proceduralni modul* izvede potrebne korake in povezave za doseganje nekega cilja, in sicer:

- zunanje povezave, zunanje informacije, ki jih dobimo npr. z manualnim, vizualnim ipd. modulom;
- notranje obdelave zunanjih informacij, ki se izvaja v predstavitvenem modulu.

Osnovna ideja je, da sistem sam pride do novih spoznanj, za katera še nima izdelanih pravil. To zahteva pridobitev deklarativne informacije in premislek o tem, kaj želimo dobiti. Shematsko je postopek prikazan na sliki 3.2.



Slika 3.2: Učenje

3.2.3 Pridobivanje novih spoznanj (kaj je učenje?)

Učenje označuje spremembe v sistemu, ki so adaptivne glede omogočanja sistema, da po procesu učenja bolj učinkovito izvaja identične naloge z isto populacijo.

(H. A. Simone)

Učenje bi lahko definirali s pomočjo naslednje splošne okoliščine: imamo sistem (»učenca«) in učitelja, ki izvaja (želi izvajati) določeno nalogo. Na začetku je učinek slab. Z vajo, ki je sestavljena iz posnemanja učitelja in učenja po principu napak, postane predstava vedno boljša. »Boljša« lahko pomeni hitrejša, bolj natančna, cenejša itd., odvisno od naloge. Vajo, posnemanje učitelja in ponavljanje z odpravljanjem napak, *imenujemo učenje*.

Učenec se nauči, kako opraviti določeno nalogo, in jo lahko ponovi brez ponovnega učenja. Da lahko to nalogo uspešno ponovi, jo mora transformirati. Proces transformacije zaradi učenja imenujemo *sproščanje znanja* (angl. *knowledge acquisition*).

Znanje definiramo kot interpretacijo informacij. Znanje lahko pridobimo z učenjem (pretežno faktografsko znanje) ali je lahko rezultat izkušenj (učimo se na podlagi lastnih poskusov). Lahko je napačno ali pravilno, popolno ali nepopolno itd. Vsak podatek z razlago lahko razumemo kot znanje, čeprav je v praksi zanimivo samo uporabno znanje, t. j. znanje, ki omogoča, da sistem bolje rešuje naloge z določenega problemskega področja.

Učenje je prisotno pri vseh živih bitjih, najbolj očitno pa pri ljudeh. Učenje pri živih bitjih imenujemo *naravno učenje*, če pa je učenec stroj – računalnik, to imenujemo *strojno učenje*. Namen razvijanja metod strojnega učenja je, ob boljšem razumevanju naravnega učenja in inteligence, izdelati algoritem za reševanje problema, ki zahteva posebna znanja. Mnogokrat so takšna znanja še neznanja, ali znana le ozkemu krogu strokovnjakov. Pod določenimi predpogoji, z uporabo strojnih algoritmov učenja, lahko učinkovito generiramo takšno znanje za reševanje problemov. Celo celotno evolucijo v naravi bi lahko označili kot učenje. Z *genetskim križanjem*,³³ *mutacijo*³⁴ in *reprodukcijo*³⁵ narava ustvarja vedno boljše sisteme, sposobne prilagajanja najrazličnejšim okoljem.

Evolucijske principe lahko uporabljamo tudi pri strojnem učenju in s tem usmerjamo raziskave in prostor hipotez v t. i. *genetske algoritme*.³⁶

NARAVNO UČENJE

Ljudje se učimo vse življenje. Praktično se učimo vsak dan, kar pomeni, da se naše znanje ves čas spreminja, širi in izboljšuje. Učijo se tudi živali. Sposobnost učenja je odvisna od evolucijske stopnje posameznega bitja.

³³ Genetsko križanje se v računalništvu uporablja v genetskih algoritmi in ponazarja sistem naravnega križanja vrst na matematičen način.

³⁴ Mutacije so trajne dedne spremembe genoma, ki jih povzročijo različni dejavniki (mutageni), ki so lahko kemijski (vplivi spojin in elementov, npr. azbest), fizikalni (različna sevanja), biološki (virusi). Njihovo bistvo je sprememba zgradbe in količine DNK. Mutacija je v genetskih algoritmi zelo uporabna zato, da se ne ustavimo v lokalnem optimumu problema. Če so kromosomi zapisani v dvojiškem nizu, to pomeni, da se pri mutaciji naključni gen spremeni iz 0 v 1 (ali iz 1 v 0).

³⁵ Reprodukcijski ali razmnoževanje je ustvarjanje kopije ali menjava podobnih stvari, npr. fotokopiraje ali ustvarjanje duplikata. Najbolj znana je biološka reprodukcija. V UI pa so razviti matematični modeli, s katerimi simuliramo naravne procese.

³⁶ Genetski algoritmi so računalniški model naravne evolucije, ki temelji na križanju in mutaciji. Uporabljajo se za reševanje problemov, pri katerih rešitev opišemo v obliki kodiranega zaporedja, imenovanega kromosom.

Raziskave razlag naravnega učenja je področje *psihologije učenja in psihologije izobraževanja*. Psihologija učenja upošteva pozornost, utrujenost, stopnjo pozabljanja in motivacijo, ki so vitalnega pomena za uspešen proces izobraževanja, upošteva povezave učitelj – učenec in predlaga najrazličnejše motivacijske in nagrajevalne strategije. Vse to je izjemnega pomena za človeško učenje, mnogo manj pa za strojno učenje.

Raziskovalci skušajo definirati tipe naravnega učenja in najti osnovni tip, iz katerega bi lahko izpeljali vse ostale. Večinoma so tipi učenja klasificirani v povezavi z učno kompleksnostjo, učnim gradivom in učno strategijo.

Z upoštevanjem kompleksnosti učnega procesa ločimo naslednje tipe učenja (Aberšek, 2012):

- vtiskovanje v spomin,³⁷
- pogojevanje in združevanje,
- verjetnostno učenje,
- memoriranje,
- učenje s poskusi in napakami,
- posnemanje,
- učenje z razumevanjem in vpogledom.

V povezavi z učnim gradivom definiramo naslednje učne tipe:

- senzorično učenje,
- učenje motoričnih spretnosti,
- semantično (tekstualno) učenje.

3.3 Učenje in kognitivno modeliranje

Vsi spomini, vsa naša preteklost, znanja in izkušnje nam dajejo identiteto. Mnogokrat je videti, da smo to, česar se lahko spomnimo o svoji preteklosti. Deklarativni spomin je kritičen za naša prilagajanja spremembam v svetu. Vendar samo to ni

³⁷ Vtiskovanje (angl. imprinting) je termin, ki se v psihologiji in etologiji uporablja za opis kakršnekoli oblike učenja, ki se dogodi pri določenih letih ali določenem življenjskem obdobju, ki je hiter in navidezno neodvisen od posledic vedenja.

dovolj, je le potencial, ki ga lahko izkoriščamo. Skratka, moramo ga znati tudi uporabljati. Uporaba znanja do njegovega celotnega potenciala zahteva kompleksen, preudaren proces, v katerem delamo sklepe, povzetke in napovedi na osnovi teh znanj.

Iz dosedanjih odkritij je očitno, da obstajajo različne vrste učenja. V grobem bi lahko vse vrste učenja razporedili v 4 skupine, odvisno od tega ali uporabljajo deklarativni ali proceduralni spomin kot tudi od tega, ali učenje vsebuje ustvarjanje novih simbolnih struktur ali samo vstavljanje subsimbolov v že obstoječe strukture. Preglednica 3.1 (Anderson, 2007: 92) prikazuje 2 x 2 klasifikacijo teh različnih tipov učenja.

Tabela 3.1: Taksonomija učenja

	Deklarativni	Proceduralni
Simbolno	Učenje dejstev	Usvajanje veščin
Podsimbolno	Utrjevanje	Pogojevanje

1. *Učenje dejstev:* V deklarativnem spominu lahko ustvarimo nove spomine. To je spomin v ožjem pomenu besede, to je to, kar večina ljudi razume kot spomin. To je edini način učenja, katerega posledica so novi zavestni spomini.
2. *Utrjevanje:* Ob ustvarjanju novih deklarativnih spominov lahko delamo tudi na tem, da postanejo ti spomini bolj dostopni. Ta proces učenja imenujemo utrjevanje znanja.
3. *Usvajanje veščin:* Različne oblike učenja povzročajo nove postopke dela (nova produkcijska pravila). Usvajanje veščin nas privede do stopnje rutine, ko posamezne naloge izvajamo nezavedno, torej ne razmišljamo, kako bomo aktivnost izvedli. Lep primer je slepo tipkanje, vožnja avtomobila ipd.
4. *Pogojevanje (angl. conditioning):* Z izkušnjami se naučimo, da so posamezne aktivnosti bolj učinkovite v določenih okoliščinah. Pogojevanje je splošno razumljeno kot najbolj razširjen postopek učenja npr. Pavlov eksperiment pogojevanja – učenje psa.

3.3.1 Informacijsko-teoretična ali kibernetična didaktika

Osrednji protagonist in začetnik ideje kibernetične didaktike je bil Felix Cube (1982). Ne uvršča je med samostojne didaktične smeri, ampak jo razume kot posebno metodo in del še slabo opredeljene »kritično racionalne vzgojne znanosti«.

Kibernetična didaktika temelji na kibernetičnih načelih in spoznanjih o krmiljenju in reguliranju tehničnih, organskih in družbenih sistemov. Med slednje uvrščamo tudi pedagoški sistem. (Strmčnik, 2001) Običajno je kibernetična didaktika definirana kot znanost, ki raziskuje, kako s tehnološkimi in matematičnimi pripomočki spodbuditi in voditi učne procese oziroma ravnanje nekega učnega sistema, da bi bili učni cilji optimalno doseženi. Cube jo opredeljuje kot »znanost o načelnih možnostih poseganja in konstruiranja na področju človekovega učenja«. Težišče je na vodenju, medtem ko se učnih vsebin in medijev ta didaktični model dotika le površinsko, učne cilje pa povsem obide, češ da niso stvar didaktike.

»Regulacijski krog« je osrednji pojem kibernetike za vodenje in reguliranje kibernetičnih sistemov. Prenesen na didaktično področje obsega regulacijski krog naslednje: učno načrtovanje (»regulator«), ki na podlagi primerjave med učno »je vrednostjo« in »naj bi vrednostjo« pripravi učni program, učno vodenje (»nastavitveni člen«), ki uresničuje program (učitelj, programirani učbenik, računalniški program ali drugi mediji), učenec (»regulacijska količina«), ki se prek učne kontrole (»merilno tipalo«) spreminja v skladu z učnimi cilji (»zadana vrednost«). (Strmčnik, 2001)

Učni proces je pojmovan kot posebna oblika tehničnega vodenja. V rabi je kibernetična terminologija: ne govori o učitelju in učencu, ampak o *predajniku (učitelju)* in *sprejemniku (učencu)* ali poučevalnem in učnem sistemu, za učne vsebine je v rabi izraz informacije, za učno vodenje, krmarjenje in reguliranje ravnanja itn. Kibernetična didaktika se naslanja na ugotovitve temeljnih kibernetičnih disciplin, predvsem na teorije sistema, informacije in komunikacije ter algoritmov. (Strmčnik, 1978)

Teorija sistemov se vključuje v kibernetično didaktiko tako, da pojmuje učni proces kot svojevrsten, dinamičen in odprt sistem. Sistem je celota, ki jo sestavljajo povezani deli, odnosi med njimi in struktura. Konekcijem (nevronske mreže, paralelno distribuirano procesiranje) so navdahnili spoznanja nevroznanosti o možganih in je izstopal iz tradicionalnih paradig predvsem po tem, da ni gradil kognicije z vpeljavo manipuliranja s simbolnimi paradigmami, temveč je ponudil radikalno drugačen koncept osnovnega procesiranja sistema *dub – možgani*. Osnovna ideja je, da obstaja mreža osnovnih *enot* ali vozlišč, od katerih ima vsaka neko stopnjo aktivnosti. Te enote so med seboj *povezane* tako, da aktivna enota spodbuja ostale.

Mreže ponazarjajo sistem delovanja naših možganov in z njimi povezanega duha, skupaj pa ponazarjata v našem primeru proces učenja. Poudariti moramo, da je ta proces dinamičen proces. Mreža je torej *dinamični sistem*, ki, ko dobi impulz na vходу, razširja dražljaje in zaviranja med svojimi enotami. Konekcionistični modeli, mreže, vključujejo pomembne značilnosti možganske arhitekture. Da bi lahko ocenili karakter in vrednost mrež, jih moramo opazovati med delom, med reševanjem predvsem kompleksnih problemov. (Bechtel, Abrahamsen, 2002, Bechtel, 2008) Osnovna značilnost modelov je v tem, da so sestavljeni iz preprostih enot (idealizirani nevroni), ki so medsebojno povezane. Mreža se v procesu učenja izbrane kognitivne naloge uči tako, da na osnovi učnega pravila (algoritma) spreminja moč povezav med enotami. Izbira arhitekture mreže in učnega algoritma je odvisna od tega, kakšno kognitivno nalogo naj bi z mrežo modelirali, oziroma kako nevrološko verodostojen naj bi bil model. Reprezentacije niso več statična, enostavna in z jezikom opisljiva notranja stanja. Ker je to ena od najbolj hitro razvijajočih se multidisciplinarnih ved zadnjih desetletij, bi samo za opis osnovnih izhodišč potrebovali celotno knjigo. Zato tu to le omenjamo, v poglavju o konekcionizmu pa bomo o tem povedali več. (Bechtel, Abrahamsen, 2002)

Za kibernetiko so tako tudi učitelj in učenci, učna vsebina in organizacija pouka le podsystemi celotnega vzgojno-izobraževalnega sistema. *Učitelj* je podsystem v funkciji *predajnika* in vodenja, ki informacije vsebuje, načrtuje in posreduje. *Učenec* pa je v funkciji *vodenege*, ki informacije sprejema, predeluje, skladišči in se nanje odziva oz. se po njih ravna. V tem pomenu pomeni učenje smotrno interakcijo, v kateri učenci in učitelj količinsko in kakovostno spreminjajo svoje lastnosti in ravnanja (adaptivnost sistema). Učenje označuje proces, katerega posledica so adaptivne spremembe v sistemu. Spremembe, ki so posledica procesa učenja, omogočajo, da *ista populacija* rešuje *iste naloge* hitreje in uspešneje kot pred procesom učenja. (Aberšek, 2012)

Na didaktičnem področju sta se najbolj uveljavili *teorija informacij* in *teorija komunikacij*. Pojem informacije je zelo široko zamišljen, saj pomeni vsakršno sporočilo od najpreprostejših do najzahtevnejših stvarnih in logičnih odnosov. L. N. Landa (1974) je definiral informacijo kot vse tisto »kar omejuje nedoločenost«, kar vnaša v nek sistem urejenost oziroma zmanjšuje neurejenost (entropijo). Po tej teoriji je tudi učenje preseganje entropije, ustvarjanje »reda« v glavah učencev, zmanjševanje potrebe po informiranju, prehajanje k določenemu (determiniranemu) ravnanju.

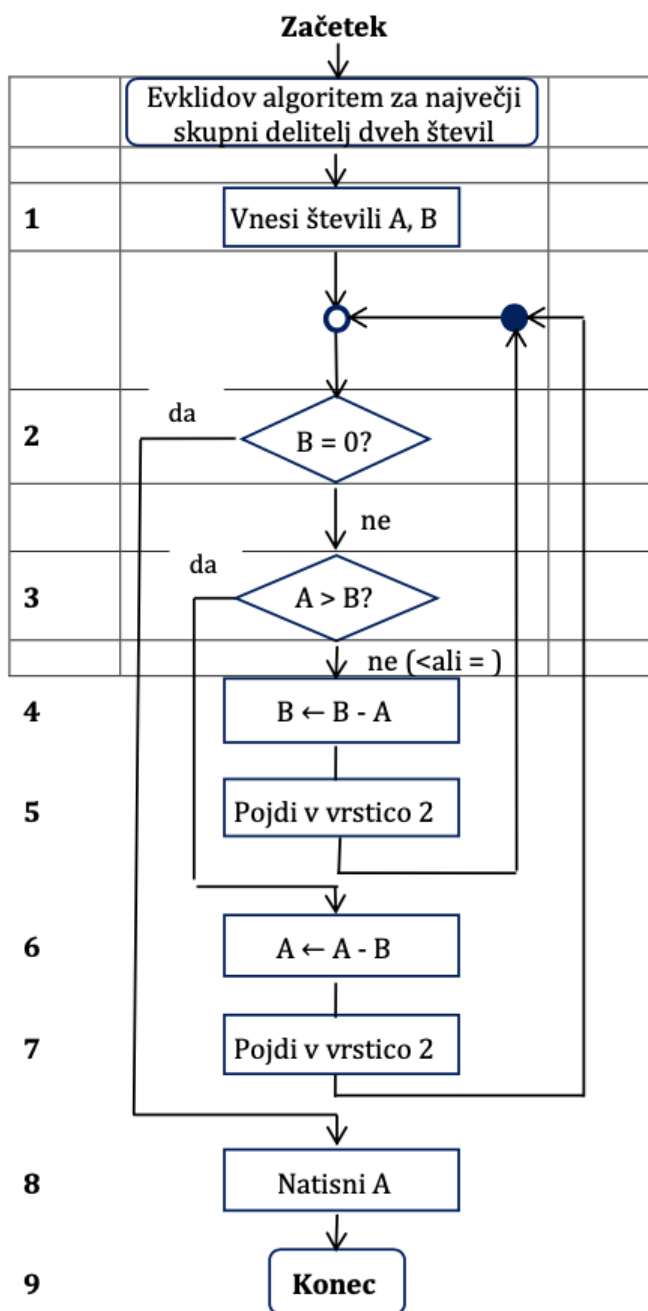
V didaktiki je našla plodna tla tudi *redundantna teorija*. Kot informacijska disciplina proučuje informativno moč sporočil, razmerje med sporočilnim (novim) in odvečnim (že znanim) delom sporočila. Ukvarja se s kvantitativno zmogljivostjo in pretočnostjo informacijskih kanalov, čutnih organov, volumna spomina, z oblikovanjem super simbolov, s sporočilno vrednostjo posameznih povednih členov, glasov, s pomenom transinformacije ter interference sporočil itn.

Teorijo informacij dopolnjuje *teorija komunikacij*. Na njeni podlagi se je razvila celo samostojna »*komunikacijska didaktika*«. Pouk definira kot komunikacijsko delovanje kot socialno interakcijo v duhu demokratične emancipatorne vzgoje in socialnega učenja »homo communicatorja«, pri čemer tudi intrapersonalna komunikacija oziroma komunikacija človeka s samim seboj in v sebi ni zanemarjena.

Sistem komuniciranja vsebuje naslednje člene, povezane v komunikacijsko verigo: izbor informacij, predajnik informacij (učitelj ali drugi mediji), kanali zvez z motnjami (periferno in centralno živčevje), sprejemnik informacij (učenci), kod (npr. z učenci sinhroniziran jezik) in cilji komuniciranja. Predajnik in sprejemnik sta povezana s sistemom primarnih in povratnih zvez.

Teorija algoritmov se ukvarja z vodenjem »avtomatiziranega« učenja. Na osnovi algoritmov se je razvil programirani pouk. Ker ne obstaja splošno sprejeta formalna definicija, kaj je to algoritem, lahko uporabimo informacijsko definicijo, da je algoritem niz pravil, ki *natančno* definirajo vrsto operacij, da je vsako končno *natančno* zaporedje napotkov oziroma inštrukcij, ki rešujejo neskončno mnogo problemov iste kategorije.

Osnovni primer algoritma je Evklidov algoritem, s katerim določamo največji skupni delitelj dveh števil ali polinomov in se glasi: če imamo naravni števili A in B , predpostavimo, da je A večji ali enak B . Če je B enak nič, potem je A rezultat postopka. Sicer pa nadaljujemo postopek s številom B in celoštevilskim ostankom deljenja A z B . Ta algoritem je prikazan kot diagram poteka na sliki 3.3. V podrobnosti se ne bomo spuščali, saj bi za to potrebovali preveč časa in prostora.



Slika 3.3: Primer Evklidovega algoritma, zapisanega z diagramom poteka

Z algoritmi na didaktičnem področju se je veliko ukvarjal že omenjeni znanstvenik Landa. Menil je, da je z algoritmično metodo tudi v šoli mogoče razvijati splošno (metodično) mišljenje učencev in sposobnosti za reševanje istovrstnih problemov, vendar le, če so izpolnjeni naslednji temeljni pogoji:

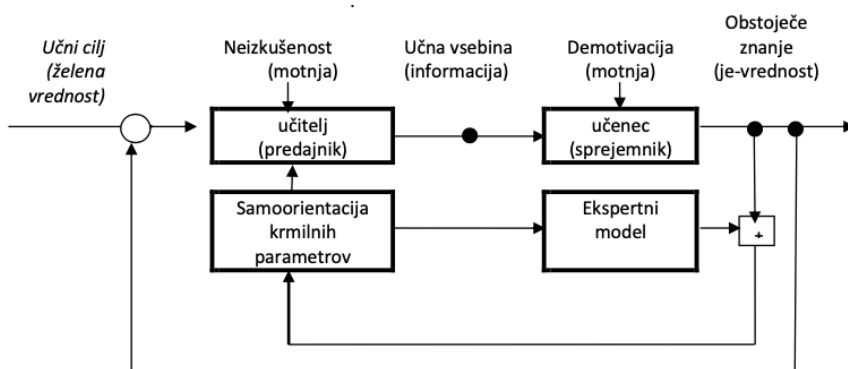
- razumljivost algoritmičnih navodil, kadar je vgrajeno vanje potrebno število nedvoumno določenih postopkov;
- z njimi mora biti možno reševati večje število enakovrstnih ali sorodnih problemov;
- zagotovljena mora biti uspešnost, če so vodila pravilno rabljena.

Hkrati je Landa svaril pred preveliko uporabo kateregakoli pouka z algoritmi. Za manj strogo sklenjene in bolj odprte učne vsebine je bolj priporočal hevristične algoritme, ki zaradi manjše določenosti puščajo več učne svobode.

Pomembna vrednost algoritmov je v zahtevi po »algoritmičnem opisu«, po *natančnem definiranju* zaporedja napotkov, kar je predpogoj za algoritmično učenje. Če želimo spremeniti nek didaktični položaj, ga moramo najprej natančno opisati, odkriti njegovo strukturo, zakonitosti, vzročno posledične in druge odnose ipd., vendar ne le v tem položaju, marveč tudi procese učenja v glavah učencev. Šele na tej podlagi je mogoče zgraditi operativne postopke (»algoritmična navodila«), ki usmerjajo procese spreminjanja tega položaja (»algoritmični proces«). Za to lahko uporabimo različna »orodja«, tudi prej omenjeno faktorsko in strukturno analizo.

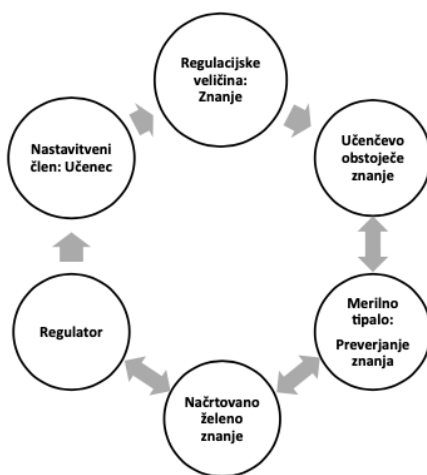
UČNI PROCES V KIBERNETIČNI DIDAKTIKI

Učni proces (slika 3.4, 3.5) razlaga kibernetična didaktika kot regulacijski krog, ki je sklenjen šele takrat, ko je doseženo znanje (stanje) ob koncu procesa (je-vrednost) enako načrtovanemu, želenemu znanju (stanju). Kibernetična teorija poučevanja se izogiba natančnemu načrtovanju ciljev. Pozornost usmerja k učnemu procesu, krmiljenju in optimiranju tega procesa, pri čemer posveča posebno pozornost odstranjevanju motenj.



Slika 3.4: Kibernetična didaktika: učni proces

Kritiki opozarjajo na težave, ki nastajajo, ko mora ekspertni model pri svojem načrtovanju predvideti idealnega učenca z njegovimi kognitivnimi in emocionalnimi (afektivnimi) lastnostmi. Računalniški program zelo težko načrtuje takega hipotetičnega učenca. Na tej točki se kaže bistvena razlika med računalniškim poučevanjem in poukom, ki ga vodi učitelj.

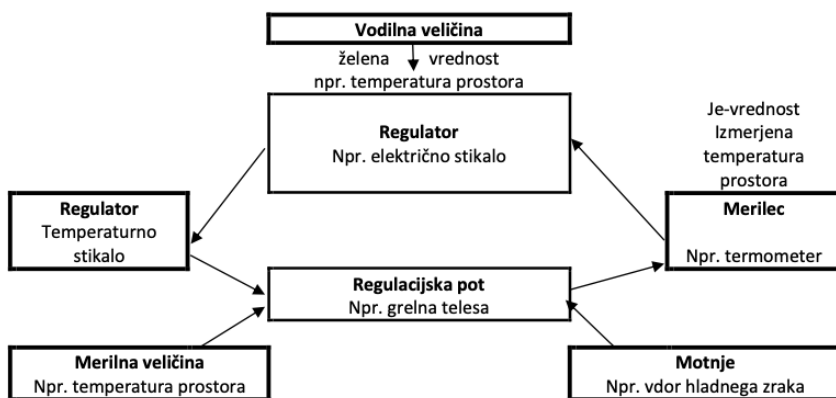


Slika 3.5: Kibernetični regulacijski krog učenja v programiranem pouku

V začetku šestdesetih let dvajsetega stoletja je pedagoški prostor z navdušenjem in velikimi pričakovanji sprejemal kibernetično pedagogiko, saj so si od nje obetali nesluten napredek na področju didaktike. Vendar je hitro sledilo razočaranje nad t.

i. *programiranim poukom*. V sedemdesetih letih dvajsetega stoletja so upali, da bo mogoče večino šolskega učnega procesa programirati in ga modelirati po vzoru tehničnih metod tega regulacijskega kroga.

Učitelju so pri tem namenili vlogo aktivnega regulatorja, ki izbere za učni proces najprimernejši nastavitveni člen (metodo pouka), s pomočjo katerega nastavitveni člen (učenec, ki se uči) usvoji nove regulacijske veličine (novo znanje). Pri tem označuje regulacijski krog obstoječe stanje (je-stanje) in načrtovano znanje (naj-znanje). Med njima pa označuje regulacijski krog še tipalo (orodje za preverjanje znanja), ki služi za ugotavljanje ali je količina je-znanja že dosegla količino naj-znanja. Šele takrat je regulacijski krog sklenjen in učni proces zaključen.



Slika 3.6: Sestavni deli in povezave kibernetiskega regulacijskega kroga

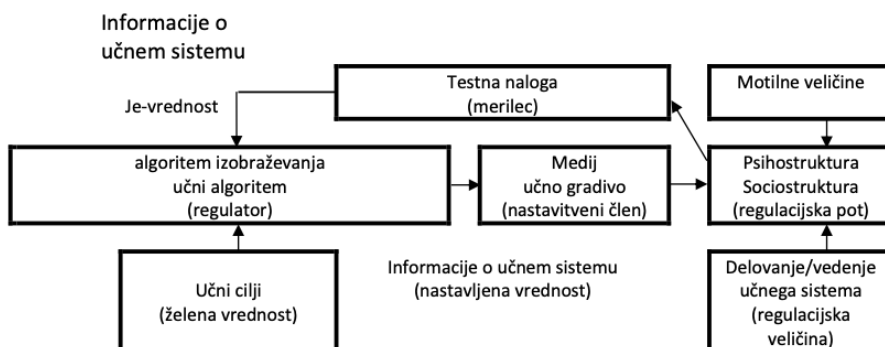
Za razumevanje in primerjavo si lahko ogledamo sliko 3.6, ki prikazuje primer kibernetiskega krmilnega kroga na principu krmiljenja temperature prostora. Frank in Mader (1971) sta želela razviti »*kibernetično pedagogiko*«, ki bi bila zasnovana na kibernetiki oz. na naravoslovnih znanostih. Kibernetična pedagogika bi bila tako znanstvena veda o tem, kako lahko vplivamo na učni proces.

V skladu s tem so izpostavljeni temeljni cilji kibernetične pedagogike:

- identifikacija in analiza poučevalnih in učnih procesov udeleženih delnih sistemov in njihova funkcija pri objektiviranju učno-vzgojnega procesa, kar pomeni prenos vseh dejavnosti na tehnične naprave oz. medije;
- analiza odnosov in medsebojnih učinkov objektiviranih in neobjektiviranih delnih sistemov učnih procesov z namenom, da bi dosegli zastavljene didaktične cilje;
- razlaganje soodnosov med različnimi oblikami delnih sistemov v vsakokratnem izobraževalnem sistemu.

Načrtovanje in izvedba ter evalvacija naj bi v čim večji meri potekala v okviru tehničnih sistemov, tj. računalniških programov.

Kibernetična pedagogika vidi pouk kot celoto komunikacijskih procesov med najmanj enim učnim programom (poučevalnim sistemom) in najmanj enim učencem (učnim se sistemom). Ti procesi se morajo podrežati krmiljenju in vodenju kibernetičnih modelov, kar ponazarja slika 3.7.



Slika 3.7: Elementi in funkcije v poučevalno učnem procesu s stališča kibernetične pedagogike

Po Franku (1993) je mogoče didaktiko formalizirati oz. objektivirati kot izobraževalni algoritem in ga izraziti kot logično matematično funkcijo petih pogojnih spremenljivk:

L – učno gradivo, M – mediji, P – psihološka struktura, S – sociološka struktura, Z – določanje učnih ciljev in B – učni ali poučevalni algoritem.

Učitelj je po Franku lahko (tehnično) realiziran tudi kot učni program (inteligentni tutor). Takšen program vsebuje *učni algoritem* B , ki ga lahko formalno zapišemo z logično – matematično funkcijo, ustreznih petih pogojnih spremenljivk:

$$B = f(Z, L, M, P, S)$$

Funkcije učnega sistema po Franku in Maderju (1971) so naslednje:

1. predstavitev relevantnih informacij, npr. učna snov, izhodiščna vprašanja, naloga, problem, navodilo, pomoč, presojanje, vrednostne sodbe;
2. sprejemanje, skladičenje in analiza informacij, ki jih posreduje učni sistem, npr. podatki o reakcijah, načini vedenja in odzivanja, odgovori;
3. evalvacija stanja, ki ga je diagnosticiral učni sistem in primerjava le-tega z načrtovano ciljno vrednostjo;
4. izbor ustreznih alternativnih učnih korakov/programov za nadaljnje vplivanje na učni sistem (povratne informacije, učne pomoči za učni sistem itd.)

Kibernetična didaktika prinaša pomembne osvežitve klasičnega didaktičnega razmišljanja. Zaradi tega je danes neupravičeno skoraj v popolni pozabi. Hkrati vsebuje ta didaktika precej pasti. Med vsemi je najbolj škodljivo omalovaževanje razlik med psihološkimi in pedagoškimi posebnostmi duševnega delovanja na eni ter tehničnimi sistemi na drugi strani. To mišljenje je udomačeno v strukturalistični filozofiji, ki podreja antropološke značilnosti modelom tehnike, češ da veljajo za človekovo mišljenje enaka organizacija in zakonitosti kot za svet tehnike. Ni tako redko prepričanje, da je s kibernetičnimi metodami mogoče objektivirati, modelirati in avtomatično upravljati tudi višje duševne aktivnosti in procese. Takrat se premalo upoštevajo posebnosti vzgojno-izobraževalnega področja, ko učenec ni le objekt, temveč prav toliko tudi subjekt lastnega vodenja in spreminjanja. V oči padejo nasprotja med smotri ljudskega učenja in kibernetičnimi učnimi potmi, se pravi med

zahtevami po razvijanju višjih, samostojnih miselnih aktivnosti in strogo vodenim učenjem, med posplošenim, sintetičnim mišljenjem in partikularnimi in analitičnimi učnimi postopki, med ustvarjalnostjo in avtomatiziranim učenjem. Ni tako preprosto vsebinsko in vrednostno učno konkretnost abstrahirati in jo zreducirati na simbole, formule in modele, ne da bi se spremenili v formalistično znanje, v izobrazbene klišeje. Danes se vsekakor ne moremo strinjati z Gilbertovim priporočilom: »Če nimate učnega stroja, nikar si ga ne nabavite ...«. Toda povsem moramo pritegniti njegovemu svarilu: »Nikoli ne dovolite, da stroj ukazuje programu«. (Gilbert v Strmčnik, 1987: 68)

3.4 Kognitivne arhitekture

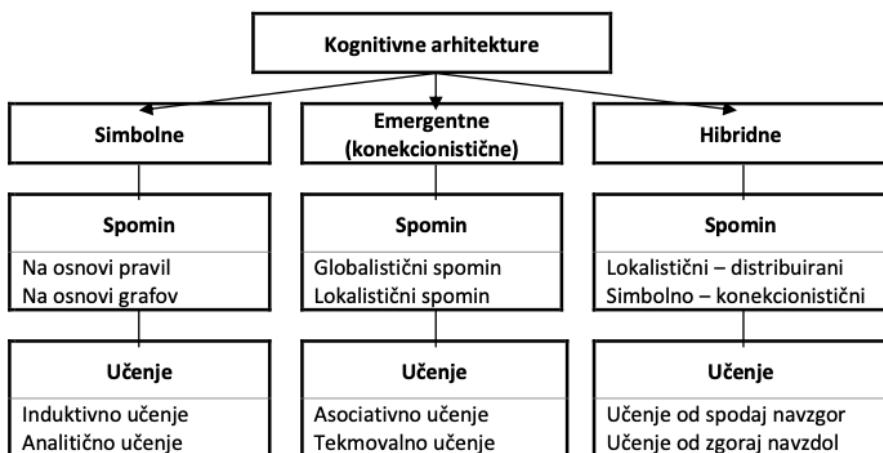
Kognitivna arhitektura je načrt za *inteligentne agente*,³⁸ ki ponuja umeten računalniški proces. Deluje podobno kot kakšen kognitivni sistem, npr. kot človek, ali deluje inteligentno pod določenimi pogoji. Kognitivna arhitektura je le ena od arhitektur inteligentnih agentov. Z izrazom arhitektura želimo poudariti, da pri modelu ni pomembno le njegovo vedenje, temveč tudi strukturne lastnosti in postopek modeliranja sistema. Pri tem ni nujno potrebno, da so vse lastnosti fizikalne, lahko so tudi virtualne in jih vključimo v fizikalne stroje (npr. v možgane ali računalnik). Kognitivne arhitekture, prikazane na sliki 3.8, so lahko naslednje: (Duch, Oentaryo, Pasquier, 2008)

- simbolni računalniški sistemi,
- emergentni in konekcionistični sistemi; sistemi nevronske mreže,
- dinamični sistemi.

Pri razvoju katerekoli kognitivne arhitekture sta najpomembnejša dva dejavnika: *spomin* in *učenje*. Pomen spomina so izpostavili (iz različnih perspektiv) mnogi avtorji, npr. Anderson (2007). Različne oblike spomina služijo za shrambo za znanje o svetu in nas samih in o vmesnih aktivnostih, medtem ko je učenje osnovni proces za oblikovanje teh znanj. Skupaj pa spomin in učenje iz začetnih znanj in veščin ustvarita nova znanja in nove veščine. Organiziranje spomina je odvisno od sheme

³⁸ Intelligentni agent je mnogokrat abstraktni funkcijski sistem, podoben računalniškemu programu. Velikokrat ko želimo poudariti, da to ni del realnega sveta (biološki sistem ali organizacija), temveč računalniški sistem, ga poimenujemo tudi abstraktni inteligentni agent (angl. Abstract intelligent agents – AIA).

predstavitve znanja. Poenostavljena taksonomija kognitivnih arhitektur, zasnovanih na teh dveh osnovnih dejavnikih, vodi do različnih modelov, ki temeljijo na treh osnovnih principih: *simbolnih*, *emergentnih* in *hibridnih*. (Duch, Oentaryo, Pasquier, 2008)



Slika 3.8: Poenostavljeni prikaz kognitivnih arhitektur

3.4.1 Klasična kognitivna znanost in simboli modeli

Kognitivisti so od behavioristov prevzeli idejo, po kateri je cilj psihologije razlaga in napoved vedenja. Prav tako so se strinjali, da je za doseganje tega cilja treba raziskovati pojave na višji ravni, kot je raven nevrofiziologije ali kakšne druge naravoslovne znanosti. Novost je v tem, da so odprli »črno škatlo« in vanjo postavili mentalna stanja kot člene v vzročni verigi razlage in napovedi vedenja. Pomembno vlogo pri oblikovanju teorije je imela računalniška metafora. M. Minsky je malce provokativno možgane označil kot meseni stroj. S strojem je mislil na računalnik – podobno, kot so pri računalniku programi (angl. software) realizirani v elektronskem stroju (angl. hardware), so tudi pri človeku kognitivni procesi realizirani v možganih. Tak pogled je bil osnova funkcionalistični teoriji duševnosti (Putnam, Fodor), po kateri so kognitivni procesi odvisni od procesov v fizičnem svetu na precej kompleksen način. So nekaj posebnega, ne da bi zato postali misteriozni duhovi. Možgani so običajen biološki realizator, kognicija pa je zadeva programske ravni. Glavni razlog, da so miselne procese obravnavali kot računske procese, je bil v tem,

da je taka obravnava omogočila (vsaj v načelu) razumevanje, kako se lahko fizične naprave vedejo racionalno. Osrednji teoretski pojem je postal pojem *reprezentacije*. To je vsebinsko mentalno stanje, ki predstavlja dogodke in stanja v zunanjem ali notranjem svetu in je simbol, realiziran v možganih. Vsebinsko določen tok misli (npr. misel »dežuje«, vodi do misli »pojdimo noter«) ustreza zaporedju teh simbolov, fizično realiziranih v možganih. Pravilom za transformacijo reprezentacij ustrezajo sintaktična pravila na ravni algoritmov, realizirana v fizičnem sistemu. (Newell, Simon, 1972)

Standardno so raziskovanja kognicije opredelili na treh ravneh: semantična (vsebinska) raven, sintaktična raven (algoritmična) in raven fizične realizacije. Prvi dve ravni sta bili funkcionalni in sta po teoriji kognitivistov prispevali k psihološki razlagi, tretja, strukturalna raven, pa naj k njej ne bi bistveno prispevala. Menili so namreč, da tudi če bi lahko »odprli glavo in pogledali v možgane«, v njih ne bi našli majhnih simbolov. Čeprav je sintaktična raven fizično realizirana, je ne moremo reducirati na fizično raven.

Znanstvene discipline, ki so preučevale fenomene na različnih ravneh, so se vključevale v skupni projekt klasične kognitivne znanosti. Filozofija in družbene vede na semantični ravni, lingvistika, psihologija, računalništvo (umetna inteligenca) na prvih dveh in nevroznanost na tretji. Pristop na prvih dveh ravneh je bil računski, zato ne preseneča, da je pomembna vloga v raziskovanju pripadla računalniškemu modeliranju – umetni inteligenci.

3.4.2 Kognitivno modeliranje

Termin kognitivna arhitektura je v kognitivno znanost vpeljal Newall kot analogijo računalniški arhitekturi (Bell, Newell, 1971), ki jo je Fred Brooks leta 1962 vpeljal v računalništvo kot analogijo arhitekture v gradbeništvu. Beseda *arhitektura* (latinsko *architectura*, iz starogrščine ἀρχιτέκτων – arkhitekton izhaja iz besed ἀρχι – 'vodja' in τέκτων 'graditelj, zidar'). *Klasična arhitektura* je ena od najstarejših oblik arhitekture in izhaja iz antične Grčije, svojo afirmacijo je doživela predvsem v antičnem Rimu. Klasična arhitektura temelji na antropomorfem preučevanju razmerij človeškega telesa ter uporabi enakih razmerij pri oblikovanju zelenega prostora. Termin arhitektura lahko pomeni:

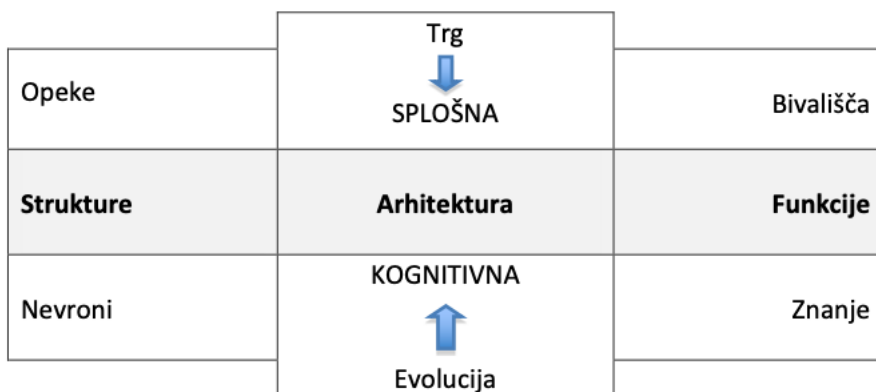
- umetnost in znanost o konstruiranju in gradnji zgradb in drugih fizičnih objektov;
- praktično delo arhitekta, pri čemer arhitektura pomeni nudenje profesionalne usluge povezane s projektiranjem in konstruiranjem zgradb ali skupine zgradb, vključno s človekovim bivalnim prostorom okoli njih;
- splošen izraz, ki opisuje zgradbe in ostale strukture;
- slog ali metodo projektiranja in konstruiranja zgradb in drugih fizičnih objektov.

Še nekoliko širša definicija vsebuje vse projektantske aktivnosti od makro nivoja (urbano projektiranje, krajinsko arhitekturo) do mikro nivoja (oblikovanje detajlov in pohištva). Arhitektura je oboje, proces in proizvod, ki upošteva planiranja, oblikovanja in konstruiranja oblik, prostora in ambienta, in se odraža v funkcionalnosti, tehnični dovršenosti ter socialni, okoljski in estetski sprejemljivosti.

Skratka arhitektura je – najbolj splošno – umetnost specificiranja strukture zgradbe na abstraktnem nivoju, zato da se zagotovi, da bo graditelj zadovoljil vse uporabnikove zahteve. Če v tem kontekstu povzamemo Brooksove besede:

»Računalniška arhitektura je, tako kot tudi vse druge arhitekture, umetnost določanja uporabnikovih potreb (zahtev) po neki strukturi in nato čim bolj učinkovito zadovoljevanje teh potreb, tako s stališča ekonomičnosti kot tudi tehnološkičnosti.« (Brooks, 1962: 110)

Iz te definicije izhaja, da arhitektura v tem kontekstu pomeni bolj rezultat konstruiranja kot celoten proces. Na enak način sta tudi Bell in Newell definirala ta pojem. Iz takšnega razumevanja izhaja tudi Newellova definicija kognitivne arhitekture, ki jo je postavil leta 1990: »Kognitivna arhitektura je statična (ali počasi se spreminjajoča) struktura, ki ustvarja okvir za neposreden proces kognitivnih predstav in učenja«. (Brooks, 1962: 111) Teh definicij je več, omenimo le še J. R. Andersenovo z leta 1983, da je kognitivna arhitektura teorija o osnovnih principih operacij, vgrajenih v kognitivni sistem, kot ponazarja naslednja slika 3.9:



Slika 3.9: Primerjava principa arhitekture v gradbeništvu in kognitivni znanosti

Preden se je pojavila ideja o kognitivni arhitekturi, so imeli znanstveniki, ki so se ukvarjali s tem področjem, dve možnosti:

- ali se osredotočiti na *strukturo* in se pričeti izgubljati v neskončnih detajlih človeških možganov (struktura, ki ima približno 100 milijonov nevronov);
- ali se osredotočiti na *funkcijo* in se pričeti izgubljati v neskončnih detajlih človekovega obnašanja.

Če hočemo razumeti duha, potrebujemo abstrakcijo, ki sega do samega bistva kognitivne arhitekture. Ta abstrakcija mora izhajati iz razumevanja povezav med strukturo in funkcijo in ne iz poudarjanja ene ali druge. A tudi o tem se znanstveniki ne strinjajo v celoti.

V disertaciji bomo zagovarjali integrativno teorijo duha (Anderson, Bothell, Byrne, Douglass, Lebiere, Qin, 2004) in skušali dokazati, da je to trenutno najboljša možna rešitev. Upoštevajoč napisano bi lahko zapisali definicijo kognitivne arhitekture v naslednji obliki: *Kognitivna arhitektura je specifikiranje strukture možganov na abstraktnem nivoju, ki razlaga, kako doseči funkcionalnost duha.* Njen cilj je razložiti pojma, kot sta:

- funkcija – delovanje duha, kar predstavlja po teoriji kognitivistov semantično (vsebinsko) in sintaktično (algoritmčno) raven, ki prispevata predvsem k psihološki razlagi;

- struktura – struktura možganov predstavlja raven fizične realizacije.

3.4.3 Razvoj kognitivne arhitekture

Če želimo govoriti o kognitivni arhitekturi, moramo pri tem upoštevati tri stvari:

- možgane,
- duh oz. funkcionalno kognicijo,
- arhitekturno abstrakcijo, ki povezuje možgane z duhom.

Mnogi avtorji so skušali bolj ali manj uspešno poiskati različne »bližnjice«. Za primer pogledjmo le tri najpogostejše.

- *Bližnjica: Klasična informacijsko procesna psihologija – IGNORIRANJE MOŽGANOV*

Allen Newell in Herbert Simon nista poudarjala pomena možganov pri specifikaciji kognitivne arhitekture. Takšen pristop je bil zelo uspešen v šestdesetih in sedemdesetih letih 20. stoletja vse do pojava konekcionizma v osemdesetih. Zagovorniki konekcionizma so poudarjali, da je procesiranje v možganih in v računalniku nekaj popolnoma različnega, saj nevroni v možganih delujejo zvezno in vzporedno, vendar počasi, medtem ko je računalnik diskreten in zaporeden in izvaja sekvence aktivnosti izredno hitro. V devetdesetih letih, se je s pojavom nevronskega upodobitvev (angl. neural imaging) pokazal pomen razumevanja delovanja možganov kot strukture, ki je osnova kognicije.

- *Bližnjica: Eliminativni konekcionizem – IGNORIRANJE DUHA*

Eliminativni konekcionizem je tip konekcionizma, ki poudarja, da je vse, kar moramo narediti, da se posvetimo možganom – le opisati moramo, kaj se dogaja v možganih na nekem nivoju abstrakcije. Možganske strukture bodo generirale vedenje ljudi in vedenje je funkcionalno.

– *Bližnjica: Racionalne analize – IGNORIRANJE ARHITEKTURE*

Osnova teh modelov je Bayesinova statistična metodologija, ki v nekem pogledu predstavlja nadomestek za kognitivno arhitekturo. Na kratko, Bayesinova metodologija zahteva naslednje:

(1) Imeti moramo niz predhodnih omejitev o naravi sveta. Te pretekle izkušnje odsevajo statistično urejenost sveta.

- (1) Ko podamo različne (pretekle) izkušnje, lahko izračunamo pogojno verjetnost.
- (2) Po vnosu podatkov lahko izračunamo naslednjo verjetnost na osnovi preteklih izkušenj – omejitev, tj. začetnih in robnih pogojev (1) in pogojnih verjetnosti (2).
- (3) Po teh preračunih uporabimo Bayesovo odločitveno strukturo in pričnemo z optimiranjem pričakovane koristi.

Vse te bližnjice niso nikoli dejansko odgovorile na osnovno dilemo: *Kako se lahko človeški duh pojavi v fizičnem vesolju?*

Oglejmo si na kratko, kako se v ACT-R, to je hibridnem kognitivnem sistemu, izogibamo takšnim bližnjicam:

- V nasprotju s klasičnimi modeli procesiranja informacij je arhitektura neposredno povezana s podatki o možganih. Vsi podatki o slikovnih obdelavah možganov so pri tem izjemno pomembni – na njih temelji celoten razvoj ACT-R.
- V nasprotju z eliminativnim konekcionizmom je arhitekturni predlog osredotočen tudi na to, kako doseči celovito funkcionalnost sistema.
- V nasprotju z racionalnimi predlogi in nekaterimi konekcionističnimi predlogi ACT-R ne zanemara tega, kako so komponente arhitekture integrirane. Še več, ACT-R je predvsem teorija o integraciji, bolj kot karkoli drugega.

3.4.4 Filozofija uma in STEAM izobraževanje

Živimo v svetu, ki ga v veliki meri krojijo različne tehnologije, zato mora vsakdo razviti določen odnos in sprejeti vedenje ter zavedanje o tehnologiji, kar na kratko poimenujemo STEAM³⁹ izobraževanje. Naši računalniki in naši pametni telefoni so tehnologija. Naš prevoz je tehnologija. Naši domovi, pisarne in šole, naša hrana in oblačila, naše ogrevanje in hlajenje, naša zabava, zaščita in onesnaževanje, proizvodnja energije kreirajo in upravljajo različne tehnologije. Podobno bodo številni kritični izzivi, s katerimi se kot družba soočamo in se bo človeštvo morda moralo soočiti še bolj intenzivno v prihodnosti, zahtevali povezovanje strokovnjakov in tehnologij po vsem svetu, npr. za iskanje trajnostne energije (Kutscher, Milford, Kreith, 2018), ukvarjanje z globalnimi pandemijami ali razvojem varstva okolja in okolju prijaznega kmetijstva ter za prehrano vse bolj intenzivnega naraščanja svetovnega prebivalstva (NAGB, 2018, ITEA, 2003, 2007). Vsa ta prizadevanja, razvoj, upravljanje in vzdrževanja našega načina življenja in človekovega ustvarjanja njegovega socialnega okolja in z njim podrejanja naravnega okolja njegovim potrebam in željam, poimenujmo s skupnim imenom *tehnologija* in izpeljano iz tega to področje definirajmo kot *tehnološka in inženirska pismenost* po analogiji vseh ostalih pismenosti. (Flogie, et.al, 2018)

Kljub svoji pomembnosti pa tehnologija ni v središču poučevanja v našem izobraževalnem sistemu, zlasti na osnovni in srednji ravni. Nekatere tehnologije se vključujejo v pouk humanistike, družboslovja, naravoslovja in matematike kot metode za podporo učenju teh predmetov. Različne tehnologije prav tako postajajo predmet študija, s ciljem razvijanja razumevanja samih tehnologij, njihovih različnih vlog in procesov inženirskega načrtovanja za reševanje človeških potreb in želja, pa tudi človeškega zavedanja. Vendar pa ni standardov znanj, ki bi presojali in jasno definirali kaj tehnološka pismeni učenci morajo vedeti o znanosti, tehnologiji in inženirstvu ter o njihovi vlogi, ki jih imajo v naših življenjih. Prav tako pa tudi ne o tem, v kolikšni meri lahko nekdo uporablja in predvsem razume (različne etične

³⁹ STEAM (angl. science, technology, engineering, visual or performing arts and mathematics) je izobraževalna strategija, ki vključuje tudi umetnost v ustaljen termin STEM. Pri naših razlagah se bomo osredinili predvsem na tri segmente te izobraževalne filozofije, to je naravoslovje (angl. science), tehnologijo (angl. technology), inženirstvo (angl. engineering) in delno tudi umetnost (angl. arts), s posebnim poudarkom na ustvarjalnost, in kreativnost in ob tem vizualizacijo, kar pomeni 3D modeliranje, ustvarjanje virtualnih modelov in virtualno realnostjo, kjer postajajo meje med inženirstvom in umetnostjo zelo zabrisane in ni več realnih prehodov med njima.

dileme) tehnologije ter razume, kako inženirji oblikujejo in te tehnologije razvijajo (Kordigel, Aberšek in Aberšek, 2021).

Tehnološka in inženirska pismenost pa bosta morali biti vedno bolj vključeni v šolski prostor tudi na elementarni in splošni srednji stopnji, od pouka o uporabi računalnikov in informacijske tehnologije pri različnih šolskih predmetih do medpredmetnih povezav, ki preučujejo vlogo tehnologije v družbi, ali tečajev (izbirnih predmetov), ki poučujejo npr. inženirsko načrtovanje. Informacijsko-komunikacijske tehnologije so postale sestavni del uporabe v akademskih, šolskih, delovnih in praktičnih kontekstih. Tehnologija in inženiring bi morala biti bistveni sestavini sodobnega izobraževanja na področju znanosti, tehnologije, inženirstva in matematike. (NAEP, 2018) Pri tem pa je pomembno to, da ta znanja ne smejo biti monodisciplinarna, temveč morajo tvoriti zaključeno interdisciplinarno (morda celo transdisciplinarno) STEAM celoto, učence naučiti povezovanja znanj različnih strok (predmetov) pri reševanju problemov, skratka načrtovanja poti od problema do rešitve tega problema ob uporabi vseh znanj in spretnosti, ki so jih učenci osvojili tekom šolanja. (Flogie, Aberšek, 2015)

Pri tem pa se ne sme pozabiti in zanemariti vprašanja morale in etike v vseh inteligentnih sistemih in tehnologijah, bodisi fizičnih sistemih z umom in telesom bodisi samo mislečih algoritmov v oblaku. (Agar, 2012) Morala izkazuje odnos človeštva do narave in družbe in se kaže kot vsota vrednot (pravil, norm, principov, kategorij, zavesti itd.), po katerih se odločamo, kaj je dobro in kaj slabo, kaj je pravično in kaj nepravično, kaj je prav in kaj narobe, kaj smemo in česa ne. V skladu s temi pravili se obnašamo in razvijamo svojo ponotranjeno zavedanje.

Tako kot velja za vse skupine, je tudi družba opredeljena s povezavami med njenimi člani, torej posamezniki, ki jo sestavljajo. Da bi razumeli družbo, moramo raziskati, kako nanjo vplivajo fizično okolje, kultura (socialno okolje) in medosebni odnosi. Vsako od teh področij proizvaja in vpliva na družbene vrednote in družbeno zavest; izobraževanje na primer vpliva na gospodarstvo in odnos do okolja. To spreminja kulturne vzorce družbe kot celote in posameznikov te družbe in njihovo zavest ter tako vpliva na celotno družbo. (PISA, 2015, Kordigel Aberšek, et. al, 2017) Temu poskušajo slediti moralni in etični premisleki pri implementaciji različnih tehnologij v družbo. Pri tem je poudarek je na vprašanjih, kot je *'ali bodo prihodnji subjekti imeli moralni status?'* (Bostrom, 2014), *'kdo je odgovoren za delovanje tehnologij in ali naj sploh*

ustvarimo pravila in zavest, ki bo imela status moralnega agenta?' (Bryson, 2018) ali *'kako umestiti tehnologije v svet, da bo deloval v skladu z našimi (človeškimi) pričakovanji?'* (Rahwan, 2018) itd.

Etika in tehnologija sta tako povezani na več ravneh:

1. „*etika pri načrtovanju*“: tehnična integracija etičnih zmožnosti in sklepanja;
2. "*etika v oblikovanju*": regulativne in inženirske metode, ki zagotavljajo analizo in oceno etičnih posledic delovanja, npr. sistemov umetne inteligence;
3. "*etika za oblikovanje*": kodeksi ravnanja, standardi in postopki certificiranja, ki zagotavljajo integriteto razvijalcev in uporabnikov pri raziskovanju, oblikovanju, uporabi in upravljanju tehnoloških sistemov.

Eden od današnjih najbolj perečih etičnih problemov ob okoljski problematiki je seveda pri razvoju tehnologij temelječih na UI. Zato moramo poskrbeti, da bo sklepanje inteligentnih agentov temeljilo na upoštevanju družbenih norm in bo v skladu z osnovnimi človeškimi načeli (moralo) ter da bodo svojo odločitev in zavedanje lahko utemeljili ob transparentnem delovanju. Le s tem bo zagotovljena blaginja in napredek v trajnostnem svetu. (Dignum, 2018) Kljub temu pa se zdi, da je filozofija nekako pozabila na nekatera pomembna vprašanja s področja etike in etičnih rešitev umetne inteligence, vprašanja, ki zadevajo moralo in ozaveščenost v izobraževanju, najbolj splošno: *Kakšni so moralni problemi pri uporabi različnih vrst UI rešitev, povezana s STEAM pismenostjo?* (Etics Commision, 2017)

Okvir tehnološke in inženirske pismenosti se osredotoča na »pismenost« kot raven znanja in kompetenc, ki jih potrebujejo vsi državljani. (PISA 2015, 2018) Od ljudi, ki so tehnološko in inženirsko pismeni, se ne pričakuje, da bodo »delali« kot inženirji ali proizvajali tehnologijo v poklicnem smislu. Zato okvir tehnološke in inženirske pismenosti ni namenjen obravnavanju tehničnega, tehnološkega in inženirskega znanja o specifičnih tehnologijah ali vrstah inženirskega strokovnega znanja, ki se poučuje v ustreznih strokovnih šolah ali ki študente pripravljajo na poklic tehnika/inženirja na srednješolskem, višješolskem/visokošolskem ali univerzitetnem inženirskem študiju. Mora pa dati vedenja, tako da lahko v četrtem razredu na primer od vseh učencev pričakujemo, da prepoznajo vrste tehnologij v svojem svetu svoji razvojni stopnji primerno, oblikujejo in preizkusijo preprost model, razložijo, kako lahko tehnologije povzročijo pozitivne in negativne učinke

(na primer v povezavi z okoljsko problematiko), ter uporabijo skupne tehnologije za doseganje ciljev in zavedanja v šoli in v vsakdanjem življenju. Od srednješolca splošnega programa pa se na primer lahko pričakuje, da bodo vsi učenci izbrali in uporabljali različna orodja in medije za izvajanje raziskav, ocenjevanje, kako dobro rešitev izpolnjuje določena merila, in razvili načrt za reševanje kompleksnega globalnega vprašanja, na primer tudi globalnega zavedanje o vplivu umetne inteligence na vsakdanje življenje ali vpliv tehnologij na naše okolje.

3.4.5 Tehnološka in inženirska pismenost

Področja tehnološke in inženirske pismenosti

Vsako ocenjevanje tehnološke in inženirske pismenosti ljudi se mora začeti z jasno predstavo o tem, kaj natančno pomeni tehnološka in inženirska pismenost, (NAEP, 2018) to pa zahteva jasne definicije tehnologije in inženiringa. V zgodovini je imel izraz »tehnologija« širši in globlji pomen in je ozko povezan z definicijami izhajajočimi in filozofije tehnologije in prav ta pomen mora biti predstavljen v našem primeru za potrebe splošnega izobraževanja:

- *Tehnologija* je kakršna koli sprememba naravnega sveta, narejena za izpolnitev človeških potreb ali želja. Po tej definiciji tehnologija zajema celoten svet, ki ga je ustvaril človek (njegovo socialno okolje), od najpreprostejših artefaktov, kot sta papir in svinčnik, do najbolj kompleksnih – zgradb in mest, električnega omrežja, satelitov in interneta. Poleg tega tehnologije niso le stvari, artefakti, ki jih ustvarijo ljudje. Vključuje celotno infrastrukturo, potrebno za načrtovanje, proizvodnjo, delovanje in popravilo različnih tehnoloških artefaktov (strojev, naprav, npr. avtomobilov), do sedeža podjetja in inženirskih šol, proizvodnih obratov, medijskih hiš in distribucijskih omrežij. Inženirji so agenti za načrtovanje številnih tehnologij, ki spreminjajo svet. Inženirji morda dejansko ne bodo izdelovali artefaktov, kot so kalkulatorji, mostovi ali letala. Razvijali pa bodo načrte in navodila za izdelavo artefaktov, kot tudi procese oblikovanja, kot so tekoče linije ali postopki, na primer za inteligentne industrijske rešitve, kot so pametne tovarne ali pametna mesta.
- *Inženirstvo* je sistematičen in pogosto ponavljajoč se pristop k načrtovanju objektov, procesov in sistemov za izpolnjevanje človeških potreb in želja. Ta

okvir opredeljuje tehnološko in inženirsko pismenost na splošen, pa tudi partikularen način.

- *Tehnološka in inženirska pismenost* je tako zmožnost uporabe, razumevanja in ocenjevanja tehnologije ter razumevanja tehnoloških načel in strategij, potrebnih za razvoj rešitev in doseganje ciljev. Tako imata tehnološka in inženirska pismenost veliko skupnega z znanstveno, matematično, pa tudi jezikovno pismenostjo, kot je na primer najbolj znana tehnologija – informacijsko komunikacijska tehnologija (IKT). Kot pri različnih oblikah akademske pismenosti tudi tehnološka in inženirska pismenost vključujeta obvladovanje nabora orodij, potrebnih za inteligentno in premišljeno sodelovanje v družbi. Orodja so različna, a končni cilj je enak. Eden posebej pomembnih sklopov tehnoloških orodij sestavljajo informacijske in komunikacijske tehnologije kot sestavni del tehnološke in inženirske pismenosti. (Kordigel Aberšek, Aberšek, 2021)

Vsebine tehnološke in inženirske pismenosti (NAEP, 2018) so organizirane v tri glavna področja tehnološke in inženirske pismenosti, od katerih vsako ustreza široki skupini znanj in spretnosti, ki jih morajo učenci poznati. To so:

- tehnologija in družba,
- oblikovanje in sistemi in
- informacije in komunikacije.

Tehnologija in družba

Ukvarja se z učinki, ki jih ima tehnologija na družbo (socialno okolje, ki ga je ustvaril človek) in naravno okolje, ter vrstami etičnih vprašanj, ki izhajajo iz teh učinkov. Znanje in zmožnosti na tem področju so ključnega pomena za razumevanje problematike razvoja in uporabe različnih tehnologij ter za sodelovanje pri odločanju o njihovi uporabi.

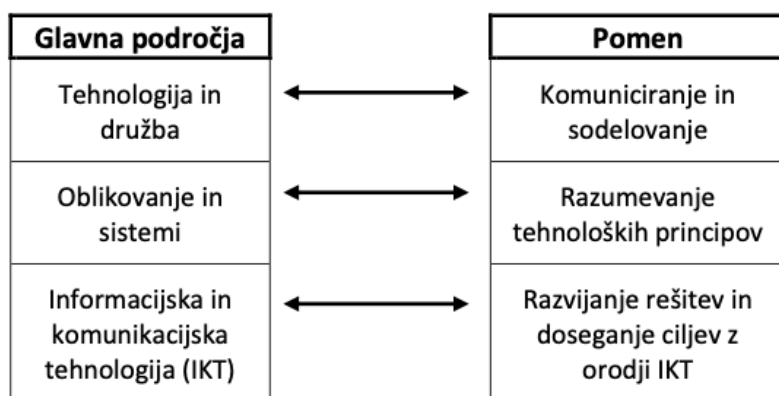
Oblikovanje in sistemi

Zajema naravo tehnologije, proces inženirskega načrtovanja, s katerim se razvijajo tehnologije, in osnovna načela ravnanja z vsakodnevnimi tehnologijami, vključno z vzdrževanjem in odpravljanjem težav. Razumevanje procesa oblikovanja je še

posebej dragoceno pri ocenjevanju tehnologij in se lahko uporablja tudi na področjih zunaj tehnologije, saj je oblikovanje široko uporabna veščina.

Informacije in komunikacije

IKT tehnologija vključuje računalnike in orodja za učenje, programske opreme, omrežne sisteme in protokole, ročne digitalne naprave in druge tehnologije za dostop, ustvarjanje in sporočanje informacij ter za omogočanje ustvarjalnega izražanja (IKT). Čeprav je le ena izmed več vrst tehnologij, je dosegla posebno mesto v splošni tehnološki in inženirski pismenosti kot tudi veljavo v splošni izobrazbi, saj sta poznavanje in enostavnost uporabe informacijsko tehnoloških orodij bistvenega pomena v skoraj vsakem poklicu v sodobni družbi.



Slika 3.10: Področja tehnološke in inženirske pismenosti

Vir: Kordigel Aberšek, Aberšek, 2021

Ta tri področja tehnološke in inženirske pismenosti, kot je prikazano na sliki 3.10, so tesno povezana. Na primer, če želite obravnavati vprašanje, povezano s tehnologijo in družbo, kot so čista voda, potrebe po energiji, komunikacije v digitalni dobi ali globalne industrijske spremembe, je pomembno razumeti kako ti tehnološki sistemi delujejo in proces oblikovanja za njihovo realizacijo. Podobno je treba znati uporabljati tudi različne informacijske in komunikacijske tehnologije za raziskovanje problemov, razvoj in komuniciranje možnih rešitev za doseganje željenih ciljev. Na vseh treh področjih tehnološke in inženirske pismenosti se od ljudi pričakuje, da bodo sposobni uporabiti posebne *načine razmišljanja in sklepanja*, ko se bodo lotili

določenega problema. Te vrste razmišljanja in sklepanja se imenujejo prakse, okvir pa določa zlasti tri vrste, ki naj bi jih ljudje pokazali, ko odgovarjajo na problemska vprašanja:

- razumevanje tehnoloških načel, na primer kaj je AI in kako deluje;
- razvijanje rešitev in doseganje ciljev in
- komuniciranje ter sodelovanje.

Razumevanje tehnoloških principov

Osredotoča se na to, kako dobro lahko ljudje razumejo in posledično uporabijo svoje znanje o določeni tehnologiji. To znanje sega od diskretnih deklarativnih dejstev in konceptov do sklepanja na višji Bloomovi (Bloom, 1956) ravni o tem, kako so dejstva, koncepti in načela organizirani v različne strukture in razmerja. Ljudje bi morali imeti možnost uporabiti svoje znanje in razumevanje o tehnologiji za razvoj razlag in napovedi, primerjav in vrednotenj.

Razvijanje rešitev in doseganje ciljev

Nanaša se na sistematično uporabo tehnološkega znanja in razumevanja, orodij in veščin za reševanje problemov in doseganje ciljev, predstavljenih v realnih kontekstih. Ta praksa vključuje postopkovne in strateške zmožnosti – vedeti, kako uporabiti preproste korake in uporabiti tehnološka orodja za obravnavo resničnih situacij, pa tudi kdaj in kje uporabiti orodja in procese.

Komuniciranje in sodelovanje

Skrbi, kako dobro lahko posamezniki uporabljajo sodobne tehnologije za komuniciranje za različne namene in na različne načine pri delu individualno ali v skupinah, z vrstniki in strokovnjaki.

O teh treh vrstah razmišljanja in sklepanja je potrebno že od najzgodnejšega otroštva naprej razvijati zavedanje, torej je treba pri posamezniku gojiti npr. odnos do IKT in uporabe in razumevanja delovanja aplikacij temelječih na umetni inteligenci. (Kordigel, Aberšek in Aberšek, 2020)

3.4.6 Učne strategije temelječe na možganskih procesih

Razmišljanja o tem, da bi morali pri poučevanju upoštevati to, kako se ljudje oz. naši možgani adaptirajo in učijo, segajo že mnogo desetletij nazaj, saj je o tem govoril že Skinner v *Teoriji programiranega pouka* (programirane sekvence), ki jo je predlagal že leta 1958. Po Skinnerju je osnovni in najpomembnejši cilj programiranega pouka *izvajanje učenja v kontroliranem okolju*. Postavil je temelje tega pouka, ki naj zasleduje tri cilje:

1. podaja naj informacije v *manjših (vsebinskih) sklopih*;
2. namenjen je za *samoučenje* in
3. omogoča *takojšnje preverjanje znanja in povratne informacije učečim*.

Kognitivna znanost in nevroznanost pa sta na podlagi intenzivnih raziskav zadnjih let ponudili in potrdili na podlagi vedenja o delovanju možganov vrsto spoznanj in na podlagi tega vrsto nasvetov, kako izvajati učenje v kontroliranem okolju kot na primer:

Nasvet 1: Pomembne stvari so vedno na začetku

Nasvet 2: Spodbujanje dosledne prakse - sistematičnost

Nasvet 3: Sprotno uvajanje novosti in z njimi inoviranje učnega procesa

Nasvet 4: Ustvarjanje veččutnih izkušenj učenja

Nasvet 5: Pohvalimo prizadevanje

Nasvet 6: Vsebino naj bo razdeljena na manjše dele

Nasvet 7: Pomagajte učencem dostopati do predhodnega znanja

Nasvet 8: Poskusite več različnih pristopov

Nasvet 9: Raznolikost

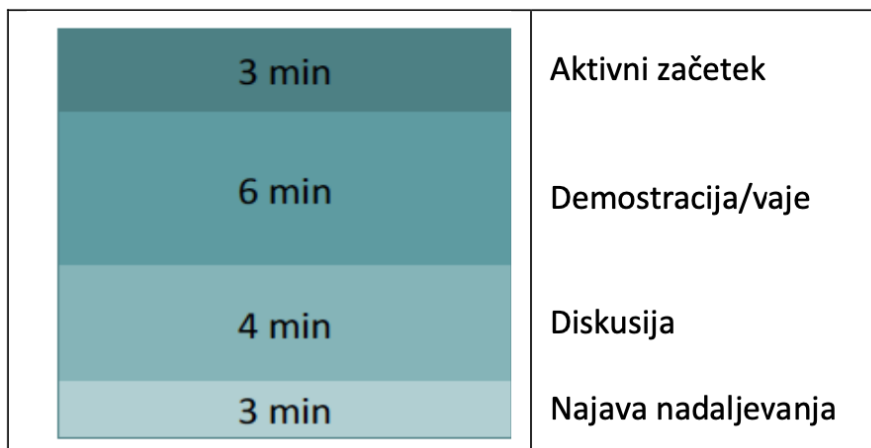
Nasvet 10: Poudarjanje pomembnosti pouka/znanja

Nasvet 11: Sprožite pravo čustvo/pozitiven odnos

Omeniti je potrebno, da se je vrsta teh nasvetov na tak ali drugačen način že v procesu učenja in poučevanja intuitivno izvajala, vendar pa sta nam sodobni kognitivna znanost in nevroznanost odgovorili tudi na vprašanje *zakaj (vprašanje je neposredno vezano z razumevanjem)*. Če se tako nekoliko poglobimo le v *nasvet 6* in na podlagi tega pripravimo primerno “učno uro”, ki ni več, tako kot v tradicionalni šoli, dolga natančno 45 minut, ampak je “učna ura” fleksibilna in sestavljena iz segmentov v trajanju le 10–20 minut. Ti segmenti učne ure so lahko krajši (npr. 10 min) ali daljši (npr. 20 min) in učna ura je prav tako lahko sestavljena iz enega ali več segmentov. Kognitivna znanost in nevroznanost sta dokazali, da človekov *delovni spomin* lahko ostane aktiviran le 10–20 minut odvisno od zahtevnosti naloge ali snovi, ki jo je potrebno usvojiti in seveda razvojne stopnje učenca. Na osnovi vedenja o tem, kako se možgani učijo, mora biti idealna lekcija torej razdeljena na nekakšne zaključene segmente v trajanju 10–20 minut in znotraj njih na tri časovne podsegmente (upoštevajmo tudi Nasvet 1):

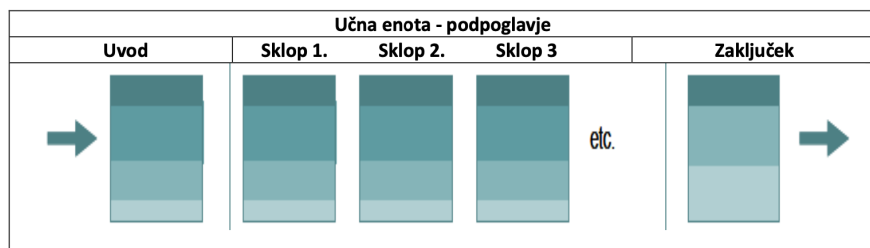
- 1. učeči si najboljše zapomnijo informacije, ki so na prvem mestu (čas 1);
- 2. nato si najboljše si zapomnijo podatke na koncu (čas 2);
- 3. najslabše pa si zapomnijo informacije, ki prihajajo v sredini (vmesni čas).

Na osnovi česar lahko zgradimo optimalen učni blok po sliki 3.11:



Slika 3.11: Učni blok

in na podlagi posameznih učnih blokov pripravimo ustrezno obravnavo določene učne vsebine/snovi:



Slika 3.12: Učna vsebina sestavljena iz učnih blokov

Med posameznimi učnimi bloki so kratki "motivacijski vložki" (Nasvet, 5, 7, 8, 9), v obliki spremembe učne aktivnosti in/ali vsebine (npr. diskusije, raziskovanja, fizičnih aktivnosti, ipd.). Na podlagi tega enostavnega izhodišča si nekoliko podrobneje pogledjmo teorijo, ki stoji za tem. (Aberšek, 2018)

V zgodnjih devetdesetih letih prejšnjega stoletja je bilo učenje, temelječe na vedenju o delovanju možganov, vse pogostejše uvajano v izobraževalne procese. Teorija učenja na podlagi možganskih procesov je usmerjena v strukturo in delovanje možganov. Dokler možganom ni prepovedano izpolnjevati svojih običajnih procesov, prihaja do učenja. Od zametkov uvedbe učenja na osnovi vedenja o tem, kako se možgani učijo (Caine, Caine, 1991) ga edukacijska znanost uvaja in spremlja z veliko pozornostjo. Čeprav so bila načela tekom let in novih spoznanj nekoliko spremenjena, so zelo dobro prestala preizkus časa. Bolj kot kadarkoli prej pa kognitivna znanost in nevroznanost zagotavljata tudi razumevanje, kaj se dejansko zgodi, ko se ljudje učijo, kar je temelj za napredno, inovativno poučevanje. (Morris, Filenz, 2003)

Učenje na podlagi vedenja o delovanju možganov ali teorija učenja, združljiva z procesi učenja v možganih, se osredotoča na koncepte, ki ustvarjajo optimalno učno okolje za maksimiranje usvajanja, doseganja in zadrževanja informacij (od delovnega do dolgotrajnega spomina). Uspešna aplikacija teh spoznanj je odvisna od vseh, ki sodelujejo v učnem procesu – razvijalcev učnih okolij, učiteljev in študentov – da razumejo strukturo in delovanje možganov in se osredotočijo na potrebe in sloge usvajanja znanja in procesiranja informacij v naših možganih. Naloga tvorcev učnih

okolij in učiteljev v tem procesu je, da za učence ustvarijo učna okolja, materiale, in poučevanje na zabaven, smiseln in diferenciran, individualiziran in bogat način. Učenje na osnovi možganskih tehnik je veliko boljši in bolj naraven kot tradicionalne tehnike predavanja. Vendar pa mora učitelj razumeti in se zavedati, kako te inovativne tehnike implementirati v napredno učno okolje. Oblikovalci izobraževalnih orodij morajo biti kreativni pri ustvarjanju možganom prijaznega učnega okolja. Učitelji se morajo pri tem zavedati, da najboljši način za učenje ni s predavanji, temveč z udeležbo v realističnih okoljih (delavnicah, laboratorijih, v naravi), ki učencem omogočajo varno preizkušanje in usvajanje novih spoznanj.

Temelji: Učenje kot celostni proces

Raziskave s področja kognitivne znanosti in nevroznanosti postavljajo pod vprašaj temeljno prepričanje o izobraževanju, ki je preživelo štiristo let. Od pisanja Descartesa v 17. stoletju se domneva, da sta telo in um ločena. Ta pristop se zdaj izpodbija na nov način. Nevroznanstveniki, kot je Antonio Damasio, jezikoslovci, kot je George Lakoff, in kognitivni znanstveniki, kot je Mark Johnson, kažejo, da so telo, možgani in um globoko povezani. Čeprav je lahko določena funkcija (kot je slisanje zvokov ali videnje obrazov) v nekaterih pogledih ločena, je bistvo tega, predvsem v šoli v procesu učenja, da je vsaka oseba povezana celota, ki sodeluje s svetom kot popoln sistem. (Morris, Filenz, 2003)

Načela učenja možganov/uma – sistemska načela naravnega učenja

Če telo, možgane in možganske procese (um) pojmuje kot dinamično enoto, potem postane mogoče prepoznati temeljna splošna načela učnega sistema. Za izboljšanje izobraževanja moramo v principu izpolnjevati naslednje štiri kriterije učnega sistema:

1. pojavi, ki jih opisujejo načela učnega sistema na temelju kognitivne znanosti in nevroznanosti, morajo biti univerzalni in uporabni za vsa človeška bitja;
2. načela, izhajajoča iz raziskav več različnih disciplin, bi morala biti interdisciplinarno povezana v smotno celoto kognitivnih znanosti;
3. načela morajo biti adaptivna in razvojno orientirana in naj predvidevajo prihodnje raziskave;

4. načela bi morala imeti ne le teoretično znanstveno raven, teoretična izhodišča, temveč bi morala biti tudi aplikativna, torej upoštevati povezave teorije s prakso in imeti pozitivne posledice na izobraževalno prakso.

Na teh kriterijih pa temeljijo načela kot rezultat velikega števila raziskav z različnih raziskovalnih področij. V bistvu ta načela kažejo, kako so številni različni človeški vidiki integrirani v celoten učni proces na različnih področjih, predmetih ali domenah. Načela so naslednja:

1. učenje je psihološko;
2. možgani/razum so socialni;
3. iskanje smisla je prirojeno;
4. iskanje smisla poteka skozi iskanje vzorca;
5. iskanje vzorca vključuje čustva;
6. možgani/razum delujejo s posamičnimi deli in kot celota hkratno;
7. učenje vključuje oboje, osredotočeno pozornost in periferno zaznavanje;
8. učenje je oboje, zavestno in nezavedno;
9. obstajata vsaj dva pristopa k spominom: arhiviranje izoliranih dejstev in veččin (deklarativni spomin) ter osmišljanje izkušenj (pretežno epizodični spomin);
10. učenje je razvojno;
11. učenje je lahko ovirano zaradi ogroženosti, povezane z nemočjo in/ali utrujenostjo;
12. vsaki možgani so edinstveno organizirani.

Takšno zaporedje načel ni nujno; pomembno pa je, da vsako načelo podpira in se povezuje z vsemi drugimi. Zato se jih pogosto prikazuje v krogu. (Caine, Caine, 1991)

Od razumevanja naravnega učenja do razumevanja inovativnega poučevanja

Preden se lotimo poučevanja je potrebno imeti trdno teoretično podlago o tem, kako se ljudje učijo, vendar pa se zastavlja vprašanje, kako to v praksi narediti? Izhodiščni odgovor bi lahko bil, da ima vsako načelo nekaj praktičnih posledic. Na primer, če je učenje psihološko, potem je učenje izboljšano, ko učenci uporabljajo svoja čutila in ukrepajo. To je le eden od razlogov, zakaj je poučevanje na podlagi vedenja o delovanju možganov tako uspešno.

Računalniška nevroznanost

Računalniška nevroznanost, kot je bilo omenjeno že v teoretičnem delu, proučuje delovanje možganov v smislu lastnosti procesiranja informacij v strukturah, ki sestavljajo živčni sistem. (Schwartz, 1990) Je interdisciplinarna veda, ki povezuje različna področja nevroznanosti, kognitivne znanosti in psihologije z inženirstvom, računalništvom, matematiko in fiziko. Računalniška nevroznanost se nekoliko razlikuje od psihološkega konekcionizma in teorij učenja, kot so strojno učenje, nevronska mreža in teorija računalniškega učenja, predvsem v tem, da poudarja opise funkcionalnih in biološko realističnih nevronov (in nevronskih sistemov/mrež) ter njihovo fiziologijo in dinamiko. Ti modeli zajamejo bistvene značilnosti biološkega sistema na več prostorsko-časovnih lestvicah od membranskih tokov, beljakovin in kemičnega spajanja do mrežnih nihanj, stebraste in topografske arhitekture ter učenja in spomina. Ti računalniški modeli se uporabljajo za oblikovanje hipotez, ki jih je mogoče neposredno preizkusiti z različnimi biološkimi in/ali psihološkimi poskusi. (Chklovskii, 2004)

3.5 Sklep

V tem poglavju smo obravnavali kognitivno znanost in povezavo le-te s kognitivno teorijo učenja ter s procesi in modeli procesiranja informacij. Učenje in pridobivanje novih znanj in veščin je poleg zavesti in svobodne volje eden od ključnih dejavnikov, ki ločujejo inteligentne od neinteligentnih sistemov. Zato smo se v tem poglavju posvetili predvsem metodam pridobivanja novih znanj s stališča, kako kognitivna nevroznanost obravnava učenje in spomin. V nadaljevanju pa smo se osredotočili na kognitivne arhitekture in se posvetili možnostim njihovega modeliranja. Pri tem smo poudarjali, da je tako kot nevroznanost tudi kognitivna znanost izrazito interdisciplinarna, in pogled s samo enega zornega kota, iz zornega kota ene posamične vede, ne more dati pravih rezultatov, najpogosteje pa daje celo popolnoma napačne in zavajajoče rezultate. Ne zadostuje, da imamo neka znanja, da imamo možnosti empiričnih raziskav, temveč je ključnega pomena razumevanje vseh pojavov.

Vsi se učimo skozi celotno življenje zavestno ali nezavedno. Vsak človek se rodi z možgani, ki delujejo kot izjemno močan procesor in je temeljni organ za človeško učenje, ki deluje po nekakšnih, sebi lastnih principih. Tradicionalno šolanje pa

pogosto zavira učenje z odvrčanjem, ignoriranjem ali kaznovanjem naravnih učnih procesov možganov. Učenje na podlagi vedenj o delovanju možganov se imenuje *kombinacija znanosti o možganih in zdrave pameti*. Morris et. all. (Morris, Filenz, 2003, 2007) je možgane imenoval "organ učenja". Zagovarjal je učenje, ki mora temeljiti na principih, kako se proces učenja izvaja v možganih samih, in temu procesu je potrebno oblikovati učinkovita in primerna učna okolja. Caine in Caine (1991) sta razvila dvanajst zgoraj omenjenih načel, ki uporabljajo vedenja o delovanju možganov pri poučevanju in učenju. Ta načela so izpeljana iz raziskovanja znanstvenikov in raziskovalcev številnih disciplin in se obravnavajo kot okvir za razmišljanje o sodobni in inovativni metodologiji poučevanja. Za kompleksno učenje pa sta Caine in Caine izpostavila tri temeljne pogoje:

1. *sproščena budnost* – stanje duha z majhno grožnjo, velikim izzivom, npr. z metodami problemskega ali raziskovalnega pouka;
2. *celovita motivacija* – večkratna, kompleksna, pristna izkušnja;
3. *aktivno procesiranje* - ustvarjanje pomena skozi procesiranje izkušenj, kjer je učenec postavljen v središče dogajanja in mu je omogočeno, da si lahko pridobiva sebi lastne izkušnje.

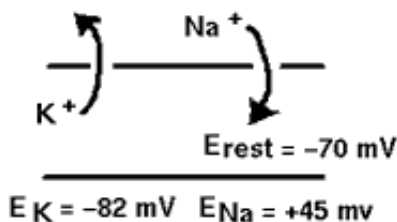
4 Nevroznanost

Nevroznanost je znanost o možganih in procesih, ki se dogajajo v možganih. Študij možganov združuje znanstvenike različnih disciplin, od filozofov, biologov, molekularnih biologov, razvojnih biologov, psihologov – tako analitičnih kot eksperimentalnih – pa fiziologov, farmakologov, antropologov, skratka le malo je akademskih disciplin, ki tako ali drugače ne sodelujejo v nevroznanosti.

V naših glavah je organ, težak približno kilogram in pol, sestavljen iz milijarde drobnih celic. Omogoča nam zaznavo sveta okrog nas, razmišljanje in govorjenje. Človeški možgani so najbolj kompleksen organ v telesu, vendar so le del celote, ki jo v telesu predstavlja živčni sistem. (Morris, Filenz, 2003, 2007) Na začetku, preden se lotimo samih možganov, pa si na kratko oglejmo celoten živčni sistem.

4.1 Nevroni

Nevron je celica. In kot večina celic ima tudi ta lipidno dvoslojno membrano. Značilnost te membrane je, da je polprepustna: večino časa ščiti to, kar je znotraj celice, pred tem, kar je zunaj nje, vendar pa občasno, pod določenimi pogoji dovoljuje tudi pretok skozi membrano, kar je prikazano na sliki 4.1.



Slika 4.1: Ionska osnova membranskega potenciala nevrona. Uravnotežen potencial K in Na ionov je nekaj kot mirujoči potencial, K ioni težijo k temu, da bi zapustili celico, Na ioni pa želijo vstopiti

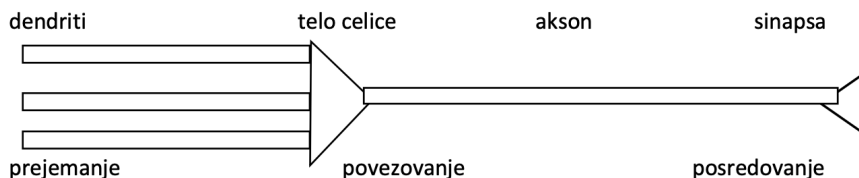
V svojem mirujočem stanju ločuje membrana nevrona električno nabite ione (prvenstveno Ca, K, in Na ione) tako, da ima notranjost nevrona, v primerjavi z zunanostjo negativni naboj. Ta razlika v naboju je znana kot mirujoč potencial membrane E_{rest} (groba ocena, - 70 mV).

Ko je nevron stimuliran, se pojavijo dramatične spremembe v naboju. Membrana postane visoko propustna za pozitivne nabite ione (K, Na), ki vstopajo v celico in s tem povzročajo nenadno povečanje potenciala membrane (tudi do + 35 mV). Ko dosežemo določen prag, se ta proces nadaljuje po celotni dolžini nevrona. To vzpostavi možnost komunikacije električnih komponent od nevrona do nevrona (tu so prisotne tudi kemične komponente). Nenadna sprememba membranskega potenciala (od - 75 mV do + 35 mV) je znana kot *akcijski potencial*. Z ustreznimi napravami lahko vstavimo elektrode v nevron in beležimo ta akcijski potencial, prikazan v grafu na sliki 3.8. Ta graf prikazuje nenaden porast membranskega potenciala, ki mu sledi oster upad in perioda obnovitve stanja.

Ne glede na to, ali so nevroni senzorični ali motorični ali so veliki ali majhni, imajo vsi skupno značilnost, da je njihova aktivnost tako električna kot tudi kemična. Njihova osnovna naloga je nadziranje celotnega stanja živčnega sistema. Kemični signali, ki jih dendriti prejemajo od aksonov, ki so z njimi v stiku, se pretvarjajo v električne signale, ki se prištejejo ali odštejejo od električnih signalov iz vseh ostalih sinaps. Način odločanja, ali bo signal prenesen naprej ali ne, je odvisen od tega, kdo je zmagovalec. Če je končna odločitev, da se signal prenese naprej, električni potencial potuje po aksonu do sinaps, kjer se pretvori v kemični signal, ki gre naprej do dendritov naslednjih nevronov, kjer se proces ponovi.

4.1.1 Dinamični nevron

Struktura nevrona odseva funkcionalno razdelitev nevrona v sprejemni, povezovalni in posredovalni predelek, kar prikazuje naslednja slika 4.2. (Morris, Filenz, 2003, 2007)



Slika 4.2: Shematski prikaz zgradbe in delovanja nevrona

Način organizacije, ki ga prikazuje slika 4.2, imenujemo *polarizacija*, ker informacije potujejo le v eno smer.

Kot vsaka struktura mora tudi nevron imeti nekakšno ogrodje. Zunanje *membrane* nevrona, sestavljene iz lipidnih snovi, so napete okoli *citiskeleta*, ki je zgrajen iz paličic, tubularnih in vlaknastih (filamentoznih) beljakovin, ki se raztezajo v dendrite in aksone. Struktura je podobna šotorskemu platnu, ki je napeto okoli cevastega ogrodja šotora. Različni deli nevrona so v neprestanem gibanju, v procesu reorganizacije, ki odseva tako nevronovo lastno aktivnost kot tudi aktivnost vseh sosednjih nevronov. Dendriti spreminjajo obliko, ustvarjajo nove povezave in opuščajo stare, iz aksonov rastejo novi končiči, ko se nevroni borijo za prevlado.

V nevronu je mnogo notranjih *predelkov*. Te sestavljajo beljakovine, ki so večinoma narejene v telesu nevrona in se do predelkov prenesejo po citoskeletu. Iz dendritov se iztezajo drobni izrastki, imenovani *dendritski trni*. Končiči aksonov se odzivajo na molekule, ki jih imenujemo *rastni faktorji*.

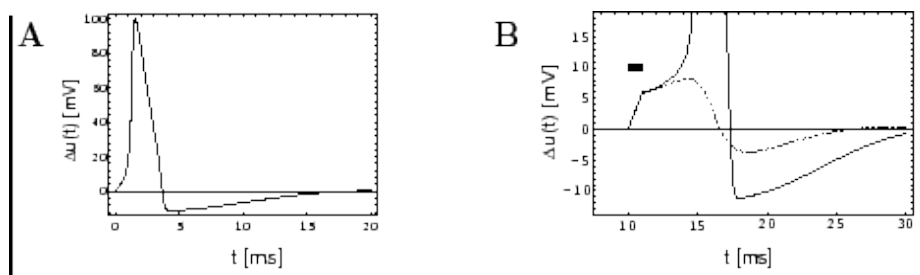
Na sprejemni strani celice so dendriti tesno povezani z aksoni sosednjih celic. Od njih jih ločijo le miniaturne špranje, široke kakšnih 20 milijardink metra. Z dendritom je lahko povezan samo en ali pa tudi na tisoče drugih nevronov. Te točke povezave imenujemo *sinapse*. Beseda izvira iz antične grščine in pomeni »spet skupaj«. Večina sinaps s celic možganske skorje je na trnih dendritov, ki omogočajo prenos signalov.

Komunikaciji med živčnimi celicami na teh kontaktnih točkah pravimo sinaptični prenos in vključuje kemični proces. Ko do dendrita pride eden od kemičnih obveščevalcev skozi špranjo med dendritom in aksonom, se v dendritskem trnu sproži miniaturni električni tok. Tokove, ki prihajajo v celico, imenujemo *ekscitacijske*, tokove, ki celico zapuščajo, pa *inhibicijske*. Če se ti tokovi v kratkem času ne seštejejo, kmalu zamrejo in se ne zgodi nič. Če pa se ti tokovi seštejejo do vrednosti, ki presega prag, nevron pošlje sporočilo ostalim nevronom.

4.1.2 Akcijski potencial

Aksoni nevronov prenašajo električne pulze, imenovane *akcijski potencial* (slika 4.3.A). Ti potujejo vzdolž živčnih vlaken kot val, zato ker membrana vsebuje *ionske kanalčke*, ki se odpirajo in zapirajo in na ta način nadzirajo pretok električno nabitih ionov.

Ko se akcijski potencial v telesu celice začne, se najprej odprejo Na^+ kanalčki. Pulz natrijevih ionov vdre v celico in že v milisekundah se ustvari nov gradient. V trenutku se napetost spremeni za okrog 100 mV. Preskoči iz negativne napetosti znotraj membrane (okrog -70 mV) na pozitivno (+30 mV). Ta sprememba odpre K^+ kanalčke, kar sproži pulz K^+ ionov iz celice, ki je skoraj tako hiter, kot je bil prej vdor Na^+ ionov. To povzroči, da se ionski potencial membrane ponovno obrne in se vrne na svojo prvotno negativno vrednost v notranjosti celice. Da se akcijski potencial zgodi, prečka membrano zelo malo ionov in koncentracija K^+ in Na^+ ionov v citoplazmi se pri tem ne spremeni veliko. Vseeno pa *ionske črpalke* vzdržujejo ravnotežni koncentraciji K^+ in Na^+ ionov, in sicer tako, da iz celic prenašajo odvečne notranje ione.



Slika 4.3:

A. Akcijski potencial

B. Učinek praga pri iniciaciji akcijskega potenciala

Akcijski potencial je kompleksen električni dogodek. Živčna vlakna delujejo kot električni vodniki. Tako akcijski potencial, ki se zgodi na enem mestu, sproži nov gradient napetosti med aktivno in mirujočo membrano v neposredni bližini. Na ta način se akcijski potencial širi in val depolarizacije se premika z enega konca živčnega vlakna na drugi konec. V nekaterih aksonih potuje akcijski potencial počasi, v drugih pa kot bi trenil. To se zgodi, ker so dolgi odseki aksonov oviti z maščobami, t. i. izolacijo, imenovano mielinska ovojnica, ki je narejena iz ovoja raztegnjenih membran glije.

Za akcijski potencial je značilno, da gre pri njem za vse ali nič, kar pomeni, da je kodiranje moči ali trajanje signala v posamični celici odvisno le od spreminjanja frekvence akcijskega potenciala, ki lahko niha s frekvenco tudi do 1000-krat na sekundo. (Morris, Filenz, 2003, 2007)

4.1.3 Od električnih do kemijskih reakcij

Akcijski potenciali se prenašajo po aksonih do specializiranih regij, imenovanih sinapse, kjer aksoni vzpostavijo stik z dendriti drugih nevronov. Sinapse so sestavljene iz presinaptičnega živčnega končiča, postsinaptične komponente, ki je pogosto na dendritskem trnu, in drobne sinaptične špranje, ki ju ločuje. Električni tok, ki je odgovoren za potovanje akcijskega potenciala vzdolž aksona, ne more premostiti sinaptične špranje. Prenos preko špranje izvršijo kemični obveščevalci, imenovani *nevrotransmitorji* ali *živčni prenašalci*. (Morris, 2003, 2007) Živčni prenašalci ali nevrotransmitorji so skupina heterogenih biokemičnih snovi, ki prenašajo informacije med nevroni preko sinaps.

SHRANJEVANJE IN SPROŠČANJE

Nevrotransmitorji so shranjeni na koncih aksonov (v živčnih končičih) v drobnih kroglastih vrečkah, imenovanih *sinaptične vezikule*. Obstajajo vezikule za shranjevanje in vezikule bližje membrani živčnega končiča, ki so pripravljene, da sprostijo svojo vsebino. Prihod akcijskega potenciala v presinaptično področje sproži odprtje ionskih kanalčkov, ki spustijo v končič *kalcijeve ione* (Ca^{2+}). Kalcij aktivira encime, ki delujejo na paleto presinaptičnih beljakovin z eksotičnimi imeni kot snare, tagmin in brevin. Nevroznanstveniki so pravkar odkrili, da te presinaptične beljakovine tekmujejo pri označevanju in lovljenju ostalih beljakovin, kar povzroči, da se

vezikule, pripravljene, da sprostijo svojo vsebino, zlijejo z membrano, se odprejo in sprostijo kemični obveščevalec iz živčnega končiča. Ti obveščevalci nato pronicajo skozi 20 nanometrov široke *sinaptične špranje*.

Elektrokemični val, imenovan tudi akcijski potencial, potuje vzdolž aksona nevrona. Ko val doseže sinapso, sproži sprostitvev nevrotransmisijskih molekul, ki se nato vežejo na molekule, ki se nahajajo v membranah drugega nevrona, na nasprotni strani sinapse.

Sinaptične vezikule se tvorijo ponovno, ko njihove membrane vsrkajo vsebino (encime) nazaj v živčni končič, kjer se ponovno napolnijo z nevrotransmitorjem za naslednja sproščanja v neprekinjenem reciklirajočem se procesu.

Ko prispe nevrotransmitor na drugo stran sinaptične špranje, kar se zgodi v manj kot milisekundi, se spoji s posebnimi beljakovinskimi strukturami na membrani naslednjega nevrona, imenovanimi *receptorji*. Tudi celice glije obdajajo sinaptično špranjo. Nekatere imajo stalno pripravljene miniaturne sesalce, imenovane *transporterji*. Njihova naloga je posesati nevrotransmitorje iz špranje. To očisti kemične obveščevalce iz špranje pred prihodom naslednjega akcijskega potenciala. Ampak nič se ne zavrže – celice glije obdelajo transmitor in ga pošljejo nazaj, da se shrani v shranjevalne vezikule živčnih končičev za nadaljnjo uporabo. Nevrotransmitorji pa se ne očistijo iz sinapse samo z gospodinjскими opravili celic glije. Včasih nevron sam črpa molekule nevrotransmitorja nazaj v svoje živčne končiče. Drugod pa razgradijo nevrotransmitor encimi v sinaptični špranji.

OBVEŠČEVALCI ODPIRAJO IONSKE KANALČKE

Interakcija živčnih prenašalcev z receptorji je podobna tisti med *ključem in ključavnico*. Vezava nevrotransmitorja (ključa) na receptor (ključavnico) po navadi povzroči odprtje ionskega kanalčka; take receptorje imenujemo *ionotropne receptorje*. (Morris, Filenz, 2003, 2007)

V glavnem razdelimo živčne prenašalce na *ekscitatorne* in *inhibitorne*. Prvi spodbujajo širjenje akcijskega potenciala po nevronih in povzročajo depolarizacijo celične membrane, drugi pa zavirajo potovanje potenciala po nevronih in membrano hiperpolarizirajo. Najpomembnejši ekscitatorni prenašalec v osrednjem živčnem

sistemu je glutamat, najpomembnejši inhibitorni pa γ – aminomaslena kislina (GABA) v osrednjem živčevju in glicin v okrajnem živčevju.

Če ionski kanalček omogoča vstop pozitivnim ionom (Na^+ ali Ca^{2+}), vstop pozitivnega toka povzroči vznburjenje (ekscitacijo). Ta povzroči spremembo membranskega potenciala, imenovano *ekscitatorni postsinaptični potencial (epsp)*. Po navadi je na nevronu veliko sinaps in v vsakem trenutku so nekatere aktivne in druge ne. Če vsota teh potencialov (epsp) doseže prag za proženje impulza, se ustvari nov akcijski potencial, ki potuje po aksonu sprejemnega nevrona. (Morris, 2003)

Glavni ekscitatorni nevrottransmitor v možganih je *glutamat*. Velika natančnost živčne aktivnosti zahteva, da ekscitacijo nekaterih nevronov spremlja zavrito delovanje (supresija aktivnosti) drugih. To omogoča *inhibicija*. V *inhibitorni sinapsi* aktivacija receptorjev povzroči odprtje ionskih kanalčkov, ki tako dovoljujejo vstop negativno nabitih ionov v celico, kar povzroči spremembo v membranskem potencialu, imenovano *inhibitorni postsinaptični potencial (ipsp)*. Ta sprememba nasprotuje depolarizaciji membrane in torej sprožitvi akcijskega potenciala v telesu sprejemnega nevrona. Obstajata dva inhibitorna nevrottransmitorja, *GABA* in *glicin*.

Sinaptični prenos je zelo hiter proces: čas od prihoda akcijskega potenciala do sinapse in sprožitve epsp v naslednjem nevronu je zelo kratek: 1/1000 sekunde. Različni nevroni morajo uskladiti svoje izločanje glutamata sprejemnemu nevronu v zelo ozkem časovnem oknu, če se hočejo njihovi epsp sešteti in sprožiti nov impulz. Prav tako mora tudi inhibicija delovati v istem časovnem intervalu, če hoče uspešno zadušiti proženje impulza. (Morris, Filenz, 2007)

4.2 Kognitivna nevroznanost

Medtem ko kognitivna znanost v najširšem pomenu zajema vse oblike znanstvenega preučevanja duha in inteligence, se kognitivna nevroznanost bolj osredotoča na vprašanje, kako se z delovanjem možganov poraja človeška duševnost. Nastanek kognitivne nevroznanosti lahko razumemo kot logičen rezultat razvoja psihologije in nevroznanosti. Vprašanje odnosa med možgani in duhom je zelo star problem, a šele teoretični napredki na eni ter tehnični razvoj na drugi strani so ob koncu 70. let 20. stoletja omogočili oblikovanje multidisciplinarne kognitivne nevroznanosti ter jo v naslednjih treh desetletjih razvili v eno najbolj ustvarjalnih in hitro razvijajočih se

ved moderne dobe. (Repovž, 2007) Za celostno razumevanje kognitivne znanosti je izrednega pomena upoštevanje njenih dveh bistvenih lastnosti (Miščević, Markič, 1998, Markič, 2010):

1. kognitivno znanost ob njenem predmetu preučevanja ključno zaznamuje njena metoda preučevanja, skratka njena kognitivna paradigma;
2. kognitivna znanost je multidisciplinarna veda, ki mora za uspešno razlago mentalnih pojavov vključevati več pristopov k preučevanju ter združevati več metodologij opazovanja in razlaganja predmeta preučevanja.

KOGNITIVNA NEVROZNANOST NEKOČ IN DANES

Zametki kognitivne znanosti segajo v trideseta in štirideseta leta 20. stoletja kot odgovor na pomanjkljivosti in končni neuspeh takrat prevladujoče behavioristične paradigme v psihologiji. Vedenjska psihologija je na podlagi prevladujočega pozitivističnega pogleda na znanost izhajala iz predpostavke, da so le pojavi, ki jih neposredno opazujemo in merimo, lahko veljaven predmet znanstvenega proučevanja. Da bi lahko polno razumeli človeško vedenje, je potrebno razumeti izračune in pretvorbe informacij, ki potekajo v umu. Poznavanje notranjih predstav podatkov in procesov, ki nanje delujejo, je ključno za razumevanje človeškega vedenja in doživljanja. Zaznavanje, spomin, pozornost, reševanje problemov in drugi kognitivni procesi so tako postali nujni predmet preučevanja psihologije. (Markič, 2010) Za uspešen vzpon kognitivne nevroznanosti sta bila ključnega pomena dva miselna preskoka. Prvi je bil povezan z razvojem računske znanosti in računalnikov. Drugi miselni preskok je predstavljalo spoznanje, da je možno z ustrezno oblikovanimi eksperimenti opazovati tudi tisto, kar je sicer očem skrito. (Markič, Bregant, 2007)

Osrednjo strategijo ustvarjanja in preverjanja modelov nevrokognitivnega sistema s pomočjo vedenjskih preizkušenj je začela kognitivna psihologija razmeroma zgodaj dopolnjevati z informacijami iz dveh sorodnih virov – računalniške znanosti in nevropsihologije. Prva je predstavljala logični korak v razvoju modelov kognitivnih procesov, eksplicitnih matematičnih modelov, ki dejansko lahko izvajajo procese, ki jih želimo razložiti, druga, nevropsihologija, pa je oblikovala spoznanje, da poškodbe različnih predelov možganov vodijo do različnih motenj kognitivnih funkcij. To spoznanje je povzročilo dve pomembni posledici. Prvič, potrdilo je predpostavko,

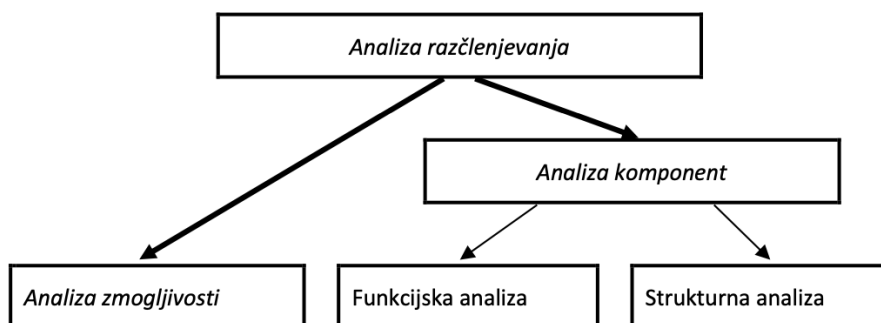
da je človeški um organiziran v obliki relativno samostojnih kognitivnih modulov, od katerih vsak opravlja svojo funkcijo v procesiranju informacij, le-te pa prehajajo od modula do modula po točno določenih medsebojnih povezavah. Drugič, s pomočjo študija vzorcev kognitivne poškodovanosti pri različnih poškodbah je bilo mogoče preverjati obstoječe modele kognitivnih procesov ter oblikovati ključne kriterije, ki jim morajo modeli ustreči. Bogastvo in pomen informacij, ki jih je podajala nevropsihologija za oblikovanje in preverjanje teoretičnih modelov kognicije, sta vodila do oblikovanja kognitivne nevropsihologije kot pomembne discipline v okviru kognitivne znanosti. Ob tem pa se je razvijala tudi nevroznanost. Nevroznanost je v času, ko je psihologija razvijala kognitivni pristop ter oblikovala in preverjala modele kognicije neodvisno od njene materialne osnove, postopoma gradila znanje o anatomiji in fiziologiji možganov.

Do ključnega premika v odnosu med kognitivno psihologijo in nevroznanostjo je prišlo z razvojem modernih metod opazovanja in beleženja strukture in aktivnosti možganov, kot so večkanalna elektroencefalografija (EEG), magnetoencefalografija (MEG), enofotonska izseвна računalniška tomografija (SPECT), pozitronska izseвна tomografija (PET) in funkcijsko magnetnoresonančno slikanje (fMR). Naštete tehnike so omogočile kompleksen celostni vpogled v aktivnost možganov, hkrati pa nevroznanost ni imela konceptualnih orodij, ki bi omogočala njihovo smiselno uporabo.

4.2.1 Kognitivna nevroznanost in vzratno inženirstvo

Pristop kognitivnega znanstvenika k preučevanju delovanja možganov in kognicije lahko primerjamo s teorijo *vzratnega inženirstva*. Osnovni princip tega pristopa je razumeti, kako deluje nek tehnični sistem. Pristop je analitičen in se pri njem uporablja strukturna in funkcionalna analiza. Vseskozi ustvarja teoretične in praktične modele, s katerimi skuša razumeti in razložiti delovanje izdelka in pogosto se zanese tudi na izračune in računske simulacije. Končni namen teorije vzratnega inženirstva je spoznati, kateri so sestavni deli, kako delujejo in kako so sestavljeni, da celota uspešno deluje. Končni cilj je delovanje razumeti do te mere, da je sposobna enak ali podoben tehnični sistem tudi poustvariti. Ker mnogokrat tehničnega sistema ne moremo v popolnosti analizirati, se v zadnjem času za te namene uporablja tudi *prediktivno vzratno inženirstvo*. Prav takšno je tudi delo kognitivnega nevroznanstvenika. (Repovž, 2007)

Ena od ključnih lastnosti pristopa kognitivne nevroznanosti je, da se ne zadovoljuje z iskanjem opisnih zakonov, ki bi podrobno opisovali lastnosti in vedenje sistema v različnih pogojih, temveč skuša preko analize sestavnih delov sistema in njihove interakcije razložiti, kaj in kako določa njegove ugotovljive lastnosti in vedenje. Strategijo razlaganja, ki jo ubira kognitivna nevroznanost, Atkinson (1998) imenuje analiza razčlenjevanja in je prikazana na sliki 4.4.



Slika 4.4: Vrste analize razčlenjevanja

V okviru kognitivne nevroznanosti se funkcijska analiza posveča opredeljevanju temeljnih kognitivnih procesov in odnosov med njimi, strukturna analiza pa se posveča opredeljevanju anatomskih komponent centralnega živčevja.

4.2.2 Nevrokognitivna znanost in nevromorfnost inženirstvo

Živi sistemi so sposobni izvajati izjemno kompleksna procesiranja signalov in informacij. V primerjavi z digitalnimi računalniki živi sistemi porabljajo veliko manj energije, so visoko adaptivni in sposobni izvajati vrsto kompleksnih, vzporednih opravil. Živi sistemi so visoko nelinearni.⁴⁰ Poskus uporabe matematike, ki naj usmerja vedenje relativno enostavnih nevronske mreže, povzroča potrebo po ogromni količini računalniške moči in časa simulacije – npr. IBM-ov Big Blue

⁴⁰ Nelinearni sistemi so sistemi, katerih vrednosti na izhodu (rezultati) niso neposredno sorazmerne z vrednostmi na vходу, skratka, ne moremo jih opisati z linearno funkcijo. Z drugimi besedami, nelinearni sistem je katerikoli problem, kjer spremenljivke, ki jo rešujemo, ne moremo zapisati kot linearno kombinacijo neodvisnih komponent. Ti problemi so zanimivi tako za matematike kot tudi za fizike in inženirje, saj je večina fizikalnih sistemov v naravi nelinearnih. Reševanje nelinearnih enačb je zelo težavno, zato se mnogokrat uporablja tudi teorija kaosa, o čemer bomo govorili v nadaljevanju.

superračunalnik je še ravno dovolj zmogljiv, da lahko z njim simuliramo delovanje majhnega dela možganov mačk, in to ne v realnem času. Vsaka nadaljnja beseda je pri tem odveč⁴¹.

Posledica tega je, da se je skoraj pred tridesetimi leti rodilo gibanje *nevromorfno inženirstvo*, ki je pričelo iskati drugačno pot za reševanje problemov modeliranja in posnemanja živih sistemov. Namesto digitalnega kodiranja so vpeljali analogna vezja, ki upoštevajo informacijo s pomočjo kontinuiranih sprememb v napetosti, prav tako kot živčne celice v svojem podpraznem stanju. Izračune je tako možno opraviti v manj korakih, s pomočjo izkoriščanja osnovne fizike silicijevih naprav. Analogno računanje enostavno podaja osnovne elemente aritmetike: seštevanje, odštevanje, eksponente in integracije; operacije, ki so sicer precej zahtevne za digitalne stroje. Kadar nevroni – naj bodo biološki ali silicijevi – računajo in sprejemajo odločitve, pošiljajo impulze vzdolž aksonov ter komunicirajo s ciljnim nevroni. Ker je kodiranje z akcijskimi potenciali energijsko drago, učinkovito kodiranje povečuje količino informacij, zabeleženih v vzorcu potencialov, z zmanjšanjem *redundance*. Energijsko učinkovitost hkrati povečujemo tudi z uporabo čim manjšega števila aktivnih nevronov. Slednje imenujemo *redko kodiranje* in predstavlja še enega od pomembnih načel pri grajenju umetnih nevronske mreže. (Morris, Filenz, 2003)

Nevromorfno inženirstvo, včasih tudi nevromorfno računalništvo, je koncept, ki uporablja sistem z zelo veliko stopenjske integracije in vsebuje elektronske analogne kroge, s katerimi simulira nevrobiološko arhitekturo živčnega sistema. Ključno pri razumevanju nevromorfne inženirstva je, kako morfologija posameznih nevronov, integriranih vezij in celotne arhitekture vpliva na predstavo informacij, na razvoj in evolucijske spremembe in kako vključuje učenje. Nevromorfni inženiring je nova interdisciplinarna disciplina, ki združuje biologijo, fiziko, matematiko, računalništvo in inženirstvo, z namenom ustvariti umeten nevronske sistem (sistem vida, govorni sistem ali avtonomne robote), pri katerem temelji dejanska fizična arhitektura in konstrukcija na bioloških živčnih sistemih. (Rothschild, 2010)

V svojih približkih so matematične modele, ki opisujejo živčne in senzorne funkcije živih sistemov v nevromorfne inženirstvu, zamenjali s fizikalnimi modeli, izdelanimi kot posebna silicijeva integrirana vezja oz. čipi. Obstaja kar nekaj

⁴¹ Za več o tem glej poglavji 6 in 7.

razlogov, zakaj so takšne platforme modeliranja superiorne pred konvencionalnim modeliranjem z uporabo digitalnih računalnikov. Če naštejemo le dva:

- prvič, fizikalne karakteristike silicijevih struktur so analogne z mnogimi karakteristikami živih sistemov, od katerih, vemo, je mnogo vedenjskih oblik tudi nelinearnih;
- drugič, pri konstruiranju fizikalnih sistemov lahko posamezne funkcije nevronske mreže in senzorske sisteme "simuliramo" v realnem času.

Vrsta različnih elektronskih komponent je sposobna posnemati delovanje možganskih celic. *Memristorji*, *silicijeve nanožičke* in *nanocenke* so le nekatere, na katere so se osredotočili znanstveniki pri sintezi umetnih nevronske mreže. V različnih raziskovalnih skupinah potekajo različne raziskave; na Univerzi Stanford, eni od vodilnih raziskovalnih inštitucij s področja nevroznanosti, so npr. razvili tako programsko kot tudi strojno opremo (čipe), ki je neprimerno bližje umetnim možganom, kot je klasična centralno procesorska enota – CPE (angl. central processor unit ali CPU). Lep primer nevromorfne računalniške strojne opreme je *Nevromreža*, ki jo je razvila skupina *Brains in Silicon* na omenjeni Univerzi v Stanfordu in je prvič pričela delovati v drugi polovici leta 2009.⁴² Nevromreža je del računalniške strojne opreme, posebej konstruirane za simuliranje živih možganov. Uporablja analogno preračunavanje, s katerim simuliramo aktivnosti ionskih kanalov, in digitalno komuniciranje s povezovalnimi vzorci. Nevromreža simulira delovanje 10^6 nevronov in 6×10^7 sinaps v realnem času. Nevroni se nabijejo desetkrat na sekundo. To pomeni resno konkurenco simulacijam izvedenim v projektu Blue Brain (glej poglavje 7), vendar se pri tem porabi manj kot 2 W električne energije. (Dethier, et al., 2011) Osnovna plošča nevromreže je sestavljena iz šestih nevro jeder, vsako izmed njih pa ima 256×256 silicijevih nevronov v 11.9×13.9 mm velikem čipu. Offčip RAM in ončip RAM (v vsakem nevro jedru) sta navpično in vodoravno povezana. Z 61 stopenjskim in 18 binarnimi programabilnimi parametri, ki so skupni vsem silicijevim nevronom, lahko nevro jedro modelira raznolikost vzorcev.

⁴² <http://www.stanford.edu/group/brainsinsilicon/pubs.html>

Nevromorfno inženirstvo bi lahko razdelili na dve skupini:

- nevromorfno modeliranje in
- nevromorfno preračunavanje.

Nevromorfno modeliranje je ozko povezano z nevropsihološkimi fenomeni. S tem širimo znanje in razumevanje o živčnem sistemu in ustvarjamo potrebna znanja o bioloških algoritmihi. *Nevromorfno preračunavanje* je osredotočeno na vprašanje, kako na osnovi nevroloških lastnosti ustvariti nevromom podobno računalniško strojno opremo, s katero bi lahko simulirali biološke procese. Skratka, kako algoritme prevesti v računalniške tokokroge.

4.3 Nevroznanost, učenje in spomin

Spomini so ključni za našo individualnost. Češar se spominjamo sami, je vedno različno od spomina drugih ljudi, tudi če smo isti dogodek doživeli skupaj. Vsi pa se, seveda po svoje, spomnimo dogodkov, dejstev, čustev in različnih veščin, nekaterih le za trenutek, nekaterih za dlje časa, nekaterih pa za vse življenje. V možganih najdemo več spominskih sistemov z različnimi lastnostmi, ki jih podpirajo različna bionevronska omrežja. Obstaja vrsta teorij, kako nastanejo novi spomini, npr. zgoraj opisana informacijska teorija na sliki 4.1. Po nevroznanstveni teoriji pa nastanejo novi spomini s pomočjo procesa sinaptične *plastičnosti*, niso pa še znani mehanizmi, ki omogočajo priklic teh spominov. (Morris, Filenz, 2003)

4.3.1 Plastičnost

Možgani se med življenjskim razvojem neprestano spreminjajo. Njihovo sposobnost spreminjanja imenujemo *plastičnost*. (Morris, Filenz, 2003) Ne gre za spreminjanje možganov kot celote, pač pa za spreminjanje posamičnih nevronov zaradi različnih vzrokov, npr. zaradi razvoja v mladosti, med učenjem ali pa tudi zaradi različnih poškodb. Obstajajo različni mehanizmi plastičnosti glede na to, kako nevroni spreminjajo svojo sposobnost komunikacije med seboj.

OBLIKOVANJE PRIHODNOSTI

Povezave med nevroni se uglasijo pri ljudeh že v zgodnji mladosti. Sinaptične povezave se začnejo spreminjati, ko se odzivamo na okolico, ustvarjajo se nove sinapse, pogostejše uporabljane povezave se krepijo, tiste, redkeje uporabljene, pa slabijo ali celo v popolnosti izginejo. Gre za princip *uporabljalj ali izgubi*. S tem principom oblikujemo prihodnost svojih možganov.

Sinaptični prenos vključuje sproščanje kemičnih nevrottransmitorjev, ki nato aktivirajo specifične proteine, imenovane *receptorji*. Merilo sinaptične moči je normalni električni odziv na sprostitvev nevrottransmitorja. *Sinaptična moč* se lahko spreminja; sprememba lahko traja nekaj sekund, nekaj minut ali vse življenje. Nevroznanstvenike zanimajo predvsem dolgotrajne spremembe sinaptične moči, ki jih povzročijo kratke periode živčne aktivnosti v dveh procesih, in sicer v:

- *dolgoročni potenciaciji* (angl. *LTP* – long-term potentiation), ki poudari sinaptično moč, in
- *dolgoročnem upadanju* (angl. *LTD* – long-term depression), ki jo slabi.

Glutamat se sprosti iz sinaptičnih končičev, prečka sinaptično špranjo in se veže na različne tipe glutamatnih receptorjev – *AMPA*, *NMDA* in *mGluR*. Nekatere glutamatne sinapse imajo tudi kainatne receptorje.

PRINCIP DELOVANJA

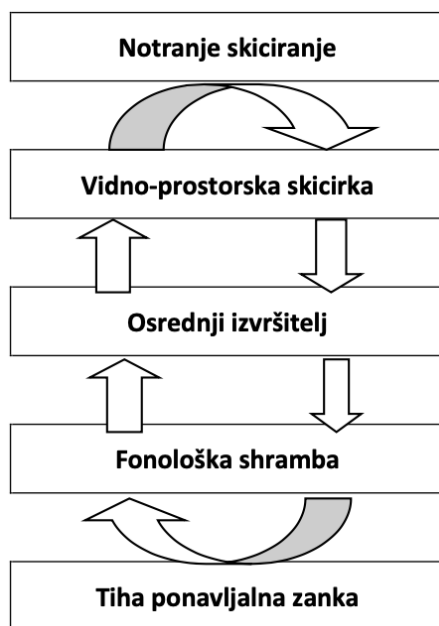
Glutamat je pogosta aminokislina, ki se uporablja po vsem telesu kot gradnik beljakovin. Glutamat je tudi nevrottransmitor, ki deluje na najbolj plastične sinapse, tiste z *LTP* in *LTD*. Glutamatni receptorji, ki so predvsem na sprejemni strani sinapse, obstajajo v štirih oblikah, prvi trije so *ionotropni* receptorji, zadnji pa je *metabotropni* receptor:

- *NMDA* receptorji so molekularni stroj za vzpostavljanje plastičnosti, stroj za učenje; transmitor se sprošča tako med osnovno aktivnostjo kot med sprožitvijo *LTP*;
- *AMPA* predstavlja naš molekularni stroj za shranjevanje spominov;

- *kainatni receptor in*
- *mGluR* (metabotropni glutamatni receptor) je metabotropni receptor.

Čeprav se vsi štirje tipi glutamatnih receptorjev odzovejo na isti nevrottransmitor, opravljajo zelo različne funkcije. Iontropni glutamatni receptorji uporabljajo svoje ionske kanalčke za sprožitev ekscitatornega postsinaptičnega potenciala, medtem ko metabotropni glutamatni receptor nadzoruje velikost in značilnosti nevroregulatornih odzivov. Vse oblike glutamatnih receptorjev so pomembne za plastičnost, toda o receptorjih *AMPA* in *NMDA* vemo največ in jih imamo pogosto za *spominske molekule*.

4.3.2 Organizacija spomina



Slika 4.5: Sistem za kratkoročno hranjenje informacij v delovnem spominu.

V možganih ne obstaja le eno področje, v katero bi shranjevali vse, kar si kadarkoli zapomnimo. *Delovni spomin* je *kratkoročni spomin*, ki aktivno vzdržuje v zavestnem stanju informacije, ki jih v našem razumu hranimo le kratek čas. Dolgoročni spomin predstavlja veliko širše in bolj pasivno skladišče informacij. (Morris, Filenz, 2003)

DELOVNI SPOMIN

Podobno kot človek uporablja beležko, da si zapiše določeno informacijo za kratek čas, ali računalnik uporablja svoj hitri spomin (RAM), imajo tudi možgani sistem, ki omogoča zanesljivo vzdrževanje in delo z manjšim številom informacij. Uporabljamo ga, da si informacijo zapomnimo tako dolgo, da lahko dojamemo tok informacij (npr. pogovora) in v mislih analiziramo te informacije ter si jih (ali pa tudi ne) nato zapomnimo. Zanesljivost je pri tem ključnega pomena, žal pa so zmogljivosti omejene. Velikokrat poudarjamo, da lahko v delovni spomin sprejmemo 7 ± 1 informaciji.

Osrednji izvršilni sistem nadzoruje tok informacij, pri čemer ga podpirata dva dodatna spominska podsistema:

1. *fonološka shramba* sodeluje s *tibo ponavljalno zanko* pri vzdrževanju besednih informacij. Gre za tisti del možganov, ki ga uporabljamo, ko govorimo s seboj. Tudi če besede ali številke preberete zapisane v vidni obliki, bo informacija pretvorjena v fonološki zapis in v tej obliki shranjena v omenjenem sistemu;
2. drugi podsistem je *vidno-prostorska skicirka*, ki vzdržuje slike ali predmete tako dolgo, da lahko z njimi v mislih izvajamo različne manipulacije.

Delovni spomin se v največji meri nahaja v *čelnem temenskem režnju* možganov. (Morris, Filenz, 2003)

DOLGOROČNI SPOMIN

Tako kot ostale vrste spomina se tudi dolgoročni spomin deli na različne sisteme, umeščene v široko razpršenih možganskih omrežjih, ki opravljajo zelo različne naloge. Informacije najprej vstopajo v senzorične sisteme, nato pa se prenašajo v bolj specializirane sisteme za njihovo procesiranje. Vidna informacija na primer iz *striatne* skorje po t. i. *ventralni* poti vstopi v *medialni* del senčnega režnja. Nadalje potuje po omrežjih, v katerih se določajo njene lastnosti, kot so oblika, barva, identiteta in prepoznavnost, vse dokler se ne ustvari spomin tega konkretnega predmeta, vključno z informacijo, kdaj in kje je bil viden. Skratka, v možganski skorji obstajajo področja, ki izluščijo *zaznavne reprezentacije* opazovanega. Z njihovo pomočjo si zapomnimo in kasneje prepoznamo stvari v svoji okolici, npr. prepoznamo znane ljudi (npr. politike) celo v karikaturah. Soroden sistem je tudi *semantični spomin*; ta

hrani vso znanje o dejstvih o svetu, ki smo ga usvojili. Vemo npr., da je Ljubljana glavno mesto Slovenije, da DNK kodira genetsko informacijo v obliki zaporedja baznih parov in tako naprej. Bistvena lastnost semantičnega spomina je, da so dejstva organizirana v kategorije – drevesne diagrame. Slednje je ključno za učinkovit priklic, saj lahko iskalni mehanizem te diagrame hitro in uspešno preiskuje. Če bi bil semantični spomin organiziran naključno, tako kot je organizirana večina stvari v človeškem življenju, bi se marsičesa zelo težko spomnili.

Prav tako se učimo tudi *veščin* in spoznavamo *čustva*. Vedeti, kaj je glasbeni inštrument, je eno, znati igrati nanj pa nekaj povsem drugega. Znati voziti kolo je uporabna veščina, zavedati se, da so ceste nevarne, pa je prav tako pomembno. Veščin se naučimo z zavestnim in aktivnim treningom in traja dlje, medtem ko je s čustvi povezano učenje lahko v nasprotju s tem zelo hitro, še posebej pri čustvih, povezanih s strahom. Oba načina učenja imenujemo *pogojevanje*. V tako učenje so vpletene posebne možganske strukture – pri učenju veščin sodelujejo *bazalni gangliji*⁴³ in *mali možgani*, pri čustvenem učenju pa *amigdaloidno jedro*.⁴⁴ (Morris, Filenz, 2003)

4.4 Celostni pristop k raziskovanju v kognitivni nevroznanosti

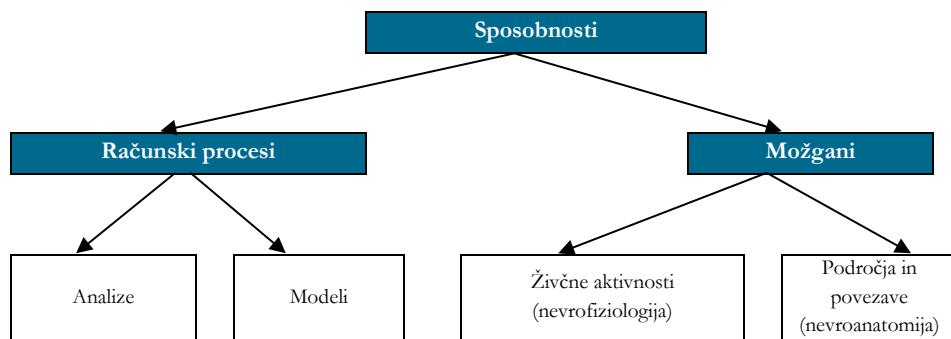
Celostni pristop k preučevanju v kognitivni nevroznanosti izhaja iz predpostavke, da je za razvoj celostnih teorij in modelov delovanja možganov in uma potreben multidisciplinaren pristop, ki ga prikazuje slika 4.6, in usklajevanje vsaj štirih pristopov znanstvenega raziskovanja.

Prvič, podrobno moramo opisati, katere sposobnosti in lastnosti mora model razložiti, ter predlagati funkcionalno arhitekturo možganov in uma, ki to omogoča. Ta cilj si je zadala *kognitivna psihologija*. Drugič, oblikovati je treba delujoče računske modele kognitivnih funkcij, ki na eni strani natančno opredeljujejo predlagane teoretične modele kognitivnih funkcij, hkrati pa se v svoji implementaciji podrejajo znanim zakonitostim delovanja živčnega sistema. To je naloga *kognitivne računske nevroznanosti*. Tretjič, seznaniti se je treba z dejansko anatomijo in fiziologijo

⁴³ Gre za bazalne ganglije (globoka jedra), ki so skupina med seboj povezanih področij, ki se nahajajo pod skorjo v globini možganskih polobel.

⁴⁴ Amigdaloidno jedro: Limbični sistem je del velikih možganov. Gre za prstanasto oblikovan predel v notranjosti možganov, ki obdaja vrhni del medmožganov in možganskega debla. Odgovoren je za nekatere vegetativne funkcije, čustvene reakcije, učenje in spomin. Limbični sistem vključuje več možganskih struktur, med katerimi je tudi amigdaloidno jedro: skupina jeder v temporalnem lobusu.

možganov, da bi lahko razložili, kako so kognitivne funkcije udejanjene v možganih. To je področje *nevroznanosti*. In končno, preučevati moramo kognitivne motnje, ki sledijo okvaram možganov. Le-te predstavljajo pomemben preizkus v teoriji ter vir teoretičnih idej in zahtev. To je naloga *kognitivne nevropsihologije*.



Slika 4.6: Kognitivno nevroznanstveni trikotnik.

Vir: Kosslyn, 1996

4.5 Sklep

V tem poglavju smo se dotaknili področja nevroznanosti, ki se je predvsem zaradi izjemno močnega tehnološkega razvoja na področju empiričnih raziskav in metod zbiranja podatkov (fMRI, EMR, BOLD, ...) v zadnjih desetletjih izrazito intenzivno razvijalo. Danes nam nevroznanost omogoča dokaj natančen vpogled v delovanje možganov, kar je temelj razvoja različnih metod modeliranja procesov in elementov tako celotnega živčnega sistema kot njegovega centralnega dela, to je možganov. Zaključimo lahko, da so možgani v svojem bistvu bionevronska mreža, katere osnovni gradniki so nevroni. Zato njim posvetimo pomemben del tega poglavja in raziskujemo, kako se informacije v možganih (nevronih) prenašajo, kako se signali pretvarjajo od električnih do kemičnih reakcij in kako se ustvarja pri tem ustrezen akcijski potencial, ki definira dostopnost do določenih informacij in verjetnost njihovega priklica v spomin, kar je osnova kakršnegakoli učenja.

5 Simbolni sistemi in konekcioniistični modeli

Simbolno modeliranje duha ima dvojje korenin:

- Prve od teh so v filozofiji. Logičen sistem je sestavljen iz procesov, ki se izvajajo s simboli. Ti simboli lahko predstavljajo stavke in povezave oz. veznike (npr. *in*, *ali*, *če* – *potem*). Splošno so cilji takšnih izvajanj enoznačna definiranja problemov (če *p*, potem *q*, ali stavek je lahko *resničen* ali *neresničen*). Ta perspektiva je splošno znana v logiki kot *modelna teoretična perspektiva*.
- Druge, znane kot *dokazna teoretična perspektiva*, niso osredotočene na povezave med posameznimi stavki in simboli, ki te predloge predstavljajo, ampak enostavno na zveze med samimi stavki.

Povezava med obema, med modelno in dokazno perspektivo, postaja zelo močna in pomembna ideja. Če je inteligenca odvisna le od logičnega razlaganja, za katerega so cilji zares ohranjeni, potem bi bilo mogoče ustvariti formalni postopek dokazovanja, ki bi imel lastnosti inteligence. Mnogokrat ni pomembno le inteligentno ukrepanje, to da se odločimo, ali je nekaj prav ali narobe, temveč moramo včasih tudi presoditi, kaj je verjetno (vendar ne nujno) resnično. To je področje *induktivne logike*. Cilj te logike je, da ustvari formalna pravila, podobna postopku teoretičnega dokazovanja

deduktivne logike, ki vodi od stavkov, ki so resnični, do stavkov, ki bi to lahko bili. Ključna predpostavka v obeh, torej v induktivni in deduktivni logiki, pa je, da pri ustvarjanju sklepov uporabljamo simbolne zapise in pri tem upoštevamo le njihovo obliko, npr. zapis p in $(q \text{ ali } r)$.

Ideja, da je inteligentni kognitivni proces bistven pri logičnem razlaganju in utemeljevanju, ima dolgo zgodovino. Hobbes je trdil, da je mišljenje primerljivo z računanjem, in predlagal, da je razmišljanje preprosto le proces formalnega računanja. (Hobbes, 1651, Bregant, 2010) To izhodiščno idejo so nadalje razvijali racionalisti kot sta Descartes in Leibniz in empiristi, kot sta npr. Locke in Hume.⁴⁵

S pojavom teorije kibernetike in računalnikov v sredini 20. stoletja se pojavijo bolj subtilni in različni pogledi na simbole in ravnanje s simboli. Po nekaterih trditvah (Haugeland, 1981) so *računalniki le naprave za vpeljevanje formalnih logičnih sistemov*. V zgodovini računalnikov je bila prelomnica možnost shranjevanja programov, sekvenc simbolov, ki natančno določajo, katero operacijo mora računalnik izvesti nad neko skupino simbolov. Razmerje med tem programom in simboli je kot zveza med formalno zapisanimi pravili npr. *modus ponens* (klasično logično sklepanje) in simboli. To perspektivo so preučevali teoretiki, kot so Dennett (2005), Fodor (1983) in Pylyshyn (1984).

Alternativo tem trditvam sta ponudila Newell in Simon s svojo hipotezo fizikalnih simbolnih sistemov. (Newell, Simon, 1972, Newell, 1980) Zanju je *računalnik fizikalni simbolni sistem*, sestavljen iz simbolov (fizikalnih vzorcev), izrazov (simbolnih struktur, povezav simbolov v fizikalnih povezavah bližnje okolice) in procesov, ki izvajajo operacije nad izrazi.

Na grobo bi lahko simbolno modeliranje procesov v človeku razdelili na dve skupini:

- *prvo*, ki zagovarja tezo, ki vodi od formalne logike k umetni inteligenci, pri čemer so računalniki naprave za manipuliranje s simboli;
- *drugo*, ki se usmerja od jezikoslovja k psihologiji, pa meni, da tudi človeški duh deluje tako, da manipulira s simboli.

⁴⁵ Descartes (2007), Leibniz (1705), in empiristi, Locke (1690) in Hume (1739–1740)

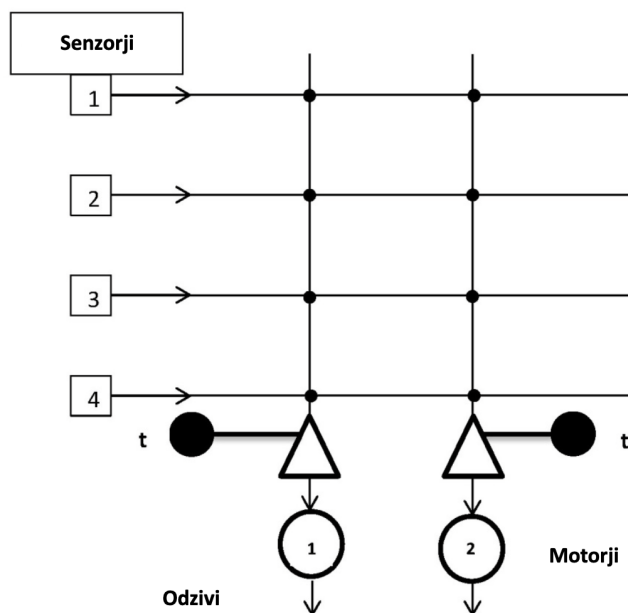
V kognitivni znanosti mnogokrat ti dve skupini združujemo in na njuni podlagi izdelujemo računalniške programe – modele ali simulacije človeške kognicije. Pri tem se zastavlja vrsta vprašanj, na osnovi katerih pridemo do zaključka, da so računalniki zares samo naprave za manipulacijo s simboli, človeški duh pa deluje popolnoma drugače.

5.1 Predhodniki konekcionizma

Spoznanja, da so možgani omrežje, je pripeljalo do razvoja mrežnih kognitivnih modelov. Zaradi kompleksnosti možganov in našega omejenega znanja seveda ni bil cilj tega razviti model aktivnosti možganov v celoti, temveč razviti model, ki predstavlja samo nekatere njihove osnovne lastnosti. Izhodišča te ideje izhajajo iz dela Warena McCullocha in Walterja Pittsa, objavljenega že leta 1943. Predlagala sta enostaven model nevronom podobnih računskih enot in demonstrirala, kako bi lahko te enote izvajale logično preračunavanje. »Formalni nevroni« so binarne enote, ki so lahko *vključeni* (1) ali *izključeni* (0). Vsaka enota prejme dražljaj in zaviralni (katalizatorski) impulz, ko je v poziciji 0. Ali bo enota preskočila v pozicijo 1, je odvisno od tega, ali bo vsota dražljajev večja od aktivacijskega praga. McCulloch in Pitts sta pokazala, kako lahko konfiguracija teh enot izvaja logične operacije *in*, *ali* in *ne*. Pokazala sta tudi, da je lahko vsak proces končnega števila teh logičnih operacij izveden z mrežo takšnih enot, imenovano *nevronska mreža*.⁴⁶ In če imamo na razpolago neskončno veliko zmogljivost spomina, potem ima lahko takšno omrežje enako zmogljivost kot univerzalni Turingov stroj. Na podlagi te ideje je v nadaljevanju nastala vrsta izboljšav izhodiščne ideje.

Njuna kasnejša dela so bila osredotočena na prepoznavanje vzorcev. Primerjala sta živali in ljudi glede na njihovo sposobnost prepoznavanja različnih verzij iste stvari –tudi ko je njihov izgled popolnoma drugačen. Eksperimente sta izvajala s pomočjo dveh mrež, kjer *prva* prepoznava *invariantne* lastnosti vzorca (lastnosti vzorca, ne oziraje se na to, kako je predstavljen) ter kjer *druga* to varianto pretvori v standardno predstavo, elementarni perceptron, kar prikazuje slika 5.1 (Bechtel, Abrahamsen, 2002: 4):

⁴⁶ Nevronska mreža (angl. neural network) je naprava za obdelavo informacij, ki deluje po vzoru človeških oz. živalskih možganov. Sestavljena je iz množice umetnih nevronov. Nevroni, osnovni gradniki nevronske mreže, so t. i. pragovne funkcije, ki imajo več različno uteženih vhodov in en izhod ter so med seboj povezani.



Slika 5.1: Elementarni *perceptron*, ki ga je raziskoval Rosenblatt (1958).

Vhodne informacije dobiva *perceptron*,⁴⁷ kot kaže slika 5.1, preko štirih senzorjev na levi in izhodne informacije so povezane z dvema pogonskima enotama spodaj. Mrežni računski enoti sta dva McCulloch-Pittsova nevrona (na sliki 5.1 velika trikotnika), katerega vsak ima inhibitorno (zaviralno) povezavo z enoto praga (na sliki 5.1 počrnjena kroga t). Vsak presek med horizontalami in vertikalami predstavlja sinapse enega senzorja na McCulloch-Pitts nevronih.

V šestdesetih in sedemdesetih letih 20. stoletja ostane v kognitivnih raziskavah dominantno in prevladujoče simbolno modeliranje in to vse do pojava *konekcionizma* v osemdesetih letih 20. stoletja.

Osredotočimo se sedaj na konekcionizem. V možganih so nevroni *živčne celice*, ki si med seboj pošiljajo živčne signale (akcijske potenciale), ta princip lahko uporabimo tudi pri drugih kompleksnih sistemih, ki delujejo podobno. V takem primeru

⁴⁷ Perceptron je najbolj razširjena nevronska mreža. Vse povezave med nevroni so pri perceptronu usmerjene naprej. Vhodne in izhodne vrednosti so lahko poljubne zvezne spremenljivke. Učenje poteka postopoma in traja, dokler ni napaka dovolj majhna.

govorimo o *mrežah formalnih nevronov*. Formalni nevroni so lahko kakršnikoli elementi nekega kompleksnega sistema, ki izvajajo podobne kolektivne procese kot (biološka) nevronska mreža. O mrežah formalnih nevronov torej govorimo takrat, ko obstaja množica nekih osnovnih elementov, tesno povezanih med seboj, ki sodelujejo ali tekmujejo ter se čim bolj usklajujejo drug z drugim. S takšno skupinsko dejavnostjo ti osnovni gradniki, ki jih imenujemo formalni nevroni, težijo k skupnemu ravnovesju. Nevronske mreže lahko tako razdelimo v dve skupini (Peruš, 2000):

- *Biološka nevronska mreža*, ki je mreža živčnih celic (pravih nevronov), ki si izmenjujejo signale prek sinaps.
- *Formalna (umetna) nevronska mreža*, ki deluje po podobnih skupinskih načelih dinamike kot biološka nevronska mreža, vendar je formalni nevron lahko neka druga informacijska procesna enota. Ta »umetni« nevron se lahko pojavlja v programski obliki (matematični nevron) ali kot silicijev nevron (modelski nevron), ki tvori silicijeve nevronske čipe.

O bioloških nevronskih mrežah smo podrobneje govorili v poglavju 4, zato se bomo v nadaljevanju posvetili le umetnim oblikam nevronskih mrež, t. i. *formalnim nevronskim mrežam*.

5.2 Genetski algoritmi za simulacijo evolucije (razvoja)

Evolucijsko računanje je smer raziskovanja umetne inteligence, ki posnema vzor iz narave, ko na podlagi *mutacije*, *križanja* in *selekcije* preživijo le primerki, ki so najuspešnejši. Osnovo evolucijskega računanja predstavljajo:

- genetski algoritmi (za optimizacijo),
- genetsko programiranje (pri razvoju programov) in
- evolucijske strategije.

Pri *genetskih algoritmih* je osnovni element (objekt) *gen*, ki podaja možno rešitev problema za en parameter. Vsi geni skupaj tvorijo *kromosom*, ki ustreza eni rešitvi problema. Geni so lahko:

- binarni vektorji,

- cela števila ali
- realna števila.

Vsaka nova ponovitev evolucijskega cikla da boljšo rešitev problema. (Aberšek, 1993) Pri genetskem programiranju je evoluciji podvržena populacija računalniških programov. Cilj je izbrati tisti računalniški program, ki najbolje rešuje dani problem. (Filipič, 1996) Evolucijski cikel sestavljajo naslednji koraki:

- generacija začetne populacije,
- ocenjevanje osebkov zadnje populacije (selekcija),
- formiranje nove generacije na osnovi genetskih operacij (križanje, mutacija).

Pričakujemo, da vsaka nova ponovitev zadnjih dveh korakov evolucijskega cikla da boljšo rešitev problema. Problem, ki ga rešujemo s pomočjo genetskih algoritmov, določa obliko populacije in cenično funkcijo, ki izbira najboljše primerke. Ta metoda se uporablja za reševanje optimizacijskih problemov, kot je na primer razporejanje proizvodnih zmogljivosti.

Študije simuliranja evolucije so povezane z nekakšno verzijo *genetskega algoritma*, ki ga je razvil John Holland (1975/1992), da bi z njim raziskoval naravo adaptivnih sistemov.⁴⁸ Holland je simuliral tri procese, kritične za biološki razvoj:

- *mehanizem dedovanja*, ki lahko proizvede potomca, podobnega njihovim staršem;
- postopek, ki vpeljuje *v proces reprodukcije različnost*;
- postopek *diferenciacije*.

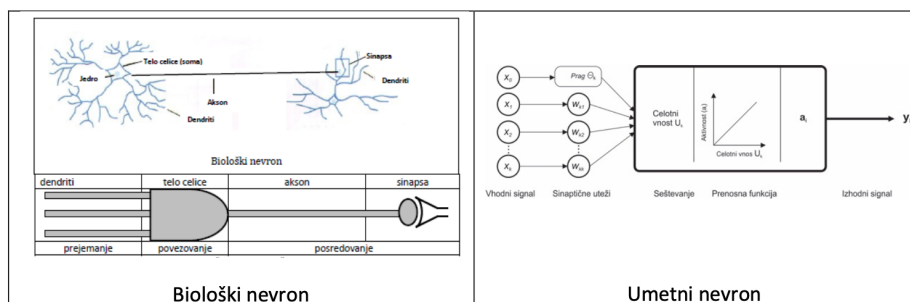
V standardni sliki biološke evolucije je mehanizem dedovanja povezan s *kromosomi* (ki se nahajajo v genih). Različnost dosežemo s spajanjem genov ali *mutacijo* in diferenciacijsko reprodukcijo z naravno selekcijo. Pri genetskem algoritmu niz simbolov igra vlogo kromosomov, operacije kot rekombinacija in mutacija teh simbolov ponazarjajo različnost (variiranje), medtem ko se nizi reproducirajo in funkcija pripravljenosti opravlja selektivno reprodukcijo. Genetski algoritem tako proizvaja nove in nove generacije. V vsaki naslednji, novi generaciji se izbere najbolj

⁴⁸ Glej tudi Goldberg (1989).

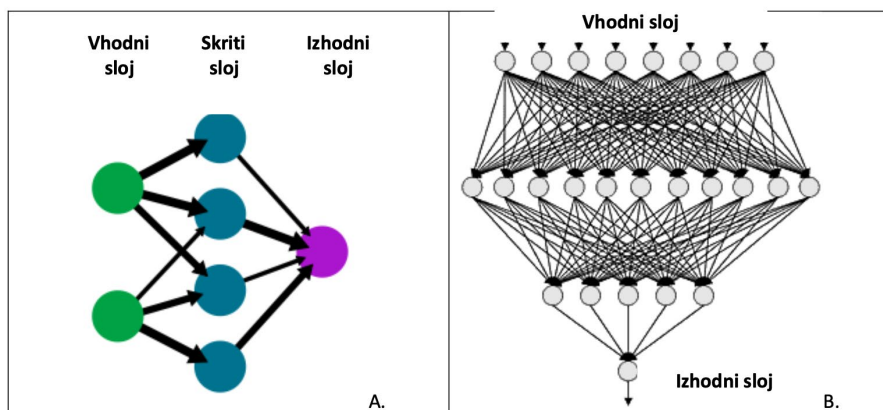
uspešen niz, ki postane starš, iz katerega izpeljemo naslednjo generacijo. Pri potomcih te generacije preverimo njihovo »zdravje« in najboljši (najuspešnejši nizi) postanejo starši naslednje generacije. V sodobnih raziskavah zadnjih let je dan poudarek predvsem na razumevanju bioloških sistemov in procesov. (Aberšek, Popov, 2004, Aberšek, Flašker, 2004)

5.3 Mrežni sistemi

S pomočjo različnih matematičnih modelov so znanstveniki s področja umetne inteligence skušali biološki nevron zapisati v obliki matematičnega modela. Na osnovi tega je nastal umetni nevron, ki je osnovni gradnik nevronske mreže. Simbolna primerjava med biološkim in umetnim nevromom je prikazana na spodnji sliki. (Aberšek, Borstner, Bregant, 2014)



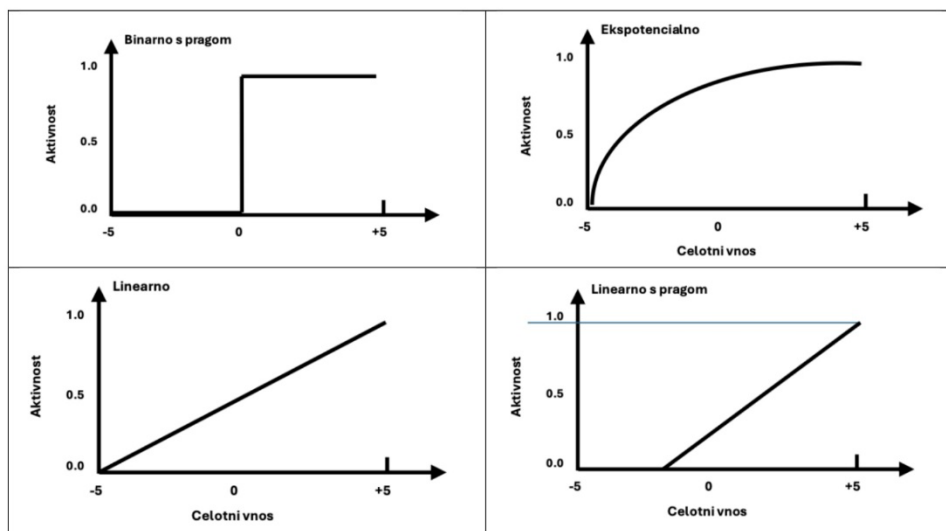
Slika 5.2: Shematski prikaz tipičnega biološkega in tipičnega umetnega nevrona.



Slika 5.3: Enostavna nevronska mreža - A in večslojna nevronska mreža - B

Medsebojna povezava večjega števila nevronov ustvari nevronske mreže. *Formalna nevronska mreža*, tudi umetna nevronska mreža, je, v nasprotju z *biološko nevronske mreže* (npr. našimi možgani), naprava za obdelavo informacij, ki deluje po vzoru človeških oz. živalskih možganov. Sestavljena je iz množice umetnih nevronov. Primer je prikazan na zgornji sliki.

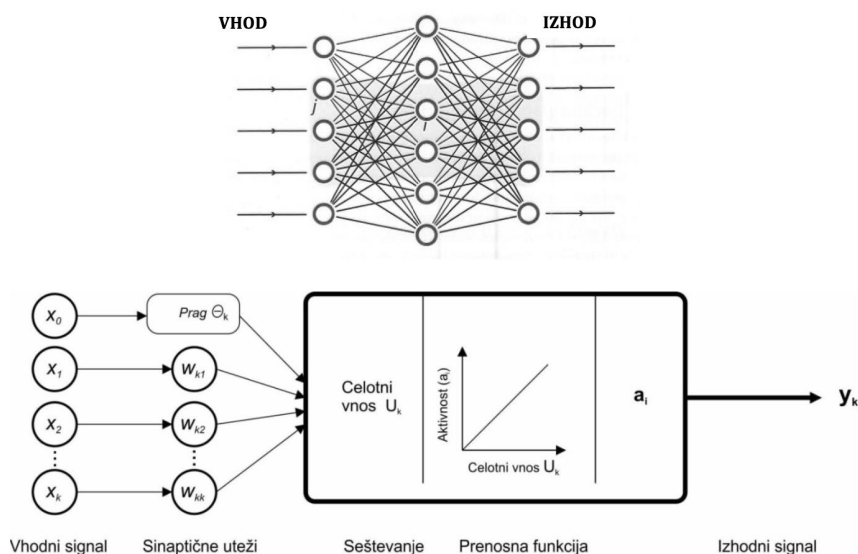
Nevroni, osnovni gradniki nevronske mreže, so t. i. pragovne funkcije, ki imajo več različno uteženih vhodov in en izhod ter so med seboj povezani. Najenostavnejše nevronske mreže imajo le en nivo oz. sloj, pri bolj zapletenih nevronske mreže pa je nivojev lahko več. Po povezavah umetni nevroni drug drugemu pošiljajo električne dražljaje oz. signale. Če je vsota vhodnih signalov dovolj velika, pride do vžiga nevrona, kar pomeni, da se na izhodu pojavi signal. Povezave med nevroni lahko imajo poljubno jakost in strukturo, kar prikazuje slika 5.4. Jakost izhodnega signala pri umetnih nevronih definiramo z *aktivacijsko funkcijo*, funkcijo, ki definira izhodni signal na osnovi vsote vseh vhodnih signalov, kar je podobno kot akcijski potencial pri naravnih bioloških nevronih, ki se ustvari po elektro-kemičnih zakonitostih. Aktivacijsko funkcijo avtorji nevronske mreže izbirajo med vrsto različnih možnosti. Nekatere izmed njih so prikazane na spodnji sliki 5.4.



Slika 5.4: Akcijski potencial umetnega nevrona.

Uteži vhodov posameznih nevronov, povezave med nevroni ter prag, pri katerem nevron na izhodu odda signal, se oblikujejo med učenjem. To pomeni, da se navedeni parametri nevronske mreže spreminjajo toliko časa, dokler ni nevronska mreža zmožna optimalno rešiti nekega problema.

Spodnja slika 5.5 predstavlja splošno strukturo konekcionistične mreže, na spodnjem delu slike je ponazorjen pogled na dogajanje znotraj posamezne enote i .



Slika 5.5: Računske operacije, ki se izvajajo v posamezni enoti konekcionističnega modela.

Proces obdelave informacij bi tako lahko razdelili na tri korake:

- (1) integriranje vseh vhodnih impulzov predhodnih nivojev in tako ustvarjanje izhodnega impulza;

$$\text{Celoten vnos} = \sum_{j=1}^n w_{ij} a_j \quad (5.1)$$

- (2) uporaba aktivacijske funkcije, s pomočjo katere spremenimo celoten vnos na aktivacijski nivo;
- (3) izhod aktivacijskega nivoja je vhod enot naslednjega nivoja. (Bermudez, 2010: 235)

Bistvo nevronske mreže je v tem, da med učenjem same ugotovijo pravilo, ki povezuje izhodne podatke z vhodnimi. To pomeni, da se lahko naučijo tudi več in bolje kot učitelj oziroma človek. Ko je nevronska mreža naučena (kar lahko traja dlje časa v primerjavi s človekom), deluje tudi v okoliščinah, s katerimi v procesu učenja ni imela opravka. To pomeni, da lahko rešuje tudi naloge, kjer ne obstaja rešitev v obliki zaporedja korakov (kot npr. pri računalniških algoritmičnih simbolnih sistemov). Pri tem obstaja seveda večja nevarnost nepredvidljivega delovanja. (Aberšek, Borstner, Bregant, 2014)

5.4 Konekcionistične mreže

Konekcionizem (nevronske mreže, vzporedno procesiranje)⁴⁹ so navdahnili spoznanja nevroznanosti o možganih. Konekcionistični modeli – nevronske mreže – vključujejo pomembne značilnosti možganske arhitekture. Da bi lahko cenili karakter in vrednosti teh mrež, jih moramo opazovati med delom, med reševanjem predvsem kompleksnih problemov. (Bechtel, Abrahamsen, 2002) Konekcionistične mreže pri tem ne delujejo natančno tako, kot delujejo nevroni in nevronske procese, temveč skušajo zajeti delovanje možganov na bolj abstraktni ravni (začetnika takih raziskovanj sta bila že omenjena McCulloch in Pitts, 1943/1990).

Konekcionistične mreže so zapleteni sistemi enostavnih enot (idealiziranih nevronov), povezanih v celoto, ki se dinamično prilagajajo okolju. Nekatere mreže imajo na tisoče enot, vendar tudi tiste z le nekaj enotami lahko izvedejo neverjetno kompleksno in bistroumno operacijo. To se dogaja predvsem zato, ker se procesiranje izvaja vzporedno in interaktivno, v nasprotju z zaporednim procesiranjem, ki smo ga vajeni pri simbolnih sistemih (digitalnih računalnikih). Vsaka enota ima določeno aktivacijsko vrednost, ki jo preko različno močnih vezi posreduje drugim enotam, in s tem pripomore k povečanju ali zmanjšanju aktivacijskih vrednosti teh drugih enot. Ves proces se odvija vzporedno in ne potrebuje nobenega osrednjega dela za nadzor. Mreža se v procesu učenja izbrane kognitivne naloge uči tako, da na osnovi učnega pravila (algoritma) spreminja moč povezav med enotami. Izbira arhitekture mreže in učnega algoritma je odvisna od tega, kakšno kognitivno nalogo naj bi z mrežo modelirali oziroma kako nevrološko verodostojen naj bi bil model.

⁴⁹ Za podrobnejše razlage glej razdelka 5.9 in 5.10

Konekcionistične nevronske mreže lahko imajo različne arhitekture. Razlikujemo jih glede na naslednje:

- *Topologijo*, torej glede na to, kako so razvrščene enote v mreži. Ločimo eno-, dvo- ali večslojne mreže (za globoko učenje).
- *Povezavo med enotami*, kjer ločimo simetrične in nesimetrične povezave, pri tem so simetrične povezave dvosmerne.
- *Funkcijo aktivacije in izhodno funkcijo*, od česar je odvisno obnašanje sistema. Aktivacijska funkcija določa, kako se bodo vhodni podatki preslikali v aktivacijsko stanje enote, medtem ko izhodna funkcija preslika aktivacijo enote v izhodno enoto. Aktivacijska funkcija je lahko deterministično binarna, deterministično zvezna ali stohastična.
- *Učno pravilo*, sistem, se prilagaja okolju s pomočjo učenja, ki poteka po izbranem učnem pravilu. S pomočjo učnega pravila sistem postopno spreminja moč povezav med enotami. Proces učenja lahko poteka *nenadzorovano* (npr. upošteva Hebbovo pravilo (Hebb, 1949) ali *nadzorovano*, npr. upošteva tekmovalno pravilo ali algoritem učenja (delta pravilo oz. angl. backpropagation)⁵⁰.

V nevronskih mrežah vsaka enota deluje le v svojem lokalnem okolju in dobi informacije samo od enot, s katerimi je povezana, skratka tako, kot je tudi delovanje urejeno v bioloških nevronskih mrežah. Informacija, ki jo enota posreduje, je električni signal, izražen s številsko vrednostjo, in se nanaša na aktivacijsko stanje enote. Skratka, edino, kar posamezna enota pozna, je lokalno stanje, ne ve pa, kaj dela celoten sistem, saj je znanje sistema porazdeljeno po enotah in se nahaja v utežeh povezav med enotami. Tudi pomnilnik, ki je pri digitalnem računalniku poseben sistem, je v konekcionističnih sistemih del procesa in je porazdeljen med posamičnimi enotami. Čeprav vsa pravila delujejo lokalno, pa v mrežah prihaja do globalnega sodelovanja, kar je glavna značilnost nevronskih mrež. (Peruš, 2000, Markič, 2010, (Aberšek, Borstner, Bregant, 2014)

Pomembne teoretske razlike med konekcionističnimi in simbolnimi modeli so v drugačnem pojmovanju predstavitve znanja. Predstavitve znanja v konekcionizmu niso več statična, enostavna in z jezikom opisljiva notranja stanja. Ne tvorijo več

⁵⁰ Za podrobnejšo razlago glej poglavje 5.16.2

nujno simbolnega sistema, so podsimbolne in porazdeljene od konteksta odvisne skupine enot. Prav tako znanje ni več predstavljeno v formalnem zapisu jezika ali logike, temveč je shranjeno v utežeh povezav nevronske mreže kot posledica učenja. Konekcionisti se od klasične kognitivne znanosti razlikujejo tudi v zavračanju ostre delitve na funkcionalno in strukturalno raven. Pogosto je le od interesa raziskovalca odvisno, kako bo ocenil, na kateri ravni raziskuje, in je zato nemogoče oceniti, da določena raven ne pripomore k razlagi izbranega fenomena. Ob tem je v primeru, ko sta dve teoriji podobni na kognitivni ravni – ko imamo dve teoriji, ki obe enako dobro razlagata in napovedujeta kognitivne pojave – prav nevrološka verodostojnost tista, ki odloča o izbiri ene od obeh teorij.

Konec osemdesetih let 20. stoletja sta Fodor in Pylyshyn z argumentom iz sistematičnosti misli spodbudila diskusijo o tem, ali konekcionizem predstavlja alternativni pristop k modeliranju na kognitivni ravni ali je zgolj implementacija simbolnih modelov. (Fodor, 1983) Slednje je bilo s strani »klasikov« slišati kot nekaj, kar je za razlago vedenja manj zanimivo. Kasnejši razvoj je šel v veliki meri prav v načrtovanje računskih modelov nevroznanosti, ki veliko bolj kot konekcionizem upoštevajo značilnosti možganov in je z njimi mogoče modelirati posamezne predele možganov in interakcije med njimi. Začela so se zapostavljati vprašanja o možnostih redukcije na nevroznanost. (Churchland, 2007)

5.4.1 Perceptron

Do leta 1960 je napredek pri razvoju strojne inteligence na obeh področjih modeliranja tako pri mrežnih kot tudi pri simbolnih modelih enakomerno napredoval. Ta enakomeren razvoj se je po tem letu spremenil, predvsem po zaslugi izjemne finančne podpore DARPA – Defense Departments Advance Research Project Agency – simbolnim modelom strojne inteligence na temelju računalniških programov. Rezultat tega je bilo leta 1969 objavljeno delo Marvinja Minskyja in Symorja Paperta z naslovom *Perceptron* (1969). Le nekateri znanstveniki so še vztrajali pri mrežnih sistemih, npr. Anderson (1983) ali Grossberg (1976) in še to ob minimalni finančni podpori. Minsky in Papert sta analizirala tako potenciale kot omejitve mrežnih modelov z uporabo matematičnih orodij in ugotovila, kateri perceptroni bi lahko bili oz. ne bi mogli biti uporabljeni (npr. ločevanje med parnimi

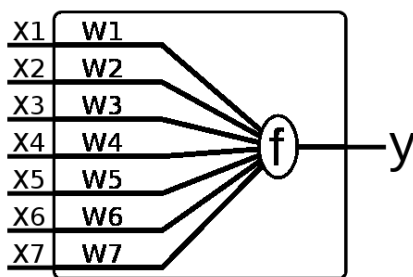
in neparnimi števili, ki jih ne moremo izvesti z mrežami, ali pa *izključujoči ali*⁵¹), o čemer bomo podrobneje razpravljali v poglavjih 6 in 7.

Problemi bi bili morda lahko rešljivi z dvoslojnimi mrežami, a vprašanje, kako bi reševali kompleksnejše probleme, vseeno ostaja. Z nekaterimi omejitvami bi lahko mreže usposobili, da bi izvedle želene izhode iz podanih vhodnih podatkov, vendar bi morali razviti postopek asociacije vhodnih podatkov z želenimi rezultati na izhodu. Ta asociativizem je nekaj, kar so snovalci modernega kognitivizma, kot so Newell, Anderson, Hinton in drugi, zavračali.

Perceptron pravzaprav predstavlja mračno obdobje konekcionizma. Perceptron je oblika nevronske mreže, ki jih je leta 1958 razvil Frank Rosenblatt (Rosenblatt, 1962). Poenostavljeno povedano je perceptron binarni razvrščevalnik, ki matrično preslika vhodno vrednost x (realni vektor) v izhodno vrednost $y = f(x)$ (posamezno binarno vrednost):

$$y = f(x) = \begin{cases} 1 & \text{če je } wx + b > 0 \\ 0 & \text{drugje} \end{cases} \quad (5.2)$$

kjer je w vektor realnih vrednosti uteži, $w \cdot x$ je točkovni produkt (ki ga računalniki utežno seštevajo) in b je konstanta, ki je neodvisno od vhodnih parametrov, kar je prikazano na sliki 5.6.



Slika 5.6: Simbolni prikaz perceptrona, kjer so ustrezne uteži uporabljene na vhodnih podatkih in povzročajo utežno vsoto, ki vpliva na izhodno funkcijo

⁵¹ Logična operacija izključujoče ločevanje, ki ga mnogokrat imenujemo tudi »izključujoči ali« (zapisano s simboli XOR, EOR, EXOR) je tip logičnega ločevanja dveh operandov (količin, na katerih se delajo računske operacije), s pomočjo katerih pridemo do pravilnega rezultata, če ima natančno eden od operandov pravilno vrednost. Z drugimi besedami bi to lahko izrazili kot: »eden ali drugi, vendar ne oba«, ali simbolno: $p \text{ XOR } q$.

Kot večina raziskovalcev UI je bil tudi Rosenblatt optimističen, ko je napovedal, da se bo perceptron sposoben učiti, odločati in prevajati različne jezike. (Rosenblatt, 1962) Intenzivne raziskave o perceptronski paradigmi so se pojavile v šestdesetih letih 20. stoletja. Vendar je prišlo do nenadne zaustavitve z objavo knjige *Perceptron*, ki sta jo objavila Minsky in Papert leta 1969. V njej ugotavljata, da obstajajo resne omejitve o tem, kaj perceptron zmore in ugotovila, da so bile Rosenblattove napovedi precejšnje pretiravanje. To je povzročilo, kot smo že omenili, da so se konekcionistične raziskave v naslednjem desetletju skoraj popolnoma prekinile.

5.5 Konekcionizem in simbolni sistemi

Minsky in Papert sta analizirala tako potenciale kot omejitve mrežnih modelov z uporabo matematičnih orodij in ugotovila, kateri perceptroni bi lahko bili oz. ne bi mogli biti uporabljeni (npr. ločevanje med parnimi in neparnimi števili, ki jih ne moremo izvesti z mrežami, ali $p \text{ XOR } q$), o čemer bomo podrobneje razpravljali v poglavju 6. Problem bi morda lahko rešili z dvoslojnimi mrežami, vprašanje pa vseeno ostaja – kako bi reševali takšne bolj kompleksne probleme?

Z nekaterimi omejitvami bi lahko mreže usposobili, da izvedejo želene izhode iz podanih vhodnih podatkov, vendar bi morali razviti postopek asociacije vhodnih podatkov z želenimi rezultati na izhodu. Ta asociativizem je nekaj, kar so snovalci modernega kognitivizma zavračali.

5.6 Dinamika konekcionističnih modelov

Poskusi, da bi izdelali inteligentne stroje, so povzročili revolucijo v mišljenju o mišljenju. Kot pravijo kognitivni znanstveniki, predstavlja računalnik naše novo zrcalo. Računalniška metafora je odprla nove revolucionarne poti v raziskovanju človekovih spoznavnih procesov. Možnost modeliranja človekovega spoznavanja je namreč omogočil šele nastop računalnika. Pred tem so obstajali zgolj opisi miselnih procesov, ki so temeljili na notranjem uvidu, na introspekciji. (Horgan, Tienson, 1996)

Področje umetne inteligence je vpeljalo nov pojmovni aparat ter omogočilo proučevanje kognitivnih procesov s povsem novih vidikov. Pri opisu in razlagi miselnih procesov smo začeli uporabljati nove pojme, ki so prihajali iz računalništva,

kot na primer program, procedura, algoritem, hierarhičen nadzor, shema. Ob tem, da je imel računalnik vlogo nosilca metafore, je po drugi strani predstavljal orodje, s pomočjo katerega so potekale simulacije določenih miselnih procesov, kot na primer reševanje problemov, priklic iz spomina, prepoznavanje vzorcev. Osnovna hipoteza kognitivne znanosti, da sta tako um kot računalnik informacijska sistema, ki obdelujeta informacije, je dobila s konekcionistično teorijo novo vsebino. Novi teoretski izrazi tudi natančneje opisujejo in razlagajo spoznavne procese, ki potekajo v možganih.

Vlada prepričanje, da je organizacijo in interakcije kognitivnih sil mogoče najboljše razložiti s pomočjo matematične teorije dinamičnih sistemov – TDS, ki je dolgo igrala osrednjo vlogo v fiziki. (Horgan, 1997) Matematični koncepti, tehnike in rezultati, ki so nastali v okviru te teorije, so začeli igrati pomembno vlogo tudi v konekcionistični kognitivni znanosti in v nevroznanosti. Matematični opisi konekcionističnih modelov oziroma nevronske mreže tako prevzamejo metodologijo opisa dinamičnih sistemov, kakršni so običajni, denimo, v fiziki. Sistem je pri tem opisan kot večdimenzionalna pokrajina, katere vsaka dimenzija predstavlja aktivacijsko vrednost določenega posamičnega nevronskega vozlišča. Določeno stanje, to je vzorec, ki ustreza reprezentaciji, se pojavi kot umiritev na neki točki v celotni pokrajini sistema.

Ljudje vse od začetka sveta izkazujemo brezmejno željo po razumevanju sveta okoli sebe. Védenje, kaj je svet, se je s človekovim razvojem spreminjalo. V začetku je bil svet le bližnja okolica, kmalu pa se je pričel človek obračati k zvezdam in sanjati o neskončnosti vesolja, ki je postajalo njegov svet. Vse skozi je želel spoznati in tudi razumeti tako vse posamične sile narave kakor tudi njihovo vzajemno delovanje. Vsa ta opažanja je hotel formalizirati, zapisati v obliki enačb, ki bi predstavljale medsebojno delovanje naravnih sil. Z njimi ne le da bi lahko preračunaval le trenutno stanje sistema (kompleksne strukture, kot so vesolje ali človek) kot celote ali njegovih sestavnih delov, ampak bi lahko napovedoval tudi prihodne dogodke na osnovi strogih logičnih in matematičnih dokazov. Pri tem se mu je, hote ali nehote, zastavljala vrsta večinoma dvoumnih vprašanj:

- Kako naj se lotimo reševanja problema razumevanja sveta?
- Kako smo lahko prepričani v pravilnost naših spoznanj?

- Ali obstajajo stvari, katerih človeški um na sedanji stopnji razvoja ni sposoben dojeti?

René Descartes je v Platonski tradiciji podal argument, da je prevladujoča zahodna znanstvena metoda dokaza aksiomska oz. geometrična metoda. Ta pristop se začne s postavitvijo hipotez, domnev, brez dokazov, temelječ na nepreverjenih aksiomih ali postulatih, ki pa so splošno sprejeti, ker so očitno resnični. Dokaz je izpeljan filozofsko, od hipoteze do sklepa z uporabo vrste pravil logičnega sklepanja, ki nam dovoljujejo, da korak za korakom predelamo vsako izjavo, ki ji sledi naslednja glede na strogo logiko, tako dolgo, dokler dokaz ni popoln. Vsaka teorija ali predlog o tem, da iz določene hipoteze sledi določen sklep, mora biti dokazana na ta način, kar velja tudi za druge podobne hipoteze, tako da vse znanje izhaja iz premis, predpostavk, aksiomov, postulatov in hipotez. To nam služi kot učinkovit način ohranjanja tolikšne količine gotovosti, kolikor se je v kopičenju znanja le da, prav tako se lahko prepričamo, da so naše teorije med seboj skladne in si ne nasprotujejo.

Nadaljnje izkušnje bi lahko, in bi morale, voditi do poznejšega prečiščenja teorije, s tem, da se držimo konstantne zahteve po tem, da se mora teorija skladati z opazovanji realnega sveta, da mora biti bolj ali manj veren odraz realnega problema. Te razlagalne teorije vodijo do zbirke enačb, ki združene opisujejo obnašanje naravnih sistemov v našem vesolju skozi čas, prostor, v prostoru stanja ali kako drugače. Na ta način razvijamo matematične modele, s katerimi predstavljamo obnašanje resničnega sveta, ki ga opazujemo v naravnih sistemih okolja.

Klasični ali newtonski pristop, ki je verjetno najmočnejše vplival na mišljenje zahodnega sveta, pravi, da smo zmožni razumeti fizične zakone narave, ki vplivajo na naravne učinke, do tolikšne mere, da lahko zapišemo končno število enostavnih linearnih enačb, ki vsebujejo končno število spremenljivk ali določilnih dejavnikov, ki vplivajo na obnašanje zapletenih naravnih sistemov, ki so skoraj brez odstopanja dinamični sistemi. Ta pristop pravi, da lahko definiramo začetne pogoje, ki so vključeni v sistem simultanih enačb, ki oblikujejo obnašanje naravnega sistema in se odražajo v natančnih, absolutnih napovedih kateregakoli prihodnjega stanja sistema v času, prostoru ali v čemerkoli drugem.

Obstaja zelo ekstremno mišljenje, da se določeni pojavi v naravi, še posebej tisti v krajših časovnih razmikih in na krajših razdaljah, pojavljajo popolnoma naključno. Naključje nasproti klasičnemu determinizmu (ko so vsi dogodki določeni z vsoto predhodnih dogodkov in naravnih zakonov) je postala priljubljena polarizirana dvojnost, o kateri se veliko razglablja. Trenutno najbolj priljubljena teorija, ki obravnava to naključnost, je teorija kaosa, ki predstavlja pomemben korak pri razumevanju tega, da so stvari veliko bolj zahtevne od klasične Newtonove deterministične interpretacije. Zapleteni naravni dinamični sistemi, ki jih preučujemo, so oblikovani matematično s kaotičnimi sistemi, sestavljenimi iz več simultanih nelinearnih integralno-diferencialnih enačb z več spremenljivkami, ki kažejo na določene lastnosti kaosa, vključno z občutljivo odvisnostjo od začetnih pogojev. Ti abstraktni matematični modeli kaotičnih sistemov, ki jih vidimo v našem naravnem svetu, so lahko sestavljeni iz nešteto enačb, ki lahko imajo nešteto spremenljivk ali določilnih dejavnikov, vsak od teh dejavnikov pa lahko ima poljubno število pogojev, vključno z razlago vseh spremenljivk, vseh izpeljank, vseh integralov in vseh funkcij, skupaj s primernimi koeficienti za vsak pogoj, ki so določeni z njihovo močjo ali težo.

Torej, če bi želeli nadaljevati s klasičnim determinističnim pristopom, moramo meriti, začeti pa moramo z neskončnim številom začetnih pogojev, kakor tudi z neskončnim številom robnih pogojev, ki morajo biti ves čas prisotni in ki tudi vplivajo na rezultat ali prihodnje stanje sistema.

Verjetno je za človeka nemogoče, da bi popolnoma dojel in razumel vsa razmerja sil, ki vodijo obnašanje sistemov naravnih pojavov v vesolju, da bi jih lahko na ta način zajel v enačbah in jim poiskal ustrezne rešitve. Obstajajo mnenja, da popolnoma naključnih pojavov ni in da so zagotovo določeni s kompleksnimi naravnimi procesi. Naključje v naravnem vesolju v bistvu ne obstaja, obstaja samo kot matematični model, ekstremni ali idealni primer na enem koncu spektra razumevanja, ki ga je realnosti vsilil človek. Ob teh razcepih pride do kaotičnega nenapovedljivega vedenja sistema. Najpomembnejši od teh modelov je splošna logistična enačba, katere grafična predstavitev vključuje množico Mandelbrot⁵² in neskončno število Julijevih

⁵² Množica Mandelbrot je v matematiki množica točk v kompleksni ravnini, katere meja tvori fraktal, ki ga definiramo kot "hrapavost ali fragmentirano geometrijsko obliko, ki se lahko deli na dva dela, vsak od njiju je pomanjšana kopija celote, poimenovana tudi samopodobnost. Imenuje se po francosko-ameriškem matematiku Benoïtu B. Mandelbrotu.

množic,⁵³ ki igrajo pomembno vlogo pri razumevanju splošnega obnašanja naravnih kompleksnih dinamičnih sistemov, kadarkoli se dva ali več dejavnikov sreča na nelinearen način, kar se odraža v atraktorjih, odbojnikih, stalnih točkah, periodičnih in nestalnih ciklih itd., dolgoročno pa v obliki ponavljalne »analize povratne informacije«. Te vrednosti imajo meje v dimenzijah fraktala, kar kaže na neskončno stopnjo zapletenosti, ki je na delu v naravnem svetu, kakor je to prikazano v teh grafičnih predstavitvah.

5.7 Prepoznavanje vzorca in znanje

Mreže so naprave za preslikavo enega razreda vzorcev v drugi razred. To naredijo tako, da kodirajo statistične zakonitosti v utežne povezave, ki jih lahko modificiramo na osnovi izkušenj. (Bechtel, Abrahamsen, 2002: 89–90) Tako so konekcioniistične mreže izjemno primerne za modeliranje človeške kognicije. Ustvarjanje vzorcev s pomočjo mrež pomeni:

- preslikava vzorca je temelj vrste človeških sposobnosti;
- konekcioniistične mreže izvajajo preslikave na posebno ugoden (primeren) način.

Preslikave so pravzaprav izjemno širok koncept, zato moramo ločevati med različnimi tipi preslikav:

- *prepoznavanje vzorca* je preslikava posebnih vzorcev v bolj splošne (to je prepoznavanje posameznikov kot primerkov razreda);
- *kompletiranje vzorca* je preslikava, ki nekompletne vzorce preslika v kompletne vzorce istih vzorcev (npr. besedo *pr_dmet*, preslika v kompletno besedo *predmet*, ali pri vizualni zaznavi spoznamo dele nekega objekta, ki ga nato dopolnimo v končno sliko);
- *transformacija vzorcev* je preslikava enega vzorca v drug soroden vzorec (npr. besedo *come* transformiramo v preteklik *came*);

⁵³ Julijeva množica je množica, sestavljena iz vrednosti funkcij, pri katerih tudi najmanjša motnja lahko povzroči drastične spremembe vrednosti te funkcije. Njeno vedenje označujemo kot kaotično. Ime je dobila po francoskem matematiku Gastonu Juliji.

- *povezovanje (združevanje) vzorcev* je poljubna preslikava nekega vzorca v druge, nesorodne vzorce (kot je npr. naloga združevanja parov, ki je ena od pomembnih metod pri tradicionalni psihologiji učenja);
- *avtoasociacija* je preslikava vzorca samega vase (kjer npr. iz velikega števila vhodnih podatkov preslikamo (opišemo) vzorec v poenostavljeni obliki).

Vseh pet tipov preslikav dokazuje uporabnost konekcionističnih modelov človeških kognitivnih predstav. Pri ljudeh igra prepoznavanje vzorca najpomembnejšo vlogo pri občutkih in zaznavah. Ko opazujemo neko sceno, se vzorec svetlobe kodira na očesni mrežnici in se nato dekodira na različne načine v živčnem sistemu, kar je osnova za prepoznavanje.

5.8 Ponovno rojstvo – novi konekcionizem

V zgodnjih osemdesetih letih 20. stoletja je kognitivna znanost ponovno odkrila Rosenblattove ideje, število konekcionističnih raziskav je pomembno naraslo in v devetdesetih letih 20. stoletja so postale vse pomembnejše. Vzrok je bila vrsta naključij:

- Prvič, pojav vplivnih modelov z novo arhitekturo in novimi tehnikami uporabljanja večslojnih mrežnih naprednih metod in opisovanja nelinearnih sistemov, kar je bilo neposredno uporabno za kognitivno modeliranje procesov.
- Drugič, nekaj pomembnih ljudi, kot je npr. znani fizik John Hopfield, je zastavilo svojo besedo za kognitivno znanost.
- Tretjič, na »prava« mesta so prišli pravi ljudje, kot sta npr. Anderson in Hinton.
- Četrtrič, kognitivni znanstveniki so začeli ponovno ugotavljati prednosti konekcionističnih modelov pri nevronskih in podobnih arhitekturah.
- Petič, povedano je vzbudilo splošno pozornost.
- Šestič, veliko ljudi je pričelo ugotavljati omejitve simbolnih modelov, zato so se pri njih in posledično tudi pri drugih pričeli pojavljati dvomi o tem, ali so lahko sistemi pravil zares sposobni opisovati človeško obnašanje. Pričelo se je poudarjati krhkost, nefleksibilnost, izpostavljanje probleme pri izkustvenem učenju, neprimernost za posploševanje, probleme pri določanju domen in

sistemov na osnovi pravil. Za človeški duh pa vemo, da je brez omenjenih omejitev in problemov.

Resničen boj med simbolnimi in konekcionističnimi strujami se je tako začel sredi osemdesetih let 20. stoletja. Rezultat je bil, da so se meje med obema strujama pričele »mehčati«. Npr. John R. Anderson (1996) je uporabil v svojem sistemu *ACT-R* lokalistično mrežno arhitekturo za dolgoročni spomin in sistem produkcijske arhitekture za iskanje rešitev. V *SOAR* arhitekturi (Laird, Newell in Rosenbloom, 1987) so uporabili produkcijski sistem za obe nalogi. Newell je v svoji mojstrovini *Unified Theories of Cognition* (1990) izpostavil, da je osnovni smisel konekcionističnih modelov prav v njihovi enostavnosti. Na konekcionistični strani so nekateri razvili različne *hibridne* modele, tako da so implementirali določene izračune na osnovi pravil v kolektivistične arhitekture, s čimer so združili prednosti obeh.

5.9 Sodobne nevronske mreže

Nevronske mreže so danes verjetno, splošno priznано, najboljši model simuliranja možganskih procesov na mikroskopskem nivoju. V posplošeni, funkcionalistični različici so tudi najuspešnejši model duševnih procesov, vsaj če zanemarimo zavest in takšnosti (lat. *qualia*). (Peruš, 2000, Aberšek, 2012) Med seboj se nevronske mreže razlikujejo po zgradbi, po vrsti učenja (nadzorovano in nenadzorovano) in po vrsti signalov (analogni in digitalni). Tako v najbolj razširjenem pojmovanju izraz »nevronske mreže« ponazarja model nekega kompleksnega sistema, kjer je mnogo približno enakih elementov povezanih med seboj z namenom medsebojnega delovanja (npr. perceptron, samoorganizirane preslikave (angl. Self-Organizing Maps), pa tudi pri Kohonenovi in Hopfieldovi nevronske mreži (asociativni pomnilnik)). V tem primeru govorimo o mrežah formalnih nevronov, ki so lahko različni elementi nekega kompleksnega sistema, ki izvajajo podobne kolektivne procese kot biološke nevronske mreže, ki smo jih podrobneje opisali že v poglavju 3.

V nadaljevanju se bomo osredotočili predvsem na umetne, formalne nevronske mreže. Ločili bomo več vrst mrež formalnih nevronov. Najbolj znane nevronske mreže so:

1. *Modularna nevronska mreža*; formalna nevronska mreža je skupek (modul) pravih nevronov. V to skupino spadajo:
 - *Kvantna mreža*; formalni nevron je nek kvantni element (»točka«) in ni nujno, da je kvantni delec. Kadar hočemo posebej poudariti, da gre za mrežo kvantnih »točk« (matematičnih oziroma modelskih) globljo od kvantnih delcev, govorim o *subkvantni mreži*.
 - *Virtualna mreža*; *mreža posplošenih nevronov*, kjer so formalni nevroni višje organizacijske oblike (gestalti) bioloških nevronov ali drugih materialno udejanjenih formalnih nevronov.
 - *Konceptualna* oziroma *semantična mreža*; označba za mreže konceptov, idej, duševnih vzorcev. Za monista je to podmnožica virtualne mreže, za dualista pa bi bila nekaj neodvisnega.
2. *Mnogoslojna fraktalna mreža*; ta izraz ne označuje hierarhično urejene mreže formalnih nevronov iste vrste, temveč sestavljeno mrežo formalnih nevronov različnih vrst oziroma velikosti. To je mreža formalnih nevronov, kjer na različnih ravneh oziroma velikostnih skalah nastopajo formalni nevroni različnih vrst. V tej skupini omenimo še *nevrokvantno mrežo*, ki je mnogoslojna fraktalna mreža, kjer na makroskopski ravni nastopajo pravi nevroni (živčne celice), na mikroskopski ravni (denimo znotraj njihovih aksonov) pa kvantni formalni nevroni (delci, spini, kvantne »točke«). (Peruš, 2000)
3. *Silicijevе nevronske mreže*; so mreže modelskih nevronov, ki so analogna vezja, čipi, ki prenašajo informacijo s pomočjo kontinuiranih sprememb v napetosti, prav tako kot živčne celice v svojem podpražnem stanju. Izračune je tako možno opraviti v manj korakih, s pomočjo izkoriščanja osnovne fizike silicijevih naprav. Analogno računanje enostavno podaja osnovne elemente aritmetike: seštevanje, odštevanje, eksponente in integracije, operacije, ki so sicer precej zahtevne za digitalne stroje. Ti nevroni – naj bodo biološki ali silicijevi – računajo in sprejemajo odločitve, pošiljajo impulze vzdolž aksonov ter komunicirajo s ciljnimi nevroni. (Rothschild, 2010, Dethier et al., 2011)

5.10 Umetne nevronske mreže

Za preučevanje učenja in spomina pogosto uporabljajo umetne nevronske mreže (UNM). Najpogostejše so udejanjene kot programska oprema na konvencionalnih digitalnih računalnikih. Sestavlja jih večje število enostavnih procesnih enot, ki so gosto povezane v omrežje. Najenostavnejša oblika UNM so *enostavni asociatorji*, ki jih sestavljajo sloji medsebojno enosmerno povezanih vhodnih in izhodnih enot. Asociativni spomin je zabeležen s pomočjo spremembe moči povezav med sloji, tako da mreža ob vnosu vhodnega vzorca na izhodnih enotah poustvari z njim povezan shranjeni vzorec. Bolj kompleksne UNM so *povratne nevronske mreže*. Te so sestavljene iz enega samega sloja, v katerem je vsaka enota povezana z ostalimi ter vse enote delujejo kot vhodne in izhodne. Sliši se nekoliko nenavadno, a takšna zasnova omogoča mreži, da hrani vzorce in ne le parov elementov. Takšno *avtoasociacijsko mrežo* je mogoče dekodirati z rekurzivnim iskanjem shranjenega vzorca. Preizkusi so pokazali, da je lahko v mreži s 1000 elementi shranimo približno 150 vzorcev brez prevelikih napak v priklicanih vzorcih. (Morris, Filenz, 2003)

UNM so možganom podobne v načinu, kako hranijo in obdelujejo informacije. »Znanje«, ki ga obdelujejo, je shranjeno v mreži sami. Spomin ni shranjen posebej, kot to velja za digitalne računalnike, kjer je aritmetični procesor ločen od spominskih naslovov. Spomini se shranjujejo na način, ki omogoča njihovo *vsebinsko naslavljanje*. V UNM so informacije shranjene v utežeh posamičnih povezav, podobno kot sinapse med učenjem spreminjajo svojo moč. Ob tem UNM niso programirane za izvajanje določenih postopkov. Posamičen nevron v mreži je neumen ter se zgolj odziva v skladu z vsoto uteženega vnosa. Kljub temu pa je nevrone mogoče naučiti opravljanja pametnih zadev. Nevronska mreža je mogoče učiti s pomočjo učnih pravil, ki spreminjajo moči povezav med posameznimi nevroni. Pogosto pravilo temelji na primerjanju dejanskega izhodnega vzorca mreže, ki je posledica podanega vhodnega vzorca, z želenim vzorcem. »Napake«, ki jih razkrije primerjava, so nato podlaga prilagajanja uteži povezav, tako da je dobljena izhodna vrednost vedno bližja želeni. Mreža na ta način postopoma zmanjšuje napako do minimuma. Postopek je učinkovit, a počasen.

Napake v mreži so pomembne – brez njih učenje namreč ni možno. Ta lastnost učenja je pogosto spregledana. Pretirano trenirane mreže, ki ne delajo napak, bi se odzivale le na eno vrsto vhoda. Takšne mreže metaforično poimenujemo »babičine

mreže», s čimer namigujemo na analogijo s teoretično predpostavljenimi, a verjetno bolj mitskimi »babičnimi celicami«, ki naj bi se odzivale le takrat, ko vidimo babico, in naj bi pri tem nikoli ne zagrešile napake. V resničnem svetu to ni posebej uporabno, saj bi za vsako stvar, ki bi se je naučili, potrebovali svojo mrežo. V nasprotju s tem je prednost UNM prav v njihovi sposobnosti *posploševanja* odzivanja na vhodne vzorce, tudi na tiste, ki jim v postopku učenja nikoli niso bile izpostavljene. Sposobne so videti odnose, zajeti asociacije in odkriti zakonitosti v vzorcih. Ob tem pa so odporne na napake – prav tako kot pravi možgani. Shranjeni vzorec so sposobne obnoviti tudi kadar je vhodni vzorec nepopoln ali mu je prištet šum. To so zelo pomembne lastnosti bioloških možganov, ki jih izkazujejo tudi UNM.

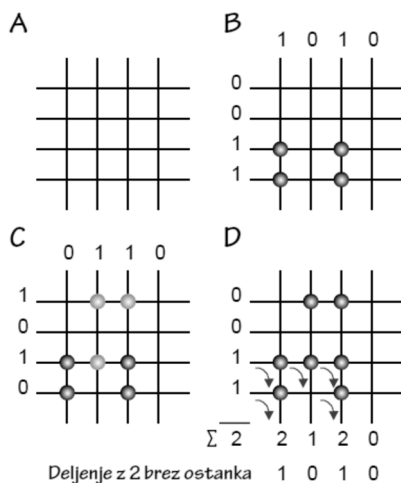
5.11 Paradoks moderne računske tehnologije

Paradoks današnjih UNM je, da so simulirane matematično na digitalnih računalnikih. Slednje omejuje njihovo uporabo v vsakodnevnih situacijah, saj so simulacije časovno zahtevne in UNM tako ne morejo delovati v realnem času. UNM se zdijo pravšnje za vožnjo avtomobila ali upravljanje letala, saj delujejo robustno tudi ob prisotnosti šuma in v primerih, ko katera od enot v mreži odpove. Kljub temu pogosto uporabljeni ekspertni sistemi, kot na primer avtomatični piloti, temeljijo na digitalnih računalnikih, programiranih s konvencionalno deterministično programsko opremo, ki za zagotavljanje varnosti potrebuje sistem z več varnostnimi podsistemi. Če se pri upravljanju letala pojavijo nepredvidljive težave, se ekspertni sistemi s tovrstnimi situacijami ne zmorejo spoprijeti in pilot mora prevzeti nadzor. Današnji algoritmi za učenje UNM so v tovrstnih nujnih in izrednih primerih prepočasni. Ko bomo vedeli več o delovanju možganov, bomo sposobni zgraditi tudi kompleksnejše nevronske mreže, ki bodo dosegale sposobnosti pravih možganov. Ena od rešitev bodo verjetno tudi silicijeve nevronske mreže, pri katerih je sedaj problem še njihova samoučljivost. Če bi bili silicijevi nevroni sposobni učenja, česar trenutno ne zmorejo, bi bili razrešeni številni problemi, ki smo jih našli.

5.12 Primer: matematična uganka

VSEBINSKO NASLOVLJIV PORAZDELJEN SPOMIN

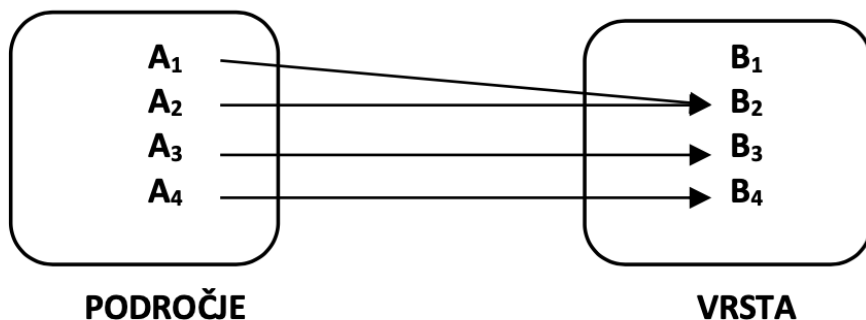
Zamislimo si niz žic, ki tečejo vodoravno in se srečajo s štirimi, ki tečejo navpično, na mestih, kjer se srečajo, pa so stikala (slika 5.7 A). Ta matrika naj predstavlja spomin. Informacije so v njej zabeležene v obliki binarnih števil, kot npr. 0011 (vertikalno na sliki 5.7 B) in 1010 (horizontalno na sliki 5.7 B). Stikala uredimo tako, da se prižgejo vsakič, ko 1 sreča 1 (kar je prikazano na sliki 5.7 B). Ta stikala hranijo paritev teh dveh števil. Matrika lahko hrani tudi druga števila na vrhu omenjenih dveh (prvi primer ostane trajno v spominu), kot na primer 1010 in 0110, seveda odvisno od nastavitve (uporabljenih pravil – Boolovih funkcij). Končna matrika, npr. naj ima prižganih sedem stikal, kot prikazuje slika 5.7 C. Če sedaj ponovno prikažemo prvo število – 0011 – končnemu stanju matrike in uredimo, da je tok prisoten v navpičnih žicah vsakič, ko je prižgano stikalo (slika 5.7 D), bo tok, ki prihaja na dnu navpičnih črt, sorazmeren številu 2120. To ni število, s katerim je bilo 0011 prvotno parjeno. Vendar, če delimo 2120 s skupnim številom enic v številu, ki smo ga uporabili za priklic ($0+0+1+1$ je enako 2), z uporabo deljenja celih števil (kjer zanemarimo ostanek), končamo z 1010. Matrika si je torej »zapomnila«, da sodi 0011 skupaj z 1010, čeprav je na njej shranjeno dodatno sporočilo. Da resnično deluje, lahko preverimo tudi z drugim parom števil.



Slika 5.7: Simbolična predstavitev spomina

To je vrsta spomina, ki ga pripisujemo možganom. Informacij ne hranijo na določenih lokacijah, kot to velja za osebni računalnik, ampak so te razpršene po mreži, shranjene kot spremembe v sinaptičnih povezavah. Priklicati jih je možno z naslavljanjem njihove vsebine. Težava je, da se tovrstni spomin zelo hitro zasiči, še posebej, če so na razpolago le štiri žice. S 1000 pari žic, lahko matrika hrani veliko prekrivajočih se parov sporočil brez večjih motenj med njimi. (Morris, Filenz, 2003)

Pokomentirajmo to uganko nekoliko bolj »znanstveno«. Na sliki 5.8 je prikazan primer funkcij preslikave. Puščice prikazujejo, katera točka s področja bo preslikana v drugo točko vrste. Imamo različne možnosti, tudi takšne, da se dve točki področja preslikata v eno točko vrste. Z vpeljavo binarne Boolove algebre dobimo le niz {prav, narobe}, kot je prikazano v preglednici 5.1. Vsaka binarna Boolova funkcija ima lahko vrednost le PRAV ali NAROBÉ.



Slika 5.8: Prikaz funkcije preslikave

Tabela 5.1: Boolova funkcija

Področje	Vrsta
NAROBÉ, NAROBÉ	
NAROBÉ, PRAV	NAROBÉ
PRAV, NAROBÉ	PRAV
PRAV, PRAV	

Boolove funkcije so lahko različne, lahko so oblike IN (angl. AND), ALI (Angl. OR), pa tudi IZKLJUČOJOČI ALI (ang. XOR). V zgornjem primeru smo uporabili funkcijo IN, kar pomeni, da dobimo PRAV, če sta obe rešitvi pravilni (eden IN drugi), za vse ostalo pa NAROBÉ, kar prikazuje preglednica 5.2.

Tabela 5.2: Primer za Boolovo funkcijo IN

A	B	A IN B
PRAV	NAROBE	NAROBE
NAROBE	PRAV	NAROBE
NAROBE	NAROBE	NAROBE
PRAV	PRAV	PRAV

Preglednica 5.3: Primer za Boolovo funkciji XOR

A	B	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

Zapišimo PRAV = 1 in NAROBE = 0 in dobimo zgornjo uganko po sliki 5.7. Za primer XOR dobimo rezultate prikazane v zgornji preglednici 5.3.

5.13 Nevronske mreže in umetni možgani

Resnični možgani so zdrizasti. Njihove živčne celice, žile in precati so napolnjeni s tekočino in so zgrajeni iz lipidnih membran, proteinov in veliko vode. Opis možganov se zdi trdno zasidran predvsem v biologiji in medicini. Obstajajo pa tudi drugi načini razmišljanja o njih, načini, ki so pritegnili zanimanje matematikov, fizikov, inženirjev in računalniških znanstvenikov. O možganih razmišljajo s pomočjo enačb, izdelave računalniških modelov in celo naprav, ki oponašajo delovanje pravih nevronov v naših glavah. (Morris, Filenz, 2003)

Resnični možgani so izredno prilagodljivi. Sposobni so razpoznavati rokopis, ki ga niso videli še nikoli pred tem, razumejo govor popolnih tujcev, sposobni so reševati probleme, ki so jim popolnoma novi. Kljub stalnemu odmiranju celic, delujejo razmeroma dobro skozi vse življenje in so se še v starosti sposobni učiti novih veščin. V primerjavi z njimi so današnji roboti in računalniško vodeni sistemi zelo dobri le pri opravljanju ozkega razpona nalog, za katere so bili načrtovani, a se zelo težko spoprijemajo z nepredvidljivimi okoliščinami in nenačrtovanimi problemi.

Kot smo že omenili, prave možgane sestavljajo tesno povezane biološke nevronske mreže. Nevroni potrebujejo energijo, mreže pa prostor. Naši možgani obsegajo približno 100 milijard živčnih celic, 3,2 milijona kilometrov »žic«, 10^{13} povezav, vse

pa se nahaja v prostornini 1,5 litra, ima maso približno 1,5 kg in potroši borih 10 Watov energije. Če bi teoretično lahko zgradili takšne možgane s pomočjo silicijevega vezja, bi predvidoma za njihovo delovanje potrebovali 10 MW energije, dovolj, da bi s to energijo napajali manjše mesto. Toplota, ki bi jo sproščali takšni možgani, bi bila tolikšna, da bi te možgane takoj stopila.

Izziv, s katerim se soočamo, je odkriti, kako možgani delujejo tako učinkovito in ekonomično, ter uporabiti podobne principe pri ustvarjanju možganom podobnih sistemov. (Morris, Filenz, 2003)

5.14 Gradnja silicijevih možganskih vezij

Izjemno pomembna lastnost možganov je energijska varčnost. *Energijski stroški* prenosa signalov od ene živčne celice do druge so verjetno predstavljali ključni dejavnik pri evoluciji možganov živih bitij. Približno 50–80 % celotne energijske porabe možganov se porabi pri prenosu akcijskih potencialov vzdolž živčnih vlaken ter pri sinaptičnem prenosu. Preostanek se porabi za proizvodnjo in vzdrževanje, kar velja za vse vrste možganov živih bitij.

V primerjavi s hitrostjo digitalnih računalnikov je hitrost živčnih impulzov v možganih zelo nizka, saj znaša le nekaj metrov na sekundo. Razlog je, da so v nasprotju z digitalnimi računalniki, ki so grajeni za zaporedno procesiranje, možgani grajeni kot vzporedna omrežja. Večina živčnih celic se povezuje neposredno z nekaj tisoč drugimi živčnimi celicami. Da jim to lahko uspe, možgani izkoriščajo svojo tridimenzionalnost – možgani, plasti celic, se zgibajo v gube in povezave tesno spletejo v snope. V primerjavi z njimi so povezave med skromnim številom silicijevih nevronov omejene z dvodimenzionalno strukturo čipov in vezij. V nasprotju z možgani je neposredna komunikacija med silicijevimi nevroni zelo omejena, kar je delno možno premostiti z izkoriščanjem velike hitrosti konvencionalne elektronike; impulze iz umetnih nevronov je možno s pomočjo tehnike multipleksiranja združiti in jih prenašati po isti žici. Na ta način skušajo inženirji v siliciju simulirati povezanost bioloških omrežij. (Morris, Filenz, 2003)

Da bi zmanjšali moč in hkrati povečali hitrost, so inženirji, navdihnjeni z nevroni, sprejeli biološko strategijo uporabe *analognega* namesto *digitalnega kodiranja*. Carver Mead (2002) je prenos nevrobiologije v tehnologijo opisal z izrazom *nevromorfno inženirstvo*.

5.14.1 Silicijeva mrežnica

Možgani sesalcev prejemajo vizualne informacije samo takrat, ko oči zaznajo spremembe v sliki, torej v gibanju. Na ta način se zmanjša potrebno število informacij poslanih v možgane, kljub temu pa le-ti dobijo dovolj informacij, da ugotavljajo, kaj se dogaja v okolici. Umetne mrežnice delujejo na enak način. Večina umetnih mrežnic v preteklosti je vsebovala implantat zunanje kamere v očesu in ustrezen zunanji računalnik. Te je v novejšem času nadomestil silicijev čip, ki sta ga razvila Kareem Zaghloul iz Univerze v Pennsylvania in njegov kolega Kwabena Boahen z Univerze Stanford (Zaghloul et al., 2003, 2005, 2007), in predstavlja enostavno umetno verzijo biološke nevronske mreže, ki beleži svetlobo ter avtomatsko prilagaja svoj izhod spremembam v odvisnosti od osvetljenosti. Povezuje se z dvema silicijevima nevronoma, ki – podobno kot resnični nevroni v vidni skorji – izločata informacijo o kotu med črtami ter kontrastnimi mejami na sliki mrežnice. Silicijev čip, ki posnema nevronske mreže realne mrežnice, je pomemben del bionskih oči, ki bodo slepim ali slabovidnim povrnile vid. Čip je sestavljen iz svetlobnih senzorjev, ki delujejo zelo podobno kot živci realne mrežnice, avtomatski filter pa filtrira zaznane vizualne podatke in pošilja le potrebne podatke za ustvarjanje slike sveta. Ta čip je mogoče vstaviti v oko in ga neposredno povezati z živcem, ki prenaša signale v možganski vizualni korteks. Pri tem je impresivno prav to, da deluje pri razpoznavanju gibanja na enak način kot realna (živa) mrežnica.

Nevrone v tem čipu imenujemo »*nevroni seštej in sproži*«. Nevromorfni inženirji jih uporabljajo pogosto. Svoje ime so dobili, ker seštejajo utežen vnos, kodiran v obliki napetosti, ki prihajajo na njihove sinapse, ter se odzovejo s proženjem akcijskega potenciala zgolj kadar napetost preseže postavljeni prag. Silicijevi nevroni so zgrajeni iz tranzistorjev. A namesto, da bi jih uporabljali kot stikala s pomočjo povečevanja napetosti do zasičenosti – enako kot v konvencionalnem digitalnem sistemu –, so tranzistorji uporabljeni v podpraznem razponu. V tem območju se obnašajo podobno kot celične membrane dejanskih nevronov. Dodatni tranzistorji priskrbijo aktivne prevodne lastnosti za posnemanje od napetosti in časa odvisnih ionskih

kanalčkov. Ta majhen vidni sistem je prototip za mnogo kompleksnejše umetne vidne sisteme, ki jih že razvijajo. Uspeva mu prikazati, kako je možno v realnih razmerah izredno hitro procesirati vhod z veliko šuma in oblikovati enostavne odločitve. Uspeva mu opravljati tisto, za kar je bil oblikovan – sporočati orientacijo črte v prostoru. Nevroznanstveniki že uporabljajo ta enostavni silicijev vidni sistem za preizkušanje opreme in trening študentov. Najpomembnejše lastnosti umetnih mrež so, da delujejo v resničnem svetu, v dejanskem času in porabljajo zelo malo energije. (Morris, Filenz, 2003)

Za izdelavo silicijevega čipa so se uporabljale tehnologije, ki se uporabljajo tudi pri izdelavi vseh ostalih čipov v računalniški industriji. Mere čipa so 3,5 x 3,3 mm. Sestavlja ga 5760 silicijevih foto tranzistorjev, ki nadomeščajo na svetlobo občutljive nevrone žive mrežnice. Ti tranzistorji so povezani z več kot 3600 tranzistorji, ki posnemajo živčne celice za prenos informacij o svetlobi in ki posredujejo te informacije možganom v nadaljnjo obdelavo. V celotnem čipu je 13 različnih tipov tranzistorjev, s katerimi posnemamo delovanje različnih tipov realnih živčnih celic. (Zaghloul et al., 2003, 2005)

Danes govorimo že o drugi generaciji umetnih brezžičnih silicijevih mrežnic (angl. Artificial silicon retina – ASR). Gre za čip premera 2 mm in debeline približno 0,02 mm. Sestavljen je iz približno 3500 mikroskopskih sončnih celic, imenovanih tudi »mikrofotodiode«, s svojim lastnim napajanjem. Mikrofotodiode so konstruirane tako, da pretvarjajo svetlobno energijo zaznavane slike v tisoče drobnih električnih impulzov, ki nato stimulirajo delovanje celic mrežnice. (Chader et al., 2009)

5.15 Računalniška nevroznanost

Ko se je konekcionizem razvijal, sta vse pogostejše prihajali v ospredje kognitivna psihologija in UI. Pri tem se uporabljajo nevronske mreže kot medij za modeliranje človeškega duha. Konekcionizem postane presek dveh v preteklosti neodvisnih smeri raziskav uporab mrež, nastalih v osemdesetih letih 20. stoletja. Osnovni cilj konekcionizma je bil ustvariti globlji in bolj kompleksen vpogled v kognicijo, bolj kot je bilo to možno s pravili in simbolnimi predstavami. Oboje, nevronska verodostojnost in računalniška moč, postaneta za konekcioniste zelo atraktivni, predvsem za modeliranje in razumevanje duhovnosti.

V osemdesetih in devetdesetih letih je vrsta kognitivnih znanstvenikov (vključno s konekcionisti, ki so modelirali procese) in nevroznanstvenikov (vključno računalniškimi nevroznanstveniki) iskala sebi lastne poti.

Razviti so bili različni modeli možganskih procesov. Naštejmo le štiri (Bechtel, Abrahamsen, 2002: 309–341):

1. vizualno procesiranje,
2. vloga hipocampusa v spominu,
3. simuliranje disleksije pri modelih mrež za branje,
4. računalniška moč modularnih struktur korteksu.

Očitno je, da konekcionistične mreže pomembno vplivajo na kognitivno znanost. Na področju teh raziskav ne uporabljajo *vs*i termina *konekcionizem* in vse terminologije, ki je s tem povezana. Nekateri imajo raje termine, kot so *nevronske mreže* ali *umetne nevronske mreže* (angl. artificial neural networks – ANN), ki so mnogo bolj nevtralne, vendar nosijo v sebi močan spomin na korenine mrež, katerih zametki izhajajo iz nevroznanosti sredine 20. stoletja. (Bechtel, Abrahamsen, 2002).

V nadaljevanju si bomo ogledali odgovore na vprašanje: *Kaj je ož. bi morala biti zveza med konekcionizmom in študijami možganov?* Odgovore na to vprašanje lahko iščemo v dveh smereh, dveh razlagah.

V prvem valu raziskav sta McCulloch in Pitts definirala idejo enostavnega nevrona. S stališča nevronskega sistema kot računalniškega sistema je ta ideja povzročila intenziven razvoj področja, ki ga danes poznamo kot *računalniško nevroznanost* (angl. *Computational neuroscience*). Pri tem so različni raziskovalci uporabljali najrazličnejše matematične, inženirske in simulacijske tehnike. Od teh so najbolj znani modeli nevronske mreže prve generacije. (Kordigel Aberšek, 2012)

Če se nekoliko bolj osredotočimo na računalniško nevroznanost, lahko ugotovimo, da so pri *drugem valu* raziskav kognitivne psihologije in umetne inteligence pričelo uporabljati nevronske mreže tudi kot medij za modeliranje človeške kognicije.

5.16 Kognitivizem in konekcionizem

Konekcionizem je izstopal iz tradicionalnih paradigem predvsem po tem, da ni gradil kognicije z vpeljavo manipuliranja s simbolnimi paradigmi, temveč je ponudil radikalno drugačen koncept osnovnega procesiranja sistema *razum – možgani*. Osnovna ideja je, da obstaja mreža osnovnih *enot* ali vozlišč, od katerih ima vsaka neko stopnjo aktivnosti. Te enote so med seboj *povežane* tako, da aktivna enota spodbuja ostale. Mreža je *dinamičen sistem*, ki, ko dobi impulz na vходу, razširja dražljaje in zaviranja med svojimi enotami. V nekaterih tipih mrež se ta proces ne ustavi, dokler se ne vzpostavi *stabilno stanje*.

Da bi razumeli konekcionistični sistem, ki izvaja kognitivno nalogo, je potrebna predhodno ustrezna razlaga. To najpogosteje naredimo tako, da opazujemo začetne aktivnosti v sistemu, ko specificiramo problem in opazujemo reševanje problema, ko se vzpostavlja stabilno stanje.

Oba, konekcionistični in simbolni sistem, lahko opazujemo kot računski sistem, vendar je vseeno očitna razlika v konceptih uporabljenega računanja. Pri simbolnih sistemih je uporabljena računska zveza izvedena s transformacijo simbolov po določenih pravilih (npr. aritmetika: učimo pravila, s katerimi izvajamo operacije, določene z nekim simbolom, npr. s simbolom $+$ ali $-$, na drugih simbolih), kar vse skupaj zapišemo v ustreznem programu.

Konekcionistični pogled na računanje je popolnoma drugačen. Osredotočen je na vzročne procese, pri katerih se enote druga za drugo spodbujajo in zavirajo in jih pri tem ne usmerjajo ne shranjeni simboli ne pravila.

5.16.1 Simboli ali mreže v kognitivni znanosti

V kognitivni znanosti obstajajo velike diskusije o tem, ali naj se arhitektura poimenuje *simbolna* ali *konekcionistična*. V starih časih so bili simboli povezani s fizičnimi objekti in so ustvarjali vizualno predstavo nečesa, npr. križ simbolizira krščanstvo. Nasproti temu danes simbole definirajo tudi kot nekaj, kar se npr. uporablja v matematiki in logiki, kjer izračuni temeljijo na operacijah s simboli. Tako je nastala ideja o manipuliranju s simboli. V računalništvu so se pojavili programski jeziki, kot so Fortran, Basic, Pascal, C, C++, Python ipd. Pomemben korak, predvsem

za področje kognitivnega modeliranja, je bil McCarthyjev simbolni manipulacijski jezik LISP, kar je bila osnova tudi za zgodnjo umetno inteligenco. Od takrat ima simbol v kognitivni znanosti le ohlapno povezavo z izvirnim pomenom besede.

Skratka, ne obstaja konsenz o vlogi simbolov pri razlagi duha in o tem, kako naj bo koordiniran z znanji o možganskih procesih. Oglejmo si nekaj razlogov in proti – razlogov o razlagalni vlogi simbolov ali mrež. S + bomo označili, da obstaja zveza, z – pa, da ne obstaja.

1. + Simboli, – Mreže

V tem primeru je položaj simbolov podoben, kot je njihov položaj v LISP-u. Nasprotovanja temu so, da duh ni LISP program in ti simboli ne morejo razlagati duha. Pri tem obstajata dve tradiciji:

- lingvistična tradicija (Noam Chomsky in Jerry Fodor);
- informacijsko-procesna tradicija (Newell in Simon), kar je povezano z bližnjicami, ki so opisane zgoraj (poudarjanje nepomembnosti fizičnih procesov pri realiziranju teh simbolov).

2. – Simboli, + Mreže

To je eliminativni konekcionizem, saj poskuša eliminirati simbole pri razlagi kognicije.

3. + Simboli, + Mreže

Implementacijski konekcionizem verjame, da so konekcionistični izračuni organizirani tako, da dobimo simbolne rezultate. Pri tem imata oba, torej konekcionizem in simbolna karakterizacija pomembno razlagalno vlogo (Smolensky, 1995).

4. – Simboli, – Mreže

Nekateri raziskovalci zavračajo obe, simbolno in konekcionistično razlago. Pri tem ali ponujajo drugačne rešitve ali pa kar tako zavračajo razlago človeškega duha. Zgodovinsko lahko sem prištevamo funkcionalizem in nekatere variante behaviorizma, npr. tistega Skinnerjevega, ki ponujajo druge rešitve.

Anderson (2007) meni, da so vsi ti poskusi razlag kot boj z mlini na veter, da so brezploдне diskusije, saj ni soglasja niti o tem, kaj simboli pomenijo, kaj šele, da bi dosegli soglasje o čem drugem, bolj kompleksnem.

5.16.2 Bionevronske mreže in konekcionistični modeli

V nadaljevanju si bomo podrobneje ogledali stične točke med konekcionističnim modeliranjem in študijami možganskih sistemov in pri tem osvetlili nekatere razlike med konekcionističnimi formalnimi mrežami in biološkimi nevronskimi mrežami in razlike od klasičnih simbolnih modelov do biološkega realizma računalniško podprte nevroznanosti. (Kordigel Aberšek, 2012)

Mnogi praktiki in opazovalci konekcionističnega modeliranja si zastavljajo vprašanje: Kako lahko konekcionistične mreže (formalne mreže) konkurirajo resničnim nevronskim (bio)mrežam možganov? (Peruš, 2000) Spodbujene z originalnimi nevroni (bionevronskimi mrežami možganov) formalne mreže zagotavljajo nekatere očitne podobnosti:

- prisotnost večjega števila enot, umetnih nevronov z različnimi in spremenljivimi stopnjami aktivnosti;
- širjenje aktivnosti po povezavah med enotami (katerih uteži so istovetne sinaptični učinkovitosti);
- adaptivno plastičnost;
- sortirano odzivnost in
- toleranco do napak.

Vendar pa, če problem pogledamo podrobneje, lahko uvidimo, da obstaja tudi ogromno razlik. *Prvič*, obstajajo bistvene razlike med enotami (umetnimi nevroni) in bionevroni, od katerih so tri še posebej pomembne:

- enotam je vnaprej dodeljena aktivacijska vrednost, medtem ko nevroni posredujejo vlak impulzov (angl. spike train), ki sproti ustvarja to aktivacijsko vrednost;
- konekcionistične mreže uporabljajo posamične naprave za seštevanje utežnih aktivnosti zato, da dobijo vhodni signal mreže, medtem ko so povezave v

- možganih vzpostavljene kemično, s pomočjo različnih nevrottransmiterjev z različnimi lastnostmi;
- v konekcionističnih mrežah za vhod v enoto vrednosti preprosto seštevamo, pri biološki nevronih uporabljajo bolj zapletene mehanizme.

Drugič, obstajajo razlike v celotni arhitekturi. V formalnih konekcionističnih mrežah je vsaka enota enega sloja povezana z vsemi enotami naslednjega sloja. Pri živčnem sistemu se vzorci povezav razlikujejo v odvisnosti od tipa celice in njene lokacije. Na primer, v hipokampusu so nekatere povezave bolj redke kot druge.

Tretjič, obstajajo razlike v principih učenja. Če je *delta pravilo*⁵⁴ (*angl. backpropagation ali delta rule*) učenja eden od najbolj razširjenih konekcionističnih algoritmov učenja, ne obstaja nikakršen dokaz, da takšen mehanizem velja tudi za procese učenja v živčnem sistemu. Celo nasprotno, učenje sproščajo neprimerno bolj lokalni signali.

Prikažimo delta pravilo na enostavnem primeru 3-slojne mreže (samo en sloj je skrit):

```

Initialize the weights in the network (often randomly)
Do
  For each example e in the training set
    O = neural-net-output (network, e) ; forward pass
    T = teacher output for e
    Calculate error (T - O) at the output units
    Compute delta_wh for all weights from hidden layer to output layer ; backward pass
    Compute delta_wi for all weights from input layer to hidden layer ; backward pass
  continued
  Update the weights in the network
Until all examples classified correctly or stopping criterion satisfied
Return the network

```

```

Inicializiraj uteži v mreži (pogosto so naključne, uporablja se neka funkcija ali metoda generiranja
naključnih števil, npr. metoda Monte Carlo)
Naredi
Za vsak primer e v nizu urjenja
  O = izhod nevronske mreže (mreža, e); prepustnica za naprej
  T = učiteljev izhod za e
  Izračunaj napako (T - O) na izhodnih enotah
  Izračunaj delto_wh za vse uteži od skritega do izhodnega sloja; prepustnica za nazaj

```

⁵⁴ Delta pravilo ali pravilo širjenja napake je splošna metoda učenja umetnih nevronske mrež. Z njim lahko izračunamo ali določimo katerikoli izhodne vrednosti za podane znane vhodne vrednosti.

Izračunaj δ_i za vse uteži od vhodnega sloja do skritega sloja; nadaljevanje prepustnice za nazaj
 Obnovi uteži v mreži
 Dokler niso vsi primeri označeni za pravilne ali je pogoj za zaustavitev zadovoljen
 Vrne se v mrežo

Kot izhaja že iz imena pravila samega se napaka (in zato tudi učenje) širi od izhodnih vozlišč proti vhodnim, skratka *vzvratno*. Tako, tehnično rečeno, se delta pravilo uporablja za računanje gradienta napake mreže z upoštevanjem spremenljivih uteži vozlišč mreže, podobno, kot se dogaja pri bioloških nevronskih mrežah in akcijskem potencialu. Ta gradient nato skoraj vedno uporabimo v preprostem *stohastičnem algoritmu*, s pomočjo katerega iščemo uteži, ki minimizirajo napako. Mnogokrat termin »vzvratno« uporabimo v bolj splošnem pomenu, da z njim opišemo celoten postopek, ki spremlja tako računanje gradienta, kot tudi njegovo uporabo stohastičnega gradienta. Metoda nam omogoča hitro konvergiranje rešitve do zadovoljivega lokalnega minimuma napake.

Vzvratne mreže so v bistvu večslojni perceptroni (po navadi z enim vhodnim, enim skritim in enim izhodnim slojem). Da bi lahko skriti sloj služil svojemu namenu, morajo vsebovati večslojne mreže nelinearno aktivacijsko funkcijo, saj če večslojne mreže uporabljajo samo linearno aktivacijsko funkcijo, potem je to enako, kot bi uporabljali enoslojne linearne mreže. Najpogosteje uporabljena nelinearna aktivacijska funkcija vsebuje logistično funkcijo⁵⁵, softmax funkcijo⁵⁶ ali Gaussovo funkcijo.⁵⁷

To, da je delta pravilo s stališča bionevronov in nevroznanosti dokaj neverjetno, je spodbudilo mnoge komentarje in poudarilo razlike. (Philips, 1997, Rolls, Treves, 1998) Naštajmo nekatere, najpomembnejše:

- ne obstajajo dokazi, da nevronski računalnik signal napake, ki upošteva prispevke vseh ostalih nevronov, obravnava kot delta pravilo;

⁵⁵ Logistična porazdelitev je družina zveznih verjetnostnih porazdelitev z dvema parametroma (lokacije in merila). Njena zbirna funkcija verjetnosti je logistična funkcija, po kateri je dobila ime. Po obliki spominja na normalno porazdelitev, ima samo močnejše repe (večjo sploščenost).

⁵⁶ Softmax aktivacijska funkcija je nevronska transferna funkcija, ki v nevronskih mrežah prenaša funkcijske izračune od izhodnega nazaj k vhodnemu sloju. Je metoda, ki je najbolj verjetno podobna biološkemu procesu. Uporablja se predvsem pri utrjevalnem učenju, predvsem za pretvorbo določenih vrednosti v akcijsko verjetnost.

⁵⁷ Normalna porazdelitev (tudi Gaussova porazdelitev) je verjetnostna porazdelitev vrednosti statističnih enot v statistični populaciji, ki je v grafični predstavitvi oblikovana v obliki zvona oziroma Normalne krivulje. Standardna normalna porazdelitev je porazdelitev vrednosti s povprečjem (aritmetično sredino) 0 in standardnim odklonom 1.

- tudi če bi nek nevron izvršil ta izračun, bi bilo verjetno neverodostojno trditi, da bi ta lahko poslal rezultate nazaj nevronom, s katerimi je povezan, ter nadalje vsem ostalim nevronom, ki so z njim povezani;
- delta pravilo deluje tako dobro v posameznih delih prav zato, ker je število skritih enot majhno zaradi možnosti posploševanja, vendar pa nevronska arhitektura ne dovoljuje impliciranja informacij od spodaj navzgor;
- kot algoritem, ki nadzira učenje, delta pravilo zahteva, da je ciljni izhodni vzorec določen v vsakem koraku. To so informacije, ki so načelno neznane znotraj sistema. Kot je na primeru prikazano, lahko mreže napovedujejo svoje naslednje vhodne podatke kot del izhodnih podatkov. Čeprav je to obetavna ideja, ki jo moramo empirično še preveriti, ni nikjer dokazov, da realni možgani zares tako delujejo. (Bechtel, Abrahamsen, 2002)

Ko so se pojavili prvi konekcionistični modeli, so bili poimenovani kot kognitivni in ne kot možganski modeli. Ljudi je vznemirjalo predvsem to, da imajo ti sistemi bistvene prednosti pred klasičnimi simbolnimi modeli kognicije – predvsem zaradi fleksibilnosti in adaptivnosti; in sicer tako, da so se upoštevale pri njih zelo splošne karakteristike živčnega sistema znotraj neke, ponovno zelo splošne večnamenske arhitekture. Vendar so imeli prvotni mrežni sistemi tudi veliko pomanjkljivosti. *Mreže so lahko večslojne ali pa se lahko učijo – nikakor pa ne zmorejo obojega!*

Ko je bil razvit algoritem po delta pravilu, je to zelo vplivalo na nadaljnji razvoj. S pomočjo algoritma za večslojne mreže je postal način učenja vedno bližji človeškim sposobnostim, tako da so postale mreže vedno resnejši tekmeč simbolnim modelom. Vse te prednosti v računski uspešnosti in psihološki realnosti so rezultat predvsem biološke realnosti. Ko se je nevroznanost (in še posebej kognitivna nevroznanost) v devetdesetih letih 20. stoletja razbohotila, so postala vse bolj izrazita vprašanja:

- Kako lahko konekcionistične mreže sploh opravijo predvidene funkcije?
- Ali je to usodna zabloda, saj konekcionistične formalne mreže niso biološko verjetne?
- Ali so konekcionistične mreže sploh veljavne, če vsebujejo te biološko neverjetne mehanizme?

5.17 O interdisciplinarnosti v kognitivni znanosti

Informacijska oziroma računalniška metafora delovanja kognicije je, kot rečeno, v zadnjih letih izgubila povezovalno moč, ki jo je imela nekoč. Pokazalo se je, da razumnost računalnikov ne raste sorazmerno z njihovo zmogljivostjo (oziroma da sploh ne raste), kar je omajalo splošno vero v to, da ponujajo računalniki (oziroma Turingov stroj) zadovoljiv model delovanja spoznavnih procesov, čustev, še najmanj pa zavesti. Znanstveniki in filozofi so se začeli ozirati po alternativnih razlagah, po drugačnih (večinoma bolj sinergetskih) metaforah in tudi po novih epistemoloških temeljih.

Z dvomi v prvotno intuicijo o tem, kako razložiti kognitivne pojave, pa se ni zmanjšal interes za njihovo raziskovanje. Prav nasprotno. Raziskave na tem področju so vse bolj razširjene. Osrednje mesto jim namenjajo vodilne znanstvene publikacije, kot kaže pa se za raziskovanje kognicije počasi odpirajo tudi finančni viri.⁵⁸

Velika »zasluga« t. i. informacijske metafore je torej usmeritev pozornosti na pred tem zapostavljeno področje kognicije. Tudi če se izkaže, da so bili upi iz sedemdesetih in osemdesetih let 20. stoletja previsoki, je v znanstveni skupnosti že prevladal konsenz, da je to področje vredno raziskovalne pozornosti. Seveda se postavlja vprašanje: Zakaj sploh potrebujemo kategorijo »kognitivna znanost«? Kaj novega prinaša ta kategorija? Mar niso dovolj posamezne discipline, ki se na sebi lasten način ukvarjajo s kognicijo in z njo povezanimi pojavi? Kot rečeno, doživljajo nekatere od teh disciplin (v zadnjem času predvsem nevroznanost) skokovit samostojen napredek. Spet druge (npr. psihologija) imajo raziskovanje kognitivnih pojavov med svojimi osnovnimi nalogami in so se z njim začele ukvarjati že mnogo pred pojavom kognitivne znanosti. (Kordigel Aberšek, 2012)

Vprašanje o smislu obstoja kognitivne znanosti je seveda bolj retorično. Navsezadnje je področje kognicije in zavesti najbolj intimno povezano z našim bivanjem in si zato prav gotovo zasluži svojo raziskovalno disciplino, ki bi ga raziskovala kot celoto. Razumljivo je, da nobena partikularna disciplina ne more dati polnega odgovora, niti ne more zadovoljivo razložiti fenomenov, kot so npr.

⁵⁸ Raziskava Deutsche Forschungsgemeinschaft je postavila kognitivno znanost kot eno izmed dvanajstih najbolj relevantnih področij raziskovanja – klasifikacija, ki ima direkten vpliv na financiranje.

spomin, mišljenje, čustva itd. Zatorej so zgornja vprašanja bolj pozivi, ki kažejo, kam je treba usmeriti energijo: v iskanje možnosti združevanja ali/in novih načinov skupnega raziskovalnega dela. Vprašanje o smislu obstoja kognitivne znanosti kot samostojne discipline je torej predvsem metodološko vprašanje. Na tem področju se soočamo z naslednjimi dilemami:

- je mogoče z združitvijo izsledkov posameznih disciplin o istem fenomenu priti do boljšega poznavanja;
- je skupek spoznanj različnih disciplin kaj več od preprostega spiska posameznih dosežkov;
- oziroma, je sinergetsko združevanje spoznanj, v katerega bi bile vključene relevantne discipline, sploh možno (kako naj na primer sodelujeta nevroznanost in sociologija)?

Intuicija pravi, da je interdisciplinarno delo možno in smiselno. V praksi so takšni projekti redki in še tistih nekaj se mnogokrat izkaže kot larpurlartistično prizadevanje brez pravega (znanstvenega) učinka. Problemi, povezani z interdisciplinarnim sodelovanjem pod dežnikom kognitivnih znanosti, so v resnici ogromni oziroma »težki« v Chalmersovem pomenu.

Če povzamemo, se vsi raziskovalci strinjajo o osnovnem namenu kognitivne znanosti: bolje *razumeti* kognitivne pojave in procese. Od tu naprej pa se bistveno razlikujejo. Navedimo nekaj bistvenih razlik (Kordež, Markič, 2007):

1. Vprašanje, kaj sodi med kognitivne pojave. So to le pojavi, ki jih v to kategorijo uvršča klasična psihologija (učenje, odločanje, pomnjenje, mišljenje itd.)? Spadajo sem tudi čustva (zelo priljubljeno področje kognitivne znanosti v anglo-saškem svetu)? Kaj pa zavest? Pojem *zavest* je bil še nedavno tega neznanstven in so se ga vsi resni znanstveni diskurzi izogibali, v zadnjem času pa postaja raziskovanje zavesti osrednji del kognitivne znanosti.
2. Vprašanje o končnem cilju. Večina avtorjev se strinja, da takšnega cilja ni – da je raziskovanje neskončen proces in da sprotni rezultati narekujejo nadaljnjo smer. Mnogim avtorjem se zdi trenutni raziskovalni problemi dovolj težki (in natančno zastavljeni), da bi se ukvarjali še z »velikimi« vprašanji, kot je npr. zavest. Odgovori tistih, ki so cilj vseeno konkretizirali, zajemajo vse, od univerzalnega modela kognicije in razlage povezave med fizičnim in duševnim

do praktičnih napotkov za zdravljenje patologij kognitivnega aparata. Na splošno lahko ugotovimo, da si predstavniki klinične prakse ne zastavljajo »velikih« vprašanj – zanima jih, kako bi lahko izboljšali obstoječe prakse zdravljenja in diagnostike in takšne odgovore pričakujejo tudi od kognitivne znanosti.

3. Vprašanje, kaj mislimo s tem, ko govorimo o »razumevanju« kognitivnih pojavov? To vprašanje se večini avtorjev ne zdi dovolj aktualno/pomembno, da bi se mu posvečali. Kljub temu nekateri opozarjajo na različne možnosti razumevanja oziroma na različne vrste znanja, ki ga lahko pridobimo o kognitivnih pojavih. Vprašanje razumevanja se nanaša na osrednji problem, ki pesti moderno kognitivno znanost: Je možno pojave, kot so kognicija, zavest, razum razložiti algoritmično oziroma z vzročno – posledičnimi modeli? Problem pri raziskovanju kognicije je, da skoraj vsako raziskovanje bistveno spreminja raziskovani pojav. To pomeni, da je zelo težko zagotoviti neudeleženost raziskovalca pri raziskovanem procesu in da so povsem ponovljivi poskusi skoraj nemogoči. Obe našteti lastnosti (neudeleženost raziskovalca in ponovljivost eksperimentov) tvorita temelje sodobne (naravoslovne) metodologije in sta nujni za tvorbo relevantne znanstvene teorije. Večina »resnih« znanstvenih raziskav kognitivnih fenomenov poskuša ustvariti čim boljše približke objektivnih in ponovljivih eksperimentov, ne moremo pa se popolnoma izogniti vprašanju, kaj če s tem izgubljammo bistveno lastnost kognicije, o kateri nas prepričuje njeno neposredno izkustvo: njeno enkratnost in neponovljivost.
4. Je obravnava kognitivnih pojavov na enak način kot ostalih naravnih pojavov res najbolj (edina) primerna? Mar ni znanje, ki ga resnično potrebujemo, znanje »od znotraj« oziroma znanje »kako«? Tudi če predpostavljamo, da je kognitivne pojave (na čelu z zavestjo) možno razložiti z logičnimi modeli, se postavlja naslednje vprašanje: Nam takšni modeli lahko koristijo pri razumevanju samih sebe – našega izkustvenega sveta? Če ne, kako bi lahko (znanstveno) raziskovali svoje doživljanje?
5. Pri raziskovanju kognicije je zelo pomembna tudi etična komponenta. Sem sodijo vprašanja, ki si jih zastavljajo (ali pa bi si jih morali zastavljati) predvsem predstavniki klinične prakse: Kaj je patološko in kaj normalno (gre za resno vprašanje v psihoterapiji in klinični psihologiji)? Kje so meje človekove zasebnosti (sodobne metode kažejo potenciale, da lahko posegajo v

- najintimnejša doživlja sebe in sveta) in še mnoga druga, ki presegajo namen pričujočega besedila.
6. Na koncu seznama problemov pri raziskovanju v kognitivni znanosti dodajmo še povsem praktično težavo, na katero naleti prav vsaka interdisciplinarna raziskovalna skupina – problem nepoznavanja ostalih sodelujočih disciplin in njihovega strokovnega jezika. Še enkrat moramo poudariti, da pravo interdisciplinarno raziskovalno delo ni le preprost seštevek raziskav na posameznih področjih, ampak je zanj potrebna prilagojena metodologija. Ko govorimo o interdisciplinarnem združevanju, ne mislimo na sodelovanje sorodnih ved (npr. biologija in kemija) – takšno združevanje še ne pripelje do resnih teoretskih težav, saj si vede v veliki meri delijo znanstveni jezik in raziskovalne metode. Problemi, ki jih prinaša raziskovanje kognicije, ne morejo biti rešeni znotraj ozkega okvira zgolj naravoslovnih (niti zgolj družboslovnih ali humanističnih) ved, zaradi tega je ena od glavnih ovir pri celostnem raziskovanju pomanjkanje povezovalnih strategij, ki bi omogočale srečevanje celotnega spektra znanstvenih disciplin.

Zanimivo je, da se s to ugotovitvijo vračamo k osnovni ideji kibernetike, ki se je začelo v petdesetih letih 20. stoletja in ki ga mnogi vidijo kot neke vrste valilnico kognitivne znanosti. Margaret Mead, ena od ustanoviteljic tega gibanja, je nekoč izjavila:

»Še posebej želim poudariti pomembnost množice transdisciplinarnih idej, ki [...] jim pravimo 'kibernetika' – gre za obliko transdisciplinarnega razmišljanja, ki je omogočila predstavnikom mnogih disciplin medsebojno komunikacijo v jeziku, ki je bil vsem razumljiv.« (Margaret Mead v Von Foerster, 1974:10)

V začetnem obdobju kibernetike (in kognitivne znanosti) se je informacijska metafora kazala kot tisto vezno tkivo, ki lahko omogoči skupno delo različnih disciplin. Zdaj, ko se je ta metafora izkazala za nezadostno, čaka kognitivno znanost izziv razvoja pristopov za skupinsko interdisciplinarno delo. Zgodbe zadnjih let to lepo ponazarjajo, saj se niso pričele mehčati le meje med simbolno in konekcionistično strujo, temveč so se pričele ustvarjati nove povezave in novi modeli, temelječi na specializiranih podskupinah konekcionista. Najpomembnejše povezave, ustvarjene v tem obdobju, so povezave med *dinamični modeli*, o katerih bomo govorili v nadaljevanju, *utelešeno kognicijo* in *kognitivno nevroznanostjo*.

Ti novi načini razmišljanja bodo zagotovo pripomogli k vznemirljivim odkritjem v desetletjih 21. stoletja.

5.18 Sklep

Kot smo že nakazali v uvodu, obstajajo teorije, ki podpirajo diskretno, in teorije, ki podpirajo zvezno modeliranje kognitivnih procesov. Tako obstajajo trije ločeni pristopi modeliranja duha, to so *kognitivno simbolni* sistemi, *kognitivno konekcionistični*, ki za izhodišče modela uporabljajo nevronske mreže, in sistemi *modeliranja na osnovi teorije dinamičnih sistemov*. Poleg tega pa se predvsem v zadnjih desetletjih vse pogostejše pojavljajo razni mešani modeli, ki smo jih poimenovali hibridni sistemi. Ker menimo, da so se simbolni sistemi v samostojni obliki že nekoliko izpeli, se v delu osredotočamo predvsem na konekcionistične in dinamične sisteme, simbolne sisteme pa obravnavamo le v povezavi s konekcionističnimi, kot del hibridnih sistemov. Definiramo, kaj so biološke in kaj formalne nevronske mreže, pregledamo zgodovinski oris uporabe teh modelov v kognitivni znanosti, kako nevroznanost obravnava nevronske mreže, predvsem formalne, in kako na osnovi njih modelira umetne možgane. Nakazane so smernice razvoja in prikazane možnosti izgradnje silicijevih nevronskih mrež kot osnove umetnih možganov. In na koncu izpostavimo paradoks, da danes še vedno poskušamo modelirati delovanje duha (delovanje možganov), ki izkazuje izrazito zvezni proces delovanja na konvencionalnih, digitalnih računalnikih, ki delujejo izključno na diskretnem principu.

Vsak od teh pristopov h kognitivni arhitekturi poudarja posamezne značilnosti kognicije in se kaže kot primerna arhitektura zanje. Tako se simbolni klasični modeli kažejo kot precej uspešen pristop za modeliranje višjih kognitivnih funkcij, ki zahtevajo uporabo jezika, konekcionistični modeli pa pri prepoznavanju vzorcev. Pristop teorije dinamičnih sistemov in »utelešene kognicije« nas opozarja, da moramo najprej razumeti biološke rešitve za preproste naloge, kot so na primer občutljivost na svetlobo, premikanje in izbira dejanja. Po drugi strani pa za vsakega od teh pristopov poznamo tudi težke naloge, ki jim (za enkrat?) vse te teorije niso kos. Ena takih nalog je prav gotovo *zavest*.

Na koncu poudarimo, da so poskusi, da bi izdelali inteligentne stroje, povzročili revolucijo v mišljenju o mišljenju. Kot pravijo kognitivni znanstveniki, predstavlja računalnik naše novo zrcalo. Računalniška metafora je odprla nove, revolucionarne

poti v raziskovanju človekovih spoznavnih procesov. Možnost modeliranja človekovega spoznavanja je namreč omogočil šele nastop računalnika. Pred tem so obstajali zgolj opisi miselnih procesov, ki so temeljili na notranjem uvidu, na introspekciji.

6 Kognitivne hibridne arhitekture

6.1 Hipoteza fizikalnega simbolnega sistema

Hipotezo fizikalnih simbolnih sistemov, ki sta jo leta 1975 predlagala računalniška znanstvenika Herbet Simon in Allen Newell, lahko uporabimo tudi za analogijo med duhom in računalnikom. Po tej hipotezi lahko vsa inteligentna vedenja opisujemo s transformacijo fizikalnih simbolov, ob upoštevanju določenih pravil. Simon in Newell sta se naloge razlage duha lotila na zelo pragmatičen način in se pri tem nista dosti poglobljala v to, kako duh zares procesira informacije. V nasprotju z njima, pa se je filozof Jerry Fodor, prav tako zagovornik te hipoteze, problema lotil neprimerno bolj sistematično in manj pragmatično. Trdil je, da mora biti procesiranje simbolnih informacij lingvistično in razvil subtilno in sofisticirano teorijo o tem, zakaj. Trdil je, da je arhitektura duha zgrajena okrog jezika misli. V središču zelo splošne hipoteze fizikalnih simbolnih sistemov je zelo natančna hipoteza *jezika misli*, ki ostro ločuje med *sintakso* informacijskega procesiranja (fizičnim manipuliranjem simbolnih struktur) in *semantiko* informacijskega procesa (vsebino). (Fodor, 1983)

Herbert Simon in Allen Newell sta leta 1975 prejela Turingovo nagrado kot izjemno vplivna računalniška znanstvenika in pionirja umetne inteligence za svoji deli – programa *Logic Theory Machine* (1957) in *General Problem Solver – GPS* (1956). Za pridobitev te nagrade pa je bilo potrebno izvesti tudi ustrezno predavanje. Na tem

predavanju sta Simon in Newell predstavila splošen princip razmišljanja o inteligentnem informacijskem procesiranju, predstavila sta manifest, ki se je dotaknil tako študije človeškega duha, kot tudi pojava umetne inteligence. Manifest je bil to, kar danes imenujemo *hipoteza fizikalnih simbolnih sistemov* ali krajše *hipoteza simbolnih sistemov*. (Copeland, 1993/2007)

V svojem predavanju sta predlagala, da je fizikalni simbolni sistem potreben in zadosten pogoj za splošno inteligentno delovanje. Iz tega izhajata dve zahtevi:

- *Prva (zahteva o potrebi)* je, da nič ni sposobno inteligentnega vedenja, če tega ni možno predstaviti kot fizikalni simbolni sistem. Ker so ljudje sposobni inteligentnega vedenja, to pomeni, da mora biti tudi človeški duh fizikalen simbolni sistem;
- *Druga (zahteva o zadostnosti)* pa je, da ni v principu nobenih ovir za konstruiranje umetnega duha, saj imamo za to primerno orodje, to je fizikalni simbolni sistem.

Pomembnost teh zahtev je odvisna predvsem od tega, kaj naj predstavlja fizikalni simbolni sistem. Če se vrnemo k Simonu in Newellu in njuni definiciji:

Fizikalni simbolni sistem sestoji iz niza entitet, poimenovanih simboli, ki imajo fizikalen vzorec in se lahko pojavijo kot komponenta drugega tipa entitete, ki jih imenujmo *izrazi* ali *simbolne strukture*. Te simbolne strukture so sestavljene iz številnih primerov (ali znakov) simbolov, povezanih na nek fizikalen način. Sistem vsebuje v kateremkoli trenutku zbirko teh simbolnih struktur. Poleg teh struktur, pa simboli prav tako vsebujejo zbirko procesov, ki izvajajo operacije s temi simbolnimi strukturami, zato da ustvarjajo druge strukture; ustvarjanje, modificiranje, reproduciranje in uničenje. Fizikalni simbolni sistem je skratka stroj, ki v času proizvaja razvijajočo se zbirko simbolnih struktur. (Bermudez, 2010: 147)

Iz te definicije izhajajo štiri ideje v zvezi s simbolnim sistemom:

1. *Simboli so fizikalni vzorci*. Za Simona in Newella so simboli v simbolnih sistemih enaki, kot so npr. simboli – črke v abecedi. Princip je identičen kot trak s simboli pri Turingovem stroju. V realnem svetu pa to seveda ni popolnoma tako, že če si ogledamo primer digitalnega računalnika. Ta ima, poleg simbolne strukture

(programske opreme, angl. software), tudi strojno opremo (angl. hardware). V strojni opremi, tudi če bi računalnik razstavili, ne bi našli nobenih simbolov, temveč le žice, vezja, stikala. Vendar pa je takšno stikalo lahko vključeno ali izključeno, tako da bi ga simbolno prav tako lahko predstavili s simboli, npr. z 1 (vključen) in 0 (izključen).

2. *Simbole lahko kombiniramo in tako ustvarjamo kompleksne simbolne strukture.* Če nadaljujemo z jezikovno metaforo: simbole naše abecede lahko kombiniramo v besedne strukture in te nadalje kombiniramo v stavčne strukture. Proces, kako to narediti, pa ustrezajo točno določenim pravilom, v jezikovni metafori, slovničnim pravilom. Ta pravila so najverjetneje *rekurzivna pravila*⁵⁹, kar poenostavljeno pomeni, kako iz sprejemljive kombinacije simbolov tvorimo še bolj kompleksne, sprejemljive kombinacije. Turingov stroj zna sicer zna posneti le posamično celico istočasno, vendar pa je sposoben delati s kompleksnimi strukturami zato, ker te kompleksne strukture lahko gradi iz posamičnih simbolov sosednjih celic. Za to pa mora poznati samo dve reči; poznati mora vrstni red simbolov, kateri simbol sledi predhodnemu, in na nek način mora poznati zaključek, tj. s katerim simbolom se niz konča, oz. s čim dobimo kompleksni simbol.

Naslednji ideji sta bistveni za hipotezo simbolnih sistemov, tj. za idejo, da lahko reševanje problemov razumemo kot transformacijo simbolnih struktur na bazi pravil.

3. *Fizikalni simbolni sistem vsebuje tudi procese za manipuliranje s kompleksnimi simbolnimi strukturami.* V predhodnih idejah smo videli, kako iz enostavnih elementov tvorimo kompleksne simbolne strukture. V resnici pa je pomembno predvsem to, kaj sistem počne s temi kompleksnimi simbolnimi strukturami, kako iz ene izpeljuje drugo. Hipoteza simbolnih sistemov je hipoteza o inteligenci in inteligentnem ukrepanju. Pri tem ima ta hipoteza različne zahteve, tj., da je razmišljanje enostavno le transformacija simbolnih struktur po določenih pravilih. Skratka, vsak sistem, ki je sposoben transformirati simbolno strukturo na dovolj sofisticiran način, velja za inteligentnega. Za Simona in Newella je bila bistvena zahteva sposobnost reševanja problemov. Pri tem sta definirala

⁵⁹ Rekúrzija (lat. recurrere), kar pomeni teči nazaj.

prostor iskanja (angl. search-space), ki predstavlja vsa stanja, ki jih lahko dosežemo iz izhodiščnega stanja ob uporabi dopustne transformacije. Računalničarji prostor iskanja navadno predstavljajo z *drevesno strukturo*. Oglejmo si to na enostavnem primeru. GPS program uporablja tehniko reševanja problemov, imenovano *analiza ciljev in sredstev*⁶⁰. Predstavlja proces v treh korakih, ki postopno zmanjša razlike med trenutnim stanjem in končnim stanjem (ciljem):

1. razišče razlike med trenutnim stanjem in ciljnim stanjem;
2. določi transformacijo, ki bo zreducirala razlike med tema stanjema;
3. preveri, ali je izbrana transformacija uporabna v konkretnem primeru;
 - če je, potem jo uporabi in se vrni na začetni korak;
 - če ni, se vrne v drugi korak in izbere novo transformacijo.

Analiza sredstev in ciljev je analiza, ki sta jo Simon in Newell poimenovala *hevristično iskanje*. (Bermudez, 2010: 153)

4. *Procesi za ustvarjanje in transformiranje kompleksnih simbolnih struktur so tudi sami sestavljeni iz simbolov in simbolnih struktur znotraj celotnega sistema.* Prednost »splošnih« računalnikov pred npr. kalkulatorji je predvsem v tem, da lahko izvajajo več različnih operacij (programov), mnogokrat tudi istočasno, kar jim omogoča simbolna struktura, ki dekodira informacije in navodila in jih lahko uporablja v različnih programih. Pravzaprav je ravno to za Simonova in Newellova potreben pogoj za inteligentne sisteme. Za tem se nahaja pomembno teoretično odkritje o Turingovem stroju s stališča matematične logike. Turingov stroj je abstrakten model računalniške naprave, torej individualne (specializirane) naprave za reševanje individualnega (specializiranega) problema, pa ne samo to, Turing je ugotovil, da obstaja tudi poseben stroj, ki ga je poimenoval univerzalni stroj (danes to imenujemo *univerzalni Turingov stroj*). In ta univerzalni stroj lahko posnema katerikoli individualni (specializirani) stroj z uporabo individualnega algoritma. Skratka predstavlja posebno obliko računalnika za splošne namene, ki lahko simulira

⁶⁰ Analiza sredstev in ciljev (angl. mean-ends analysis – MEA) je tehnika, ki se uporablja v umetni inteligenci za kontrolirano iskanje napak in reševanje problemov pri računalniških programih. Prav tako pa se od leta 1950 uporablja tudi kot kreativno orodje pri inženirskih problemih in metodah konstruiranja. Ta metoda je tudi način, kako razjasniti neko misel pri matematičnem dokazovanju.

druge bolj specializirane stroje – računalnike. To mu omogoča, da lahko strojne tabele specializiranih Turingovih strojev pretvorimo v številke, kar nato služi kot vhodni podatek. In prav to je ustvarila hipoteza fizikalnih simbolnih sistemov pri karakterizaciji inteligentnih sistemov.

Hipoteza simbolnih sistemov je preveč temeljna, da bi jo bilo moč empirično preveriti. Vrsta strokovnjakov trdi, da je popolnoma zgrešena, saj da nima nikakršen zveze z inteligenco. Eden od najbolj izpostavljenih je John Searle s svojim argumentom iz kitajske sobe. (Searl, 1980) Kakorkoli že, če želimo narediti korak naprej pri tej ideji, potem ta zagotovo potrebuje vsaj odgovore na vprašanja, kot so: Kaj te simbolne strukture so, kako jih transformiramo in kako so te transformacije povezane z inteligenco? Skratka, pogovarjamo se o kognitivni arhitekturi, zato moramo raziskati, kako lahko služi simbolni sistem kot model človeške kognicije in kako se informacije procesirajo v človeških možganih. Skratka, govorimo o *hipotezi jezika misli*, ki jo je razvil Jerry Fodor (1983).

6.2 Manipulacije s simboli

Računalnik ni nič drugega kot stroj za manipuliranje s simboli. Vse, kar počne, je mešanje simbolov po navodilih, ki so zapisana v programu. Razložimo na kratko; za trenutek si zamislimo simbole kot vrstice krogcev in križcev na črtastem papirju:

X X O X

X O O X

O O O X

O O X X

Navedimo nekaj primerov manipuliranja s simboli, ki jih lahko računalnik izvaja:

- kopira lahko simbole iz ene vrste v katerokoli drugo vrsto;
- briše lahko katerikoli simbol v poljubni vrsti;
- zapiše nove sekvence simbolov v določeni vrsti;
- prepozna simbol, ki se pojavi na določenem mestu, npr. stroju lahko zapovemo, da kontrolira list od zgoraj navzdol, kontrolira prve tri simbole v posamezni vrsti ter se ustavi, ko najde kombinacijo OOO;

- primerja simbole v dveh določenih vrstah, kar je lahko povezano z navodilom, da napiše križec v naslednjo vrsto, če sta dve skupini simbolov različni, ali da napiše krogec, če sta enaki;
- zapiše ime posamične vrste z določeno skupino simbolov. Enostaven način, kjer stroj lahko to naredi, je pri poimenovanju simbola (npr. 1001011 je koda za 'K'. (Copeland, 1993/2007: 68)

Zgoraj opisane operacije so znane kot temeljne (ali primitivne) simbolne manipulacije. Računalnik je naprava za manipuliranje simbolov s sposobnostjo izvajanja določenega seznama temeljnih operacij. (Copeland, 1993/2007) Vse omenjeno pa si oglejmo na enostavnem primeru z začetka razvoja UI.

6.2.1 Primer – Eliza

Poskusimo analizirati simbolne sisteme na primeru Elize, računalniškega programa za enostavno procesiranje naravnega jezika. Ta program je verjetno najbolj znan program s področja UI na svetu in je, po mnenju stroke verjetno tudi najstarejši. Izdelal ga je v sredini šestdesetih let 20. stoletja Joseph Weizenbaum iz MIT (Massachusetts Institute of Technology). Glavna naloga programa je bilo posnemati pogovor med psihoanalitikom in njegovim pacientom, po vzoru *govorne terapije*,⁶¹ ki jo je razvil psiholog Carl Rogers v štiridesetih in petdesetih letih 20. stoletja. Zakaj smo izbrali prav to terapijo? Zato ker Eliza za reševanje problema ne zahteva poznavanja raziskovalnega subjekta, kar omogoča popolnoma splošne odzive, kot npr.:

- Na pacientovo trditev: »Boli me glava«, se terapevt odzove npr. z: »Zakaj ste rekli, da vas boli glava?« ali
- na pacientovo trditev: »Mama me sovraži«, se lahko terapevt odzove z: »Kdo v vaši družini vas še sovraži?«

⁶¹ Govorna terapija temelji na slabem poznavanju pacienta, tako da se psihoanalitik lahko odziva zelo splošno.

Tudi Eliza reagira na podoben način. Temelji na principu tehnike *ujemanja vzorcev* in deluje tako, da odgovore preoblikuje v novo vprašanje, npr.:

- Če natipkate: »Imel sem zelo slab dan«, se Eliza odzove z npr.: »Ali si prišel k meni zato, ker si imel zelo slab dan?«

ali

- če natipkate: »Prišel sem se pogovoriti s teboj«, se Eliza odzove z npr.: »O, k meni si prišel, da bi se pogovorila?«

Skratka, odziv je sestavljen iz dveh delov stavka. Na zgornjem primeru: prvi del stavka je zapisan v bazi podatkov programa in se glasi: »O, k meni si prišel, da*...« (zvezdica pomeni, da se stavek nadaljuje). Drugi del stavka pa izhaja iz vhodnega stavka (prišel sem se pogovoriti s teboj), ki ga program preoblikuje v: »...*bi se pogovorila«.

Če se bo vzorec »morda« pojavil v vhodnem stavku, potem program poišče odziv iz spiska stavkov, kot npr.: »Ne zdite se mi popolnoma prepričan« in prisotnost besede »vsakdo« v vhodnem stavku, usmeri program k odgovoru kot: »V mislih imate določeno osebo, ali ne?« Poleg tega je Eliza sposobna tudi konstruirati samostojne odzive, tako da preoblikuje vhodni stavek. Vsak vzorec, shranjen v programu, je povezan s spiskom transformacij, od katerih vsaka pripelje do izhodnega stavka, le če vhodni stavek vsebuje vzorec (ključno besedo) v vprašanju. Npr. eden od shranjenih vzorcev je: »me ... ti ...«. To se ujema z vzorcem v stavku: »Zakaj me ti sovražiš?« (Copeland, 1993/2007: 42) Ko program naleti na takšen stavek, samostojno izbere transformacijo z vrha spiska transformacij, povezanih z vzorcem, npr.: »Ti misliš, da te*...« ali ne. Z uporabo te transformacije, bi se stavek lahko glasil: »Ti misliš, da te sovražim ali ne?«

Ko je enkrat takšna transformacija uporabljena, se ta pomakne na konec spiska, s čimer se zagotovi, da se ne ponovi prepogosto (po navadi obstaja štiri do pet podobnih transformacij na vsakem spisku).

V programu Eliza pa je uporabljen tudi programerski trik, ki omogoča, da ustvarja poseben odziv, predvsem v situacijah, ko ni nobenega ujemanja vzorcev. Ko se pojavi vhodni stavek, v katerem je prevladujoči vzorec »moj«, Eliza, ne samo da uporabi za odziv transformacijo z vrha spiska ustreznih transformacij, temveč tudi izbere drugo transformacijo iz drugega spiska in rezultate shrani za uporabo v situacijah brez ujemanja. Tako za stavek z »moj«, lahko program shrani za nadaljnjo uporabo stavek: »Ali ima to kakršnokoli zvezo z dejstvom, da tvoj*...«.

Na opisani način se lahko posameznik pogovarja z Elizo in ta lahko spodbuja pri njem najgloblja čustva. Program nima ne sistema za učenje in si nič ne zapomni, vendar pa je kljub temu znana sociologinja Sherry Turkle zapisala, da ljudje sčasoma pričnejo ceniti zvezo z Elizo predvsem zato, ker vse, kar rečejo, le-ta spremeni v zelo osebno vprašanje. Vse to lahko privede do zelo zanimive konverzacije, ki tako Elizo dela prijazno in razumevajočo. (Turkle, 1995)

Seveda pa Eliza izkazuje precejšnje pomanjkanje razumevanja in problemi nastopijo pozneje, ko prevečkrat ponovi nekoliko modificirano obliko vhodnega stavka. V bistvu je Eliza v veliki meri odvisna predvsem od sreče, saj je celoten sistem odvisen od verjetnosti pravilne ocene vzorca in odgovarjajoče transformacije.

6.2.2 Kako Eliza programsko deluje

Kot je bilo že omenjeno, je Eliza čisti fizikalni simbolni sistem iz zgodnje faze razvoja UI. Zavedati se moramo, da je leta 1966, ko je Eliza nastala, bilo *interaktivno računalništvo*⁶² (takrat le z uporabo tipkovnice) popolnoma novo. Bilo je petnajst let pred tem, ko so postali osebni računalniki splošno znani in uporabljeni in trideset let pred tem, ko se je večina ljudi srečala s procesiranjem naravnega jezika, npr. pri orodjih za pomoč (angl. help) pri osebnih računalnikih.

Celoten programski paket Eliza je sestavljen iz dveh programov, tj. dveh delov ali dveh datotek, in to:

⁶² Interaktivno računalništvo omogoča komunikacijo med strojem (računalnikom) in človekom. V zgodnejši fazi se ja ta komunikacija vršila le preko tipkovnice, danes pa se za to uporabljajo tudi drugi elementi (npr. miška) in drugi kanali, npr. govor.

- programa ELIZA_list, ki predstavlja hevristiko programskega paketa,⁶³ sestavljeno iz različnih ukazov za manipuliranje s simboli in simbolnimi strukturami, in
- programa, datoteke skriptov;⁶⁴ BBSCHAT_list, ki predstavlja podatkovno bazo vseh možnih simbolov in simbolnih struktur (osebnih zaimkov, vprašanj, delov stavkov in stavkov, kar bi lahko primerjali z deklarativnim modulom v ACT-R). Stavki so lahko zaključeni, ali pa so le nastavki stavka, npr. »Ali ti*...«, ki se s pomočjo hevristike dopolni z vzorcem iz vhodnega stavka.

PROGRAM ELIZA

V nadaljevanju si oglejmo le nekaj osnovnih delov programa z ustreznimi komentarji.

Izvorna koda	Komentar
<pre> program eliza; const bbs_chat_file = 'bbschat.net'; {script file name} type string_12 = string[12]; { global types } string_25 = string[25]; string_80 = string[80]; string_255 = string[255]; bull_line_ptr = ^string_80; charset = set of char; const alphabet : charset = ['A'..'Z']; {global variables used in the bbs} locase : array['A'..'Z'] of char = ('a','b','c','d','e','f','g','h','i','j','k', 'l','m','n','o','p','q','r','s','t','u','v','w','x','y','z'); var case_switch : boolean; text_file : text[4096]; </pre>	Definiramo ime datoteke, kjer so skripti – vzpostavimo zvezo med dvema programoma. Definiramo splošne spremenljivke v programu skriptov.

⁶³ Kar bi lahko primerjali s proceduralnim modulom v ACT-R.

⁶⁴ Skript je v računalništvu manjši program napisan tako, da ga lahko uporablja drugi program, ki na podlagi določene hevristike izvaja predpisane aktivnosti. V UI je skript tudi struktura za predstavljanje procesnega znanja.

<pre> procedure check_time_limit; begin end; procedure clear_buffer; begin end; procedure chat_with_eliza; const max_trans = 24; {number of words to transpose} max_reply = 20; {the maximum number of replies for each set of responses} max_key = 100; {maximum number of key words} type keyrec = record word : string_25; resp_set : integer; end; reprec = record tot_resp : integer; last_resp : integer; reply : array[0..max_reply] of bull_line_ptr; end; key_word_array = array[1..max_key] of keyrec; resp_set_array = array[1..max_key] of reprec; transpose_array = array[1..max_trans] of string_12; var tmp_str : string_255; end_of_chat : boolean; this_keyword : string_255; user_input : string_255; current_response : integer; i : integer; key_no : integer; key : key_word_array; key_found : boolean; keywords : integer; </pre>	Vpeljemo funkcijo naključne izbire (<i>randomize</i>) in zaženemo program: <i>procedure chat_with_eliza</i>. Vzpostavimo začetno stanje (izbriši vse predhodne aktivnosti). Pričetek programa Definiramo število besed za menjavo=24. Definiramo maksimalno število odgovorov za vsak primer odzivov = 20. Definiramo max. število ključnih besed=100. Oblika izpisa stavka s ključno besedo Oblika izpisa odziva Ključna beseda Odziv Pretvorba Spremenljivke (njihove oblike) v programu
--	--

<pre> l : integer; my_resp : reprec; my_str : string_255; no_key : reprec; null_input : boolean; null_resp : reprec; prog_output : string_255; puncs : set of char; resp_no : integer; response : string_255; responses : integer; response_set : resp_set_array; save_key_no : integer; signon : reprec; transpose : integer; wordin : transpose_array; wordout : transpose_array; x : byte; function right(rstr : string_255; len : integer) : string_255; begin right := copy(rstr,(length(rstr) - len + 1),len); end; procedure write_line(tstr : string_255); var i : integer; procedure write_char(char1 : char); begin delay(60 + random(75)); if char1 in alphabet then char1 := locase[char1]; write(char1); end; begin for i := 1 to length(tstr) do begin if random(50) = 0 then begin write_char(chr(random(26)+65)); delay(100 + random(100)); write_char(#8); end; write_char(tstr[i]); end; end; </pre>	Postopek zapisa prve vrstice Postopek zapisovanja posameznih simbolov Postopek branja posamičnega simbola
---	--

[illegible]

<pre> end_of_chat := true else begin keywords := 0; transpose := 0; current_response := 1; my_resp.tot_resp := 0; null_resp.tot_resp := 0; no_key.tot_resp := 0; signon.tot_resp := 0; responses := 0; line_buffered := false; while not (eof(text_file) or end_of_chat) do begin if not line_buffered then begin read_script_line; rec_type := tmp_str[1]; delete(tmp_str,1,1); end; line_buffered := false; case rec_type of 'S' : add_response(signon); 'T' : add_transpose; 'N' : add_response(null_resp); 'M' : add_response(my_resp); 'X' : add_response(no_key); 'K' : add_keyword; 'R' : add_response_set; else end_of_chat := true; end; end; close(text_file); end; if end_of_chat then begin writeln('Script file error. Line ',line_no); writeln; end; end; procedure get_response; begin check_time_limit; clear_buffer; readln(user_input); tmp_str := ""; i := 0; </pre>	Postopek ustvarjanja odziva
---	--

[illegible]

<pre> ''; end; procedure conjugate_and_transpose; begin l := 1; while l <= length(prog_output) do begin for i := 1 to transpose do if (copy(prog_output,l,length(wordin[i])) = wordin[i]) then begin prog_output := copy(prog_output,1,l-1) + wordout[i] + right(prog_output,(length(prog_output) - l - length(wordin[i]) + 1)); l := l + length(wordout[i]) - 2; end; l := l + 1; end; while prog_output[1] = ' ' do begin prog_output[1] := #0; delete(prog_output,1,1); end; x := length(prog_output); while prog_output[x] = ' ' do x := x-1; prog_output[0] := chr(x); if this_keyword = 'MY' then my_str := prog_output; end; procedure show_response(var rset : reprec; var prog_out : string_255); begin with rset do begin resp_no := random(tot_resp); if tot_resp > 1 then while resp_no = last_resp do resp_no := random(tot_resp); response := reply[resp_no]^; last_resp := resp_no; i := pos('*',response); if i = 0 then write_line(response) else begin </pre>	Postopek prikazovanja odziva (izpis na monitorju)
---	--

<pre> if i <> length(response)-1 then tmp_str := copy(response,1,(i-1)) + '' + prog_out + '' + copy(response,(i+1),length(response)-i) else tmp_str := copy(response,1,(i-1)) + '' + prog_out + response[i+1]; write_line(tmp_str); end; end; writeln; end; procedure show_reply; begin writeln; if (not key_found) then begin if (my_str <> '') and (random(5) = 0) then show_response(my_resp,my_str) else begin tmp_str := ''; if null_input then show_response(null_resp,tmp_str) else show_response(no_key,tmp_str); end; end else begin current_response := key[key_no].resp_set; show_response(response_set[current_response],prog_output); end; end; begin my_resp.last_resp := -1; null_resp.last_resp := -1; no_key.last_resp := -1; for i := 1 to max_key do response_set[i].last_resp := -1; puncs := ['.', ',', ';', '!', '?']; my_str := ''; end_of_chat := false; case_switch := false; writeln('Enter QUIT to quit chatting'); writeln; </pre>	Ponoven začetek – naslednje vprašanje/odziv, vpeljemo naključno funkcijo in postopek se ponovi.
--	--

```

read_script;
if not end_of_chat then
begin
  i := random(signon.tot_resp);
  write_line(signon.reply[i]^);
  writeln;
  get_response;
  while not end_of_chat do
  begin
    find_keyword;
    if key_found then
      conjugate_and_transpose;
      show_reply;
      get_response;
    end;
  end;
case_switch := true;
with signon do
  for i := 1 to tot_resp do
    dispose(reply[i-1]);
  with null_resp do
    for i := 1 to tot_resp do
      dispose(reply[i-1]);
    with no_key do
      for i := 1 to tot_resp do
        dispose(reply[i-1]);
      with my_resp do
        for i := 1 to tot_resp do
          dispose(reply[i-1]);
        for l := 1 to responses do
          with response_set[l] do
            for i := 1 to tot_resp do
              dispose(reply[i-1]);
            end;
          end;
        end;
      end;
    end;
  end;
begin
  randomize;
  chat_with_eliza;
end.

```

Iz predstavljenega programa je razvidno, da sta osnovna stavka v programu, če že, potem stavek (angl. if – then stavek) in »do« zanka, ki omogoča ponavljanje iste aktivnosti toliko časa, da ne dosežemo želenega učinka, npr. ne najdemo ustrezne ključne besede.

ELIZA – DATOTEKA SKRIPTOV

V drugem programu BBSCHAT_list so vsi skripti, vse besede, delni stavki ali celotni stavki, s katerimi osnovni program izvaja manipulacijo simbolnih zapisov. Zgradba datoteke skriptov in princip delovanja sta prikazana v nadaljevanju.

Izvorna koda	Komentar
<pre> ; ELIZA SCRIPT FILE ; ; FORMAT: ; Sxxxxxxxx - DEFINE SIGNON MESSAGE ; Txxxxx ; Txxxxx - DEFINE PAIR OF WORDS TO TRANPOSE ; Nxxxxx - RESPONSE FOR NULL ENTRY ; Mxxxxx - DEFINE RESPONSE FOR LATER USE OF "MY" ; Xxxxxxx - DEFINE RESPONSE FOR NO KEYWORD FOUND ; Kxxxxxx - DEFINE KEYWORD ; Rxxxxxx - DEFINE RESPONSE FOR PREVIOUS KEYWORD ; ; NOTE: FOR "T" AND "K", THE LAST CHARACTER IS DISCARDED TO MAKE IT EASY TO ENTER TRAILING SPACES. ; ; FOR "R" AND "M" THE STRING FOLLOWING THE KEYWORD IN THE USER'S INPUT IS INSERTED AT THE * ; ; IF A KEYWORD HAS NO RESPONSES, THEN IT WILL USE THE RESPONSES OF THE NEXT KEYWORD WITH RESPONSES ; ; SIGNON MESSAGE </pre>	Definiranje formata S – določi sporočilo za vključitev v program (v primeru Elize sta to dva: Hi, I am Eliza. What do you want to talk about? ali: So how are you doing today?) Program z naključnim generatorjem izbira med njima. T – določi par besed, ki ju je potrebno zamenjati. N – odziv programa, če ni vhodnega odziva. M – določi odziv za kasnejšo uporabo »moj«. X – odziv, če se ne da določiti ključne besede. K – določi ključno besedo. R – določi odziv za predhodno ključno besedo.

<pre> ; SHI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT? SSO HOW ARE YOU DOING TODAY? ; ; TRANSPOSITION TABLE ; T ARE " T AM " T WERE " T YOU " T I " T YOUR " T MY " Ltd. T DAD " T FATHER " T MOM " T MOTHER " T DREAMS " T DREAM " ; ; NULL RESPONSES ; Nhow am I supposed to talk if you won't say anything? NWHAT? NCOME AGAIN? ; ; RESPONSES FOR "MY" ; TO USE IF NO KEYWORD WAS FOUND. ; MEarlier you mentioned that your*. Ltd. ; ; NO KEYWORD FOUND ; XPLEASE GO ON. XI'M NOT SURE I KNOW WHAT YOU'RE TALKING ABOUT. </pre>	Začetni stavek S – Hej, jaz sem Eliza. O čem bi se želel pogovarjati? S – Kako si danes? Tabela za transponiranje T – smo » T – sem » T – bo » T – ti » T – jaz » T – tvoj » T – moj » Ltd. T – ata » T – oče » T – mama » T – mati » T – sanje » Brez odziva (izbira se naključno) N – Kako si predstavljaš, da se bova pogovarjala, če ne rečeš nič? N – Kaj? N – Daj no? Odziv na »Moj«, ali če ni bilo mogoče najti nobene ključne besede. M – Pred tem si omenil, da tvoj* (stavek se nadaljuje z delom vhodnega stavka). Ltd. Ni bilo moč najti nobene ključne besede.
--	---

XAND???? Ltd. ; DEFINE KEYWORDS AND RESPONSES ; KCAN I " RPERHAPS YOU DON'T WANT TO*. itd; ; KI FEEL " RHOW LONG HAVE YOU FELT*? Ltd.; ; KMY MOTHER" KMY FATHER" KMY SISTER" KMY BROTHER" Ltd. RTELL ME MORE ABOUT YOUR FAMILY. ; Ltd. Ltd.; KHELLO" KHI " RHI! WHAT DO YOU WANT TO TALK ABOUT?	X – Prosim nadaljuj! X – Nisem prepričana, da vem, kaj mi želiš povedati. X – In??? Ltd. Določi ključne besede in odzive. K – Ali lahko« R – Morda pa ne želiš, da* Ltd. K – Čutim« R – Kako dolgo že čutiš*? Ltd. K – Moja mati« K – Moj oče« K – Moja sestra« K – Moj brat« Ltd. R – Povej mi kaj več o tvoji družini. K – Halo« K – Hej« R – O čem bi se želel pogovarjati?
--	---

V to datoteko lahko vnesemo poljubno število različnih stavkov, delnih stavkov, situacij ipd. Zanje je značilno, da so vnaprej premišljeni in da program izbira vprašanja (na osnovi pravil) in odgovore samo izmed že obstoječih možnosti. Podatkovne baze (baze simbolov in simbolnih struktur) se lahko dopolnjujejo in širijo le, če jih dopolni človek.

EFEKT – ELIZA

Efekt imenovan Eliza se v računalništvu uporablja kot težnja po nezavednem priznavanju, da je vedenje računalnika lahko podobno vedenju človeka. Mnogi so verjeli, da je Eliza resnična oseba, in dogodilo se je celo, da so ta program pričeli

uporabljati v psihiatričnih klinikah, kot pomoč pri zdravljenju. Najbolj trivialen primer, ki ga je navedel Douglas Hoffstaedter, je primer avtomatskih strojev (bankomatov, avtomatov za kavo ipd.), ki na koncu na zaslon zapišejo »hvala«. Najbolj nepoučen opazovalec, bi lahko mislil, da stroj izraža s tem hvaležnost, stroj pa je le napisal sprogramiran niz simbolov. (Hofstadter, 1996)

Po vseh teh izkušnjah je Weizenbaum ugotovil, da vprašanje pri vsem skupaj ni več, ali je možno narediti stroj, ki misli, temveč ali ga je sploh dovoljeno narediti. UI, tako je bil prepričan, že po svoji naravi ne bo sposobna popolnega razumevanja človekovih zahtev.

6.3 Od simbolnih sistemov do jezika misli

Fodorjeva hipoteza o jeziku misli govori o tem, da je osnovna struktura misli simbolna in prenaša informacije kot stavke v nekem notranjem jeziku misli, včasih imenovanem tudi *mentalese*. Informacijsko procesiranje transformira te stavke v jezik misli. Ta ideja je izhodiščna ideja o tem, kako duh *prejema* informacije o svoji okolici. Nekatero od teh informacij prenaša svetloba (fotonski fluks) in jih sprejema očesna retina, ali zvok (zvočno valovanje), ki ga zaznamo z ušesnim bobničem ipd. Vendar pa na splošno naše obnašanje ni v celoti *določeno* z informacijami, ki jih prejmemo, saj lahko različni ljudje sprejemajo isto informacijo različno, oz. celo ista oseba lahko sprejema isto informacijo različno ob različnih trenutkih. Ne obstajajo standardni odzivi na vzorce zvočnega valovanja, npr. klica v angleščini »Help!«. Kako procesira to informacijo Slovenec? Če se posameznik odzove na ta klic, potem se zato, ker je njegov duh na nek način dekodiral besedo v angleščini, ugotovil kaj ta želi povedati, in se odločil, kako se nanjo odzvati. To je kompleksno procesiranje začetne informacije, ki smo jo sprejeli z našim ušesnim bobničem. (Fodor, 1983)

Vendar, kako se ta informacijski proces v resnici izvaja? Kako vibracije, ki jih sprejme ušesni bobnič, privedejo do krčenja mišic, kar lahko povzroči, da bo tisti, ki je izustil ta *Help*, rešen. Informacije se morajo na nek način prenašati. Vemo, kako se prenašajo slušne informacije, kako deluje naš slušni sistem, kaj se dogodi potem s temi informacijami pa ni več tako dobro znano. Predpostavlja se, da obstaja fizikalen simbolni sistem, fizikalna struktura, ki prenaša informacije, in tako služi kot *predstava* neposredne okolice. To je naslednja osnovna predpostavka kognitivne znanosti. Skratka, informacijski proces je na svojem začetku stvar transformiranja predstav na

način, da ta končno privede do aktivnosti v živčnem sistemu, ki »dajo navodila« mišicam, da reagirajo.

Informacijski proces vsebuje mnogo različnih predstav. Celoten proces se prične s predstavo, ki nosi informacijo o vibracijah na ušesni bobnič. Nekako se ta predstava pretvori v neprimerno kompleksnejšo predstavo, ki bi jo lahko poimenovali *prepričanje*, da je nekdo v nevarnosti. To prepričanje pa je pomemben »motor« vedenja (da skočimo in pričnemo reševati). Vendar pa vse to še ni dovolj. Potrebne so še interakcije z drugimi prepričanji (kot npr. da bom sploh lahko nekoga pravočasno rešili) zato, da ustvarimo to, kar bi si lahko mislil o *namenu* reagiranja na določen način. Ta namen nato ustvari predstave, ki so povezane z motoričnimi navodili, zadolženimi za kontrolo gibov telesa.

Od vseh različnih tipov predstav, so Fodorja najbolj zanimale tiste, ki se nanašajo na prepričanja, želje in druga podobna psihična stanja. Ta psihična stanja mnogokrat imenujemo *propozicionalno stanje* (angl. *propositional attitudes*). Kadar mislimo, verjamemo ali si želimo, so naše misli, prepričanja in želje vedno o nečem (npr. želim, da bi šli na počitnice), tudi o nečem, kar ne obstaja (npr. o okroglem kvadratu), in jim lahko verjamemo ali pa ne, skratka, do njih lahko zavzamemo različna stališča. Težko pa verjamemo, da bi to veljalo tudi za zgolj fizične stvari.

Eden od Fodorejih argumentov za hipotezo jezika misli je bil tudi vpliv propozicionalnih stanj na razumevanje našega vedenja. Naslanjal se je na *psihologijo psihologije prepričanje – želja* in trdil, da je ta psihologija uspešna, zato ker je resnična. Ta pogled se velikokrat imenuje *intencionalni realizem* (angl. *intentional realism*) ali *realizem o propozicijskih stanjih* (angl. *realism about propositional attitudes*). Ključni Fodorjev argument za hipotezo jezika misli je, da je ta hipoteza edina razlaga, kako lahko deluje psihologija prepričanje – želja. (Bermudez, 2010)

6.3.1 Intencionalni realizem in vzročnost na podlagi vsebine

Intencionalni realizem obravnava prepričanja in želje kot neke vrste stvari, ki lahko povzročajo vedenje. Vendar pa je to posebna vrsta vzročnosti. Obstaja bistvena razlika med tem, kar želimo narediti, in tem, kaj naredimo zaradi zunanjih vplivov. Svojo nogo na primer premaknemo zato, ker želimo preteči maraton, ali pa zato, ker nas je zdravnik s kladivcem udaril po njej. V prvem primeru je vzrok giba želja, da

naredimo to, kar smo si zaželeli (preteči maraton). To filozofi imenujejo *vsebina* želje. Če nas pa zdravnik udari po nogi s kladivcem, reakcija nima prav nobene zveze z našo željo.

Prepričanja in želje povzročajo vedenje s tem, kako učinkujejo predstave o svetu na njihovo vsebino. Vsaka zadovoljitev interne realizacije mora razložiti, kako je bilo možno zadovoljiti ta *vzrok iz vsebine* (*angl. causation by the content*). Premik noge je racionalna stvar, če se odločimo preteči maraton in če verjamemo, da tečemo v pravo smer.

Še vedno pa je vzrok iz vsebine skrajno skrivnosten. V določenem smislu lahko predstavlja popolnoma enostaven objekt, npr. vzorec zvočnega vala, populacijo nevronov ali kos papirja. Vendar pa so predstave, za katere se zanimamo (intencionalna stanja), tudi stvari, ki v sebi nosijo posebne *semantične* povezave s svetom, skratka imajo svoj pomen. In uganka ni samo, kako imajo predstave vzročne moči znotraj sveta samega, temveč tudi, kako imajo lahko predstave vzročne moči znotraj sveta zaradi svojih vsebinskih lastnosti.

Velika prednost hipoteze jezika misli po Fodorju je, da rešuje zgornjo uganko vzroka iz vsebine. Fodor, pa tudi večina kognitivnih znanstvenikov in veliko število filozofov, meni, da so manipulacije predstav, ki jih izvajajo možgani, v celoti fizikalne in mehanske. Možgani in vse predstave, ki se nahajajo v njih, so fizikalne entitete, kar pomeni, da so lahko dovzetne le na določene tipe fizikalnih lastnosti v mentalnih predstavah. Izgovorjava besede 'mačka' ni nič drugega kot določen vzorec zvočnih valov. To zvočno valovanje ima določene fizikalne lastnosti, ki lahko imajo določene vplive na možgane. Vendar pa je dejstvo, da ti zvočni valovi predstavljajo mačko za ljudi, ki govore slovensko, popolnoma drug tip lastnosti. (Bermudez, 2010)

Imenujmo fizikalne lastnosti, s katerimi lahko manipuliramo v možganih *formalne lastnosti*, predvsem zato, ker imajo fizikalno obliko (formo). In imenujmo lastnosti, ki predstavljajo predstave, *semantične*⁶⁵ lastnosti, tako kot je semantika predstavljena v jezikoslovju. To nam omogoča pogled na problem z drugega zornega kota. Zastavimo si vprašanje: Kako so lahko možgani stroj za procesiranje informacij, če so slepi za semantične lastnosti prezentacij? In kako so lahko možgani stroj za

⁶⁵ Semantika je pomenoslovje, nauk o pomenu besed.

procesiranje informacij, če so vse, kar lahko procesirajo, le formalne lastnosti predstav?

S tega vidika je Fodorjevo delo s področja hipoteze fizikalnih simbolnih sistemov še posebej pomembno. Računalniki manipulirajo z nizi simbolov, digitalni računalniki, ki so programirani v binarni kodi, manipulirajo z nizi enk in ničel in ti nizi lahko predstavljajo karkoli. Predstavljajo lahko naravno število 1 in 0, če jih zapišemo binarno, pa lahko v obliki 10 predstavljajo število 2, v obliki 11 število 3, ali pa lahko predstavljajo nekaj popolnoma drugega, npr. serijo točk, ki predstavljajo vključiti ali izključiti. Tako lahko trdimo, da računalniki izvajajo operacije bolj s *števili* kot s *številkami*. Pri tem so števila simboli s posebno obliko in številke so to, kar ta števila predstavljajo. Posebej pomembno pa je, da so računalniki programirani tako, da manipulirajo nize enk in ničel na poseben način, ki na koncu privede do pravega rezultata nanašajoč se na interpretacijo, čeprav smatramo, da je računalnik slep za interpretacijo.

Torej, računalniki manipulirajo simbole na način, ki je občutljiv samo na njihove formalne lastnosti, a pri tem upoštevajo njihove semantične lastnosti. In kot trdi Fodor, je to natanko to, kar delajo tudi možgani. Možgani so fizikalni sistem, ki je občutljiv samo na formalne lastnosti mentalnih predstav. Vendar pa morajo stroji za procesiranje informacij (kot npr. računalniki) upoštevati semantične lastnosti mentalnih predstav. Fodorjev argument iz intencionalnega realizma in jezika misli lahko razumemo na naslednji način: predpostavimo, da morajo možgani in računalnik rešiti enak problem. Potem je popolnoma jasno, kako ga reši računalnik. Če pa želimo razumeti, kako ga rešijo možgani, to najlažje dosežemo tako, da pač možgane smatramo kot neke vrste računalnik. (Bermudez, 2010)

6.4 Računalniški model duha (zveza med sintakso in semantiko)

SINTAKSA IN JEZIK MISLI

Fodorjev način računalniškega modeliranja duha bi lahko strnili v tri zahteve (Fodor, 1983):

1. vzročnost na osnovi vsebine je v zadnji fazi stvar vzročnih interakcij med fizikalnimi stanji;

2. ta fizikalna stanja so zgrajena kot stavki in njihova stavkom podobna zgradba določa, iz česa so sestavljena in kako vplivajo ena na druge;
3. vzročni prehodi med stavki v jeziku misli upoštevajo racionalne relacije med njihovimi vsebinami.

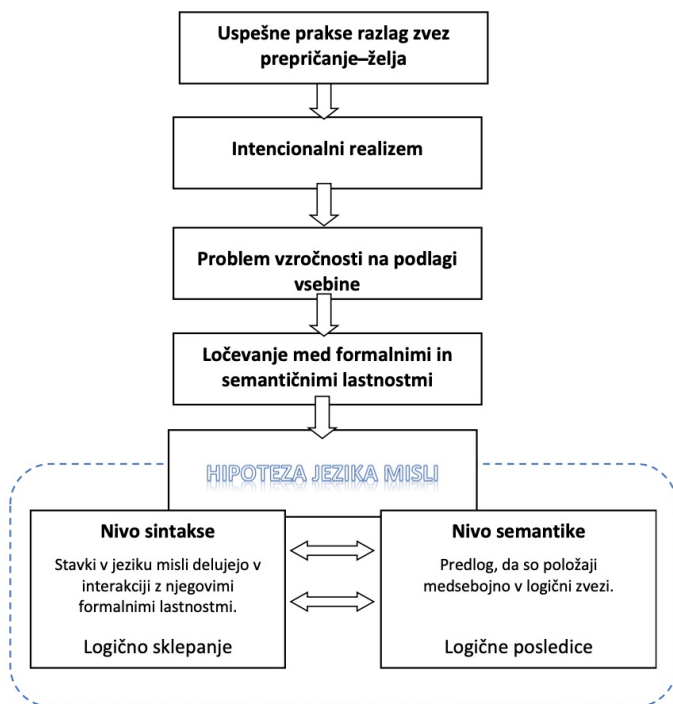
Druga in tretja zahteva predstavljata Fodorjev pomemben prispevek k reševanju problema vzročnosti iz vsebine. Iz druge izhaja njegov pogled na to, da je medij zavesti to, kar imenuje jezik misli. Po Fodorju človek razmišlja v stavkih, vendar to niso stavki naravnega jezika, kot sta angleščina ali slovenščina. Ta jezik misli je mnogo bolj podoben jeziku logike, kot je npr. *propozicijska logika*.⁶⁶

Analogija med jezikom misli in jezikom logike je bistvo Fodorjeve rešitve problema vzroka iz vsebine, kar je izpostavljeno v zahtevi 3. Fodorjev osnovni predlog je, da razumemo zveze med stavki v jeziku misli in njihovo vsebino (ali njihovim pomenom), podobno, kot to velja za zvezo med semantiko in sintakso v formalnih sistemih. Stavke v jeziku misli bi lahko obravnavali le kot čisto sintaktične, sestavljene iz osnovnih simbolov, medsebojno povezane na osnovi določenih pravil. Ali pa bi jih obravnavali s stališča semantike glede na to, kako ti predstavljajo svet (v tem primeru je to povezano z intencionalnimi stanji). Tako lahko na jezik misli gledamo s stališča sintakse ali s stališča semantike, torej s stališča formalnih zvez med fizikalnimi simbolnimi strukturami, ali pa na osnovi semantičnih zvez, ki veljajo med stanji, ki predstavljajo svet.

Vrnimo se k Fodorjevi zahtevi 3. Predpostavimo, da je vzročni prehod med stavkom in jezikom misli v bistvu sintaksa, ki upošteva čiste formalne vrednosti simbolov, ne glede na to, kaj ti simboli pomenijo. Če to drži, potem si moramo zastaviti naslednje vprašanje: Zakaj sintaktične povezave med stavki in jezikom misli, prevedenim v semantične zveze, veljajo tudi med vsebinami teh stavkov?

⁶⁶ Propozicijska logika je logika, ki opisuje in normira sklepalna razmerja med stavki (propozicijami). V propozicijski logiki je smisel stavkov abstrahiran. Propozicija (za zapis posamičnih stavkov se uporabljajo črke p, q, r ...) lahko v klasičnih, dvovalentnih logičkih sistemih zasede dve resničnostni vrednosti: resnično in neresnično. Stavki, s katerimi operira propozicijska logika, torej niso stavki vsakdanje, naravne govorice, temveč idealne entitete, ki so pomembne le iz vidika svoje resničnosti ali neresničnosti. Tako bo stavek "trava je rdeča" v logiki smatran zgolj kot nosilec resničnostne vrednosti 0, tj. neresničnega. To načelo je včasih poimenovano tudi načelo ekstenzionalnosti.

Če resno upoštevamo idejo, da je jezik misli formalni sistem, potem je na to mogoče odgovoriti enostavno. Sintaktični prehod med stavki in jezikom misli sledi semantičnemu prehodu med vsebinami teh stavkov zaradi natančno istega razloga, kot sintaksa sledi semantiki v kateremkoli pravilno konstruiranem formalnem sistemu. Poglejmo, kako to deluje na primeru. Predpostavimo, da imamo dva kompleksna simbola. Vsak od njih je stavek v jeziku misli. Vsak ima določeno sintaktično obliko. Recimo, da sta to stavka *Ga* in *Fa*. Te sintaktične oblike imajo pomen. Posamezni pomeni, ki jih imajo, so funkcija njihove oblike. Vemo, da sta »F« in »G« simbola za povedke. Naj »F« pomeni »je visoka« in »G« pomeni »ima rdeče lase«. Prav tako vemo, da je »a« ime simbola. Naj bo »a« žensko ime Georgina. Tako pomeni »Ga« – Georgina je visoka, ali pa »Fa« – Georgina ima rdeče lase. Na osnovi tega vidimo, kako lahko analiziramo zelo majhen delček misli z uporabo hipoteze jezika misli. Ker je potek sklepanja, ki nas privede do hipoteze jezika misli, dokaj zapleten, so na spodnji sliki prikazani posamezni koraki na tej poti. (Bermudez, 2010)



Slika 6.1: Simbolna struktura Fodorjeve hipoteze jezika misli.

V spodnji Preglednici 1 vidimo, kako lahko dva simbola » Fa in Ga « pretvorimo v dveh korakih v nek bolj kompleksni fizikalni simbol » $\exists x (Fx \wedge Gx)$ «. Pravila, na osnovi katerih lahko to naredimo, so popolnoma sintaktična, saj so le pravila za manipuliranje s simbolnimi strukturami. Če pa podrobneje pogledamo na zveze med pomeni, na eni strani med » Fa in Ga « in pomenom » $\exists x (Fx \wedge Gx)$ « na drugi strani, lahko ugotovimo, da ta čista sintaktična transformacija ohranja logične zveze med predlogi, ki jih zagovarjajo simboli. Če je res, da je Georgina visoka in da ima Georgina rdeče lase, potem je tudi zagotovo res, da je vsaj ena oseba visoka in ima rdeče lase.

Če sedaj vse povežemo, potem lahko prepričanja in želje smatramo kot fizikalne strukture (tj. stavke v jeziku misli) in tako lahko vzročnost in druge oblike razmišljanja razumemo v obliki vzročnih zvez med strukturami. Te vzročne zveze so uporabne le za formalne, sintaktične lastnosti fizikalnih struktur. Zato, ker je jezik misli formalni jezik z analogijami formalnih lastnosti trdnosti in celovitosti, te čiste sintaktične pretvorbe upoštevajo semantične povezave med vsebino in relevantnimi prepričanji in željami. Na tak način lahko na osnovi zdravorazumske psihologije zagovarjamo trditev duha kot računalniškega procesiranja stavkov v jeziku misli.

Tabela 6.1: Simbolne pretvorbe.

Simbol	Pravilo za pretvorbo	Pomen
1. Ga		1. Georgina je visoka.
2. Fa		2. Georgina ima rdeče lase.
3. $(Fa \wedge Ga)$	Če se kompleksna simbola » S « in » T « pojavita v predhodni vrstici, potem zapiši » $(S \& T)$ «.	3. Georgina je visoka in ima rdeče lase.
4. $\exists x (Fx \wedge Gx)$	Če so v predhodnih vrsticah kompleksni simboli z imeni simbola, potem zamenjaj ime simbola z » x « in zapiši » $\exists x$ « pred kompleksnim simbolom.	4. Vsaj ena oseba je visoka in ima rdeče lase.

6.5 Primeri simbolnih arhitektur

Obstaja vrsta arhitektur različnih skupin avtorjev, namenjenih za kognitivno modeliranje, ki skuša na različne načine razlagati človeški duh. Če omenimo le nekatere (Duch, Oentaryo, Pasquier, 2008):

- *SOAR* (angl. *State, Operator And Results*) arhitektura je zelo ozko vezana na hipotezo fizikalnega simbolnega sistema. Za reševanje problemov uporablja hevristično iskalne približke simbolnih sistemov in ne vsebuje nevronske mreže. Ker je bila to ena od prvih kognitivnih arhitektur, si jo bomo nekoliko bolj podrobno pogledali v nadaljevanju.
- *EPIC* (angl. *Executive-Process/Interactive Control*) je kognitivna arhitektura, ki simulira različne dele človekovega informacijsko procesnega sistema, kot npr. zaznavanje, kognicijo in motorično procesiranje. Še posebej pa je uporabna za ustvarjanje kognitivnih modelov s področja ustvarjanja povezav človek – računalnik.
- *CHREST* (angl. *Chunk Hierarchy and REtrieval STRucture*) je simbolna kognitivna arhitektura, zasnovana na principu omejene pozornosti, omejenega kratkoročnega spomina, in *chunkov* (gradnikov).⁶⁷ Učenje, ki je bistveni del te arhitekture, je modelirano na osnovi mrež vozlišč (*chunkov*), ki so lahko povezani na različne načine. *CHREST* se pogosto uporablja za modeliranje učenja ob uporabi velikih baz podatkov področja, kot npr. baze podatkov šahovskih iger za posredovanje šahovskih strokovnih znanj ali pa spodbujanje otrokovega govornega razvoja z uporabo baz podatkov o poenostavljenem jeziku, ki ga starši uporabljajo za pogovarjanje z dojenčki. Arhitektura vsebuje veliko sposobnostnih parametrov (npr. sposobnost vizualnega kratkoročnega spomina) in časovnih parametrov (čas učenja *chunka* ali čas potreben, da preide informacija v kratkotrajen spomin). To pa omogoča, da natančno in kvantitativno napovedujemo človekovo obnašanje.
- *CLARION* (angl. *Connectionist Learning with Adaptive Rule Induction ON-line*) je hibridna kognitivna arhitektura, ki razlikuje med implicitnimi in eksplicitnimi procesi in temelji na združevanju povezav med temi procesi. Uporablja se za simuliranje nekaterih nalog v kognitivni in socialni psihologiji. Prav tako pa se uporablja za implementiranje inteligentnih sistemov in aplikacij na osnovi umetne inteligence.
- *DUAL* je splošna kognitivna arhitektura, ki povezuje konekcionistične in simbolne modele na mikro nivoju. Temelji na decentraliziranih predstavah in emergentnem računanju. Spodbujen je bil z arhitekturo »Society of Mind«, ki jo

⁶⁷ *Chunk* - gradnik je fragment informacije, ki se uporablja v različnih multimedijskih formatih. V distribuiranem računalništvu pa je gradnik niz podatkov, ki se pošiljajo procesorju, da lahko ta izvede zahtevano informacijo (npr. ko pritisnemo v programu Word ikono poudarjeno, se izvede v gradniku zapisana operacija). Mnogokrat lahko uporabljamo tudi izraz gradnik.

je predlagal Marvin Minsky, vendar pa se je od izvirnega predloga, izvirne ideje, na mnogih področjih precej oddaljil. Računanje se pri tej arhitekturi izvaja s pomočjo interakcije številnih mikro agentov in vsak od njih je hibridna simbolno konekcionistična naprava. Ti inteligentni agenti si izmenjujejo informacije in delujejo na osnovi povezav, s pomočjo katerih se lahko učijo in se na podlagi tega modificirajo, ustvarjajo lahko koalicije, ki predstavljajo skupne koncepte in dejstva. Na osnovi te izhodiščne arhitekture se je razvila vrsta modelov: *AMBRE* – model za ustvarjanje analogij in pomnjenje, *JUDGEMAP* – model za presojanje, *PEAN* – model za percepcijo itd.

- *ACT-R* (angl. *Adaptive Components of Thought-Rational*), je hibridna arhitektura, ki je danes ena od najbolj delujočih in najbolj kompleksnih kognitivnih struktur. ACT-R se bomo podrobneje posvetili v nadaljevanju.

Če so SOAR, EPIC in CHREST simbolne arhitekture, pa je trend v zadnjih letih vse bolj usmerjen v povezave med simbolnimi in konekcionističnimi sistemi, skratka k hibridnim arhitekturam. V to skupino spadajo zgoraj omenjene arhitekture CLARION, DUAL in ACT-R. Pri svojem delu se ne bomo posvečali primerjavam med različnimi arhitekturami, saj je to v strokovni javnosti že bilo narejeno, temveč se bomo osredotočili le na to, kaj in kako se lahko iz nekega sistema, v našem primeru iz neke arhitekture, naučimo in bolje razumemo človeški duh. V nadaljevanju se bomo osredotočili le na dve od navedenih; na SOAR, ker je ena izmed prvih arhitektur in ker je ta povezana s snovalci celotnega področja simbolnih sistemov, to je Newellom, in Simonom, in ACT-R, ki je predstavnik hibridnih sistemov (Anderson et al., 2004, Anderson, 2007), saj menimo, da je le-ta trenutno vodilna in najbolj celovita hibridna arhitektura – verjetno trenutno ena od najboljših delujočih abstrakcij človeškega duha. Pri svojih izvajanjih se ne bomo osredotočil na hevristični del ACT-R-ja, ampak ga bomo uporabil le kot model, kot orodje za filozofsko razlago duha, tj. za razlago zvez med možgani in duhom kot arhitekturo, ki ta dva (možgane in duh) povezuje v neločljivo celoto.

6.5.1 SOAR

Eno prvih kognitivnih arhitektur je zasnoval Allen Newell v sodelovanju z Johnom Lairdom in Paulom Rosenbloomom. (Newell, 1980) Njeno izvirno ime je bilo SOAR (angl. State Operator And Results). Trenutna verzija je znana pod istim

imenom *SOAR*,⁶⁸ in je zelo ozko vezana na hipotezo fizikalnih simbolnih sistemov. Za reševanje problemov uporablja hevristične iskalne približke simbolnih sistemov in ne vsebuje nevronske mreže. Vse znanje je predstavljeno v celotni arhitekturi na enak način in temelji izključno na pravilih in manipulaciji in transformaciji simbolov (podobno kot v primeru Elize).

SOAR je splošna kognitivna arhitektura za razvijanje sistemov in prikazovanje inteligentnih obnašanj. Raziskovalci, tako s področja umetne inteligence kot tudi s področja kognitivne znanosti, uporabljajo SOAR za najrazličnejše naloge. Uporablja se vse od leta 1983. V tem času je bilo razvitih mnogo različnih verzij, danes je aktualna verzija 9. generacije, verzija 9.3.1 z datumom Junij 2011. Ta verzija je nadaljevanje verzije 9.3.0 iz marca 2010. Poleg tega, da so odpravili pomanjkljivosti in napake predhodnih verzij, so jo nadgradili z *utrjevalnim učenjem* (angl. *reinforcement learning* - RL),⁶⁹ epizodnim spominom (EpMem) in semantičnim spominom (SMem). SOAR naj bi bil namenjen za:

- delo, ki se pričakuje od inteligentnih agentov, od enostavnih in natančno definiranih nalog do izjemno težkih problemov brez definiranih ciljev;
- predstavitve in uporabo ustreznih oblik znanja, kot je proceduralno, deklarativno, epizodno in morebiti tudi slikovno;
- reševanja problemov z uporabo različnih metod;
- povezovanje z zunanjim svetom in
- učenje vseh značilnosti in zahtev nalog in njihova predstavitev.

Skratka, SOAR naj bi podpiral vse sposobnosti, ki se zahtevajo od splošnih inteligentnih sistemov. Zaradi kompleksnosti zastavljenih ciljev pa je dejansko SOAR sposoben reševati le omejene probleme. Največ, kar danes od njega lahko pričakujemo je, da bo sposoben aproksimirati nekatere racionalna vedenja, pri čemer se vse odločitve izvajajo s pomočjo kombinacije relevantnih znanj med samim izvajanjem. Vsaka odločitev se izvaja na osnovi interpretacije senzoričnih podatkov, vsebine delovnega spomina, ki je bila ustvarjena za reševanje primarnega problema, in relevantnega znanja, shranjenega in priklicanega iz dolgotrajnega spomina.

⁶⁸ Program je dosegljiv na spletnem naslovu: www.sitemaker.umich.edu/soar.

⁶⁹ Utrjevalno učenje (angl. reinforcement learning – RL) so zasnovali behavioristični psihologi kot področje strojnega učenja in računalništva, povezano s tem, kako agenti kadarkoli delujejo v okolju zato, da maksimirajo nekatere zapise celovitega uspeha.

Odločitve niso nikoli prehodno sestavljene v neprekinjenih sekvencah. Za lažje razumevanje si to oglejmo na primeru igre s kockami. (Ladri, Debinsky, Tinkerhes, 2011)

PRIMER: IGRA S KOCKAMI

Eden od izzivov za kognitivne arhitekture je tudi uspešna uporaba različnih oblik znanja in učenja. V nadaljevanju bo prikazana študija primera SOAR arhitekture – igra s kockami, kjer igra več igralcev. Pri tem bo glavno vlogo igralo verjetnostno sklepanje in uporaba hevrističnih simbolnih metod. Prikazana bo uporaba različnih agentov ob uporabi različnih kombinacij verjetnostnega odločanja, hevrističnega simbolnega sklepanja, nasprotujočega si modeliranja in učenja. Prikazano bo, kako SOAR mehanizem učenja na osnovi pravil v implicitnem sklepanju menjuje premišljeno sklepanje z verjetnostnim sklepanjem. Pri tem pa bo nato uporabljeno utrjevalno učenje za nadaljnje primere. Primer pravil utrjevalnega učenja v SOARU pri igri s kockami je prikazan v spodnji tabeli.

Tabela 6.2: Primer pravil v gradniku.

```
Dice Rule 1:
If the operator is to bid five 2's with no push and
    there are zero 1's and one 2 dice, and four unknown
dice then
create a numeric preference of -0.8754 for that
operator

Dice Rule 2:
If the operator is to bid five 2's pushing one 1 and
two 2's and
    the previous bid was two 1's and
    there are five dice under my cup, with one 1 and
two 2's and
    the other player has only a single die, which is
unknown, then
create a numeric preference of 0.12 for that operator

Dice Rule 3:
If the operator is to challenge and
    the bid by the previous player is two 1's and
    the player pushed and rerolled and
    there are two unknown dice and one 1 known, then
create a numeric preference of 0.38 for that operator
```

Pravila kockanja

Pravila igre so identična tistim, ki jih ljudje uporabljajo pri igranju kock. Pri igri se uporablja 5 poker kock, igra pa lahko ima na različnih koncih sveta različna imena, kot Perudo, Dudo ali Liar kocke. Igra se prične tako, da se igralci razporedijo naključno na igralna mesta (tako, kot da sedejo okrog mize). Vsak igralec ima na začetku 5 kock in posodico, v kateri meša in skriva kocke. Igra je sestavljena iz več krogov in na začetku vsakega kroga igralci zmešajo svoje kocke. Vsak lahko vidi le svoje kocke in jih pred pogledi drugih skriva s posodico. Prvi igralec je izbran naključno, potem pa mu sledijo ostali po vrsti. Med igro se *napoveduje* in *stavi*. Pri tem uspešnost igralčevih napovedi in stav ni odvisna le od igralca samega, temveč tudi od kock nasprotnikov. Zaradi te nezanesljivosti SOAR uporablja verjetnostno sklepanje, npr. pri stavi je zelo dobro, če poznamo verjetnost zmage.

Rezultati analize SOAR igre s kockami so pokazale, da so pri uporabi takšnih sistemov potrebne štiri vrste znanj:

- znanja s področja teorije verjetnosti oz. pričakovano število izračunov;
- znanja o različnih modelih in različnih nasprotnih modelih;
- ekspertna heuristika (metodologija igranja);
- znanja na podlagi izkustvenega učenja (že odigrane partije).

Dobro izdelan kognitivni agent mora torej imeti vse te zmožnosti.

6.6 Kognitivni hibridni sistemi

Kognitivno arhitekturo so na področje kognitivne znanosti vpeljali računalničarji in psihologi. Kognitivna arhitektura je podobna programskemu jeziku. Raziskovalcem daje orodje za ustvarjanje kognitivnih modelov ob uporabi »navadnih« programskih jezikov in »navadnih« zbirk orodij.

Kot smo že omenili, poznamo dve po zgledu človeških možganov diametralno ločeni, samostojni arhitekturi modeliranja duha, tj. koncepta shranjevanja in procesiranja informacij. To sta:

- *Simbolni koncept*,⁷⁰ ki je povezan s hipotezo fizikalnega simbolnega sistema. Ekstremna verzija hipoteze fizikalnih simbolnih sistemov trdi, da vse informacijsko procesiranje sestoji le iz manipuliranja in transformiranja fizikalnih simbolnih struktur. Verjetno sta tudi Newell in Simon, ko sta definirala kognitivne sisteme, imela v mislih nekaj podobnega. (Newell, Simon, 1972, Newell, 1980)
- *Porazdeljeni koncept*,⁷¹ ki je povezan z umetnimi nevronske mrežami. Skrajna verzija tega koncepta trdi, da so modeli nevronske mreže vse, kar potrebujemo za modeliranje kognicije, zaradi česar so fizikalne simbolne strukture popolnoma odveč.

Na osnovi teh konceptov sta bili v polpretekli zgodovini zasnovani dve nasprotni arhitekturi:

- čista Fodorjevska arhitektura (Fodor, 1983) z ostrim razlikovanjem med modularnim in centralnim procesiranjem;
- čista modularna arhitektura, ozko povezana z *evolucijsko psihologijo*,⁷² ki sta jo zasnovala psihologa Leda Cosmides in John Tooby (Cosmides et al., 1992, Tooby, Cosmides, 1990, 1992).

V nadaljevanju bomo skušali ta dva različna aspekta mentalnih arhitektur zblížati, pogledati, kako bi lahko simbolni in porazdeljeni koncept predstavitve duha kombinirali v enovit sistem.

Ugotovimo lahko, da razlika med fizikalnimi simbolnimi sistemi in umetnimi nevronske mrežami kljub vsemu ni tako velika, da je ne bi bilo moč premostiti. Če se ozremo na določene znane primere in modele, lahko iz njih razberemo, da sta oba sistema načelno primerna za popolnoma različne naloge (probleme) in različne načine reševanja teh problemov. Na primer tipi problemov, kot so igranje šaha ali

⁷⁰ Glej poglavje 5.

⁷¹ Glej poglavje 5.

⁷² Evolucijska psihologija je veda v družboslovni in naravoslovni znanosti, ki raziskuje psihološke kategorije, kot so spomin, zaznave in jezik, iz perspektive sodobne evolucijske znanosti. Poskuša določiti, katere psihološke veličine razvijajo prilagodljivost kot funkcionalni produkt naravne in spolne selekcije. Evolucijska psihologija zagovarja tezo, da je duh modularna struktura, podobno kot je telo, z različnimi modularnimi prilagoditvami, ki služijo različnim funkcijam. Prav tako zagovarjajo tezo, da je veliko človeških odzivov rezultat psihološke adaptacije, ki se je razvila zato, da rešuje ponavljajoče se probleme, ki jih pred človeka postavlja okolje.

konstruiranje odločitvenega drevesa iz podatkovnih baz, so izrazito primerni za simbolne sisteme. Nasprotno pa so problemi, kako si dojenček predstavlja objekte ali kako prepoznavamo vzorec, izrazito bolj primerni za porazdeljene sisteme.

Resnica pa zagotovo ni nikoli le črno – bela. Zagotovo obstaja dovolj prostora tudi za manj enostranske in bolj uravnotežene modele, ki skušajo premostiti obe skrajnosti in Andersonov ACT-R (angl. Adaptive Control of Thought – Rational) je dober primer za to. (Anderson, 2007)

6.6.1 ACT-R: kratka zgodovina

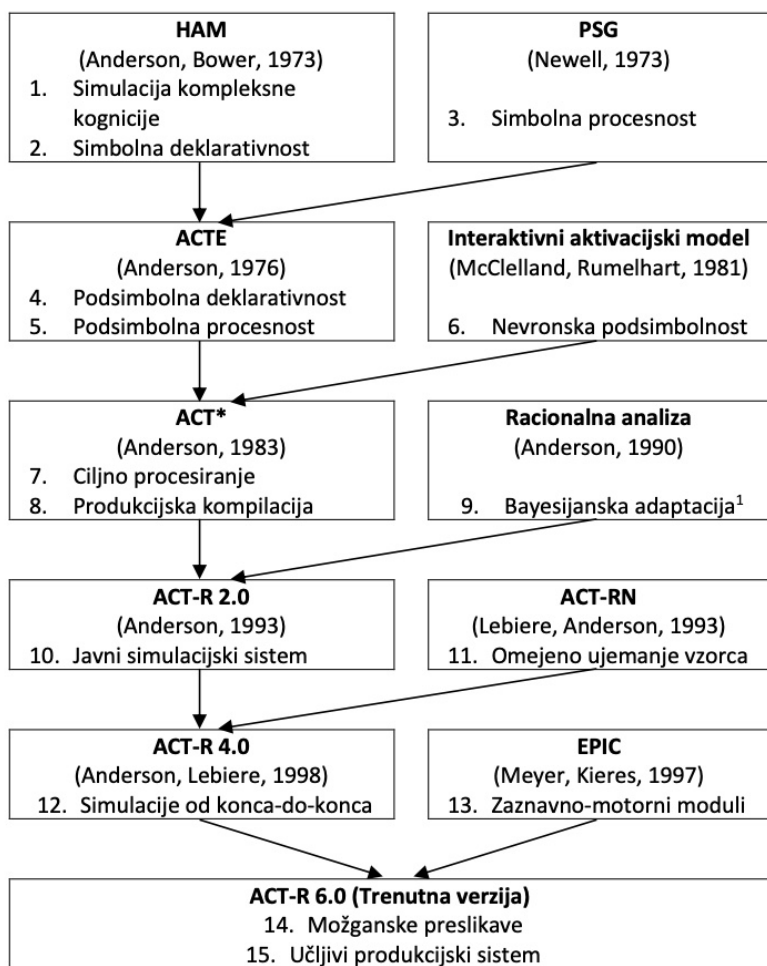
Zgodovinski oris idej, ki so pripeljale do današnjega ACT-R, je simbolično prikazan na sliki 6.2. Oglejmo si ta zgodovinski oris nekoliko podrobneje. Pričetek razvoja ACT-R sega v leto 1973, ko sta bili objavljeni dve temeljni knjigi:

- *Human Associative Memory (HAM)* (1973/1980), ki sta jo napisala J. R. Anderson in Gordon Bower. V njej sta opisala teorijo spomina. HAM je bil eden od novejših poskusov ustvarjanja resne kompleksne teorije človeške kognicije in jo opisati tako, da bo istočasno dovolj natančna in da jo bo ob tem možno tudi računalniško simulirati.
- Zbornik Carnegie simpozija, ki ga je leta 1973 uredil Bill Chase. Ta zbornik je vključeval dve prelomnici, dva Newellova članka. Prvi je bil prelomni: “You can’t play 20 questions with nature and win”. V tem članku je avtor obžaloval težnje kognitivnih psihologov, da bi razdelili svet na majhne neodvisne paradigme, vsako z njej lastnimi vprašanji in logiko. V drugem članku pa je Newell predstavil svoje odgovore na to dilemo tako, da je opisal svojo prvo teorijo procesnega sistema človeške kognicije. Ta ideja je zagotovila manjkajočo sestavino, s katero bi notranjo deklarativno predstavo o HAM-u preoblikovali v funkcionalno teorijo človeškega duha. (Newell, 1990)

Na osnovi tega je nastala prva ACT teorija, ki je poenotena teorija kognicije in temelji na človeškem asociativnem spominu (angl. HAM – human associative memory). (Anderson, Bower, 1980) Izhodiščne ideje so sposojene pri Newellu in Simonu (1973). Ta teorija predpostavlja, poleg ostalega (npr. razlikovanje med *proceduralnim*⁷³

⁷³ Proceduralni ali implicitni spomin se nanaša na obvladovanje pravil in postopkov.

in *deklarativnim*⁷⁴ spominom, znanjem), da obstajajo tudi pod-simbolne vrednosti (aktivacijske vrednosti), ki nadzorujejo dostop do deklarativnih in proceduralnih elementov. Za deklarativni spomin in za aktivacijsko vrednost je bil uporabljen aktivacijski model, ki sta ga razvila Collins in Quillian (1972). Za proceduralni sistem pa je bila uporabljena »trdna« vrednost, ki je temeljila na idejah psihologije izhajajoče iz behavioristične teorije.



Slika 6.2: Prostorske in časovne resolucije različnih orodij in tehnik v nevroznanosti.

⁷⁴ Deklarativni ali eksplicitni spomin predstavlja znanje o dejstvih in dogodkih.

Leta 1983 je izšla Andersonova knjiga, kjer je Anderson pripravil popoln opis ACT in ob tem razvil teorijo, poimenovano ACT*, ki je, ob upoštevanju nevtralnih verjetnostih predpostavk o tem, kako naj bi se ustvarjanje znanja izvajalo, razvila v ACT-R teorijo, v kateri je modelirana kognicija zato, da bi lahko razložili, kako lahko proces ustvarjanja znanja spremenimo v statistično strukturo okolja – zapišemo jo v formalni obliki. V njej je podsimbolno izračunavanje zamenjalo bolj konsistentno izračunavanje (nevronske mreže), ki je temeljilo na idejah na novo se razvijajočega konekcionizma. Najpomembnejše delo s tega področja je bil članek McClellanda in Rumelharta (1981) o interaktivnem aktivacijskem modelu.

Poudariti je treba še, da je ACT* bil pomemben tudi zato, ker sta se v njem pojavili dve novosti, ki sta še danes sestavni del sodobne ACT-R teorije. *Prva* je bila ciljno usmerjeno procesiranje, skratka kontrola kognicije od zgoraj–navzdol, kar je bilo izhodišče za današnji ciljni modul v ACT-R. *Druga* pa je bila niz idej za produkcijsko (procesno) učenje.

Obdobje od 1983 do 1993 je bilo posvečeno predvsem dvema aktivnostma:

1. Razvoju verzij inteligentnih tutorskih sistemov, imenovanih tudi kognitivni tutorji. (Anderson, 1983, 1993) To delo ni imelo bistvenega vpliva na razvoj teorije, njegov glavni rezultat je bilo boljše tehnično razumevanje, kako zgraditi uspešen procesni sistem.
2. Druga aktivnost je bila delo na racionalnih analizah kognicije. (Anderson, 1996) (Od tod izhaja tudi R v imenu ACT-R (R pomeni vpliv racionalnih analiz)).

Leta 1989 je izšla knjiga, ki je povzela teoretični razvoj zadnjega desetletja. (Anderson, 2007) Knjigi je bil priložen tudi CD, ki je vseboval prvo obsežno implementacijo teorije. Za nastanek CD-ja se je potrebno zahvaliti predvsem naraščajočemu zanimanju in razumevanju problematike kognitivnega modeliranja in temu, da je implementacijski programski jezik LISP postal standardiziran jezik. S tem je bila omogočena izmenjava programskih segmentov, kar je imelo za posledico tudi lažjo izmenjavo idej. To je bila osnova za ustanovitev uporabniške skupnosti, kar je povzročilo možnost za popolnoma novo dinamiko razvoja tako teorije kot tudi aplikacije.

Naslednja prelomnica je bilo leto 1998, ko je izšla naslednja knjiga (Anderson in Libiere, 1998) in ob njej ACT-R 4.0. Pri tej verziji moramo izpostaviti predvsem dve spremembi v arhitekturi:

- prva se odraža v kreiranju konekcionistične simulacije ACT-R za iskanje ujemanja vzorca, ki je postal bolj fleksibilen, pa tudi bolj omejevalen, ko je šlo za predpostavljjanje moči procesov, ki privedejo do ujemanja vzorca;
- druga sprememba pa je bila povezana z ustvarjanjem t. i. od konca-do-konca (angl. end-to-end⁷⁵) simulacij, pri katerih se ustvari pri izvajanju določene naloge (tipično računalniško podprto) enako okolje kot pri ljudeh.

Da bi lahko izvajali takšne »od konca-do-konca« simulacije so se pričeli ustvarjati zaznavni in motorni vmesniki. Tako je nastal ACT-R z vrsto modelov človeških zaznav in aktivnosti, kar je ustvarilo »poosebljene« ACT-R modele. Postalo je očitno, da je zaznavno-motorni aspekt pomemben tudi pri tako abstraktnih nalogah, kot je npr. algebra (npr. sledenje gledanju učenca pri kognitivnih tutorjih), zato so se ti aspekti v celoti integrirali v teorijo in pričnemo govoriti o integrirani teoriji duha. EPIC-ova⁷⁶ modularna organiziranost je pri tem zagotovo imela velik vpliv na razvoj ACT-R modularne zgradbe.

Naj omenim še en pomemben dogodek po letu 1998. To je uspešen razvoj teorije procesnega združevanja, ki je ideje iz ACT* prenesel v moderen ACT-R svet. S to teorijo je tesno povezana teorija procesnega učenja (spoznavanje pravil in postopkov), ki se uspešno ujema s teorijo deklarativnega učenja (pridobivanje informacij, usvajanje znanj). Na osnovi tega pa se je pričela razvijati teorija, kako bi se lahko takšni procesi učili z metodo dajanja navodil. Pri tem je pomembno prav to, da je v sedanji verziji ACT-R tudi mehanistična razlaga, kako napreduje posameznik od prejemanja navodila za izvedbo naloge do same izvedbe te naloge (pred tem so se samo programirali specifični procesi, brez globljega razumevanja kako in zakaj). S tem pa je bila odstranjena ena od zadnjih »magij« iz teorije.

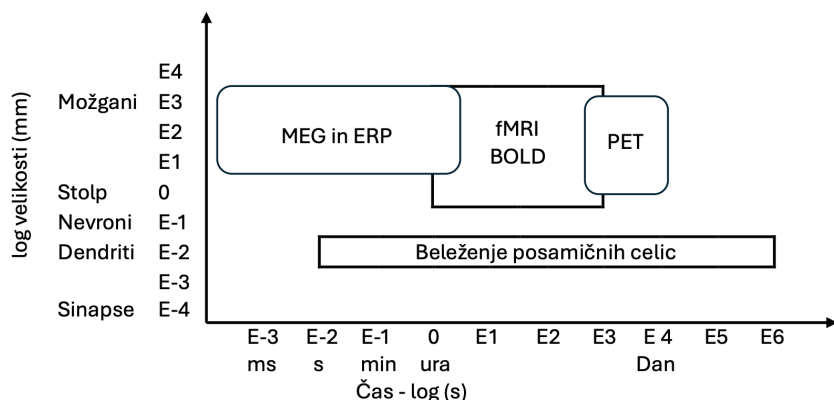
⁷⁵ Princip od konca-do-konca (angl. end-to-end principle) je danes klasičen konstrukcijski princip v računalništvu. Filozofsko izhodišče te teorije je, da poskušamo eliminirati čim več vmesnih korakov ali nivojev in s tem optimirati učinek kateregakoli procesa.

⁷⁶ Glej <http://www.umich.edu/~bcalab/epic.html>.

Naslednji pomemben razvojni korak je bila uporaba *fMRI* (angl. functional magnet resonance imaging) možganskih raziskav, skratka slikovna obdelava podatkov možganov, uporaba *BOLD* (angl. Blood Oxygen Level Dependence) in podobnih tehnik (glej sliko 6.3), kar je omogočilo neposredno potrjevanje in povezovanje posamičnih delov možganov, v katerih se izvajajo določene aktivnosti, in modulov ACT-R, ki izvajajo (modelirajo, simulirajo) sorodne aktivnosti. S tem postane ACT-R še bolj »poosebljen«. Psihološke in nevroznanstvene eksperimentalne metode danes omogočajo, da spremljamo dogajanje v možganih in jih primerjamo z matematičnim modelom. Kar je bilo leta 1990, na začetku razvoja ACT-R, še znanstvena fantastika, je danes realnost.

Nevroznanost in psihologija sta razvoju ACT-R dala nov zagon, saj sta na področje raziskav vnesla vrsto novih eksperimentalnih metod, ki omogočajo zasledovanje stanja in dogajanja na različnih nevroloških dimenzijskih in časovnih nivojih, kar prikazuje spodnja slika 6.3.

Trenutno zadnja verzija je verzija ACT-R 7, ki jo je napisal in jo vzdržuje Dan Bothell. Tej verziji se bomo posvetili tudi v nadaljevanju.



Legenda:

MEG (angl. magnetoencephalography) magnetna encefalografija

PET (angl. position emission tomography) tomografija

fMRI (angl. functional magnetic resonance imaging) funkcionalno magnetno resonančno slikanje

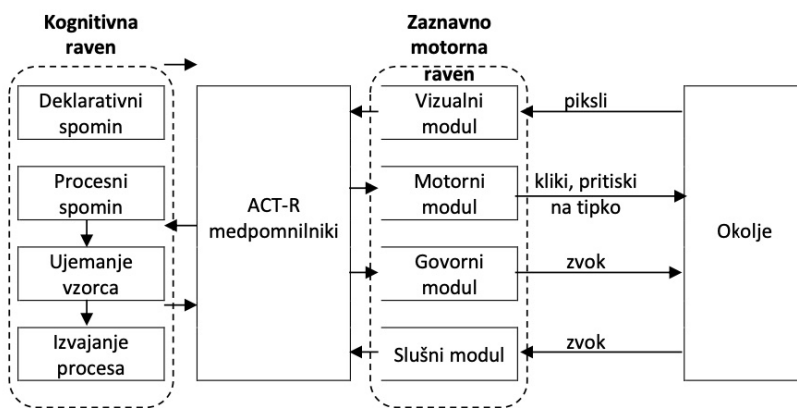
ERP (neuroimaging) uporablja EEG za merjenje možganskih aktivnosti

BOLD (blood-oxygen-level-dependent) zaznavanje različnih magnetnih lastnosti v razmerju kisika v krvi.

Slika 6.3: Prostorske in časovne resolucije različnih orodij in tehnik v nevroznanosti.

6.6.2 Hibridna arhitektura na primeru ACT-R

ACT-R (verzija 6.0) je manj homogen, kot so druge že predstavljene arhitekture, npr. *SOAR*. Prištevamo ga med hibridne arhitekture, predvsem zato, ker vsebuje oboje; *simbolno* (fizikalno simbolno procesiranje informacij) in *podsimbolno* (porazdeljeno procesiranje informacij – umetne nevronske mreže) procesiranje informacij. Vsebuje različne module, ki izvajajo različne kognitivne naloge, tip procesiranja v njih pa je odvisen predvsem od vrste in tipa naloge. Osnovna struktura ACT-R je prikazana na sliki:



Slika: 6.4: ACT-R kognitivna arhitektura.

Preden se osredotočimo na kognitivno razlago, pojasnimo zgornjo sliko z drugega zornega kota. Celotna splošna organizacija na prvi pogled deluje zelo Fodorjevsko. Ima vhodne in izhodne module, vsi ti moduli so vase zaključeni in vsi komunicirajo samo preko sistema medpomnilnikov. Vendar pa pri vsem tem tudi nekaj manjka. Ni sistema, ki bi bil, po Fodorju, odgovoren za centralno procesiranje. Po drugi strani pa je ACT-R sestavljen iz kompleksnih modulov (modulov za posamično celovito področje) in nima modulov, ki bi bili posvečeni prav posebnemu področju. Tako bi si v zvezi z ACT-R lahko zastavili najbolj naravno vprašanje: *Kako se odloča, kaj naj naredi?*

Če je določeno produkcijsko pravilo ali niz pravil aktivno, potem ni nobenih težav pri delovanju sistema. Sistem sledi »navodilom«, ki so zapisana v produkcijskih pravilih samih, izvaja aktivnosti, ki sledijo tem pravilom. Težava nastopi, ko se je

potrebno odločiti, katero produkcijsko pravilo uporabiti? ACT-R je konstruiran tako, da se proces izvaja zaporedno, korak za korakom. V vsakem danem koraku je lahko aktivno le eno produkcijsko pravilo. Večino časa pa seveda obstaja tudi vrsta drugih, lahko tudi podobnih pravil, ki bi lahko bila aktivna, če bi bila izbrana. V Fodorjevski arhitekturi to nalogo opravlja neke vrste centralni procesorski sistem, ki je simbolni sistem. V ACT-R se v nasprotju s tem proces izbire izvaja na podsimbolnem nivoju, in prav zato je ACT-R hibridni sistem. Poskusimo to razložiti nekoliko podrobneje, saj je to ključnega pomena za razumevanje njegove zgradbe in delovanja.

Nalogo izbire v želenem trenutku izvaja v ACT-R *modul za primerjavo vzorcev*, po principu najvišje koristi v danem trenutku. Ta koncept je izpeljan direktno iz *teorije racionalne izbire*,⁷⁷ zato je v imenu sistema tudi črka R, ki predstavlja to »racionalnost«.

Učinkovitost lahko razumemo zelo različno, vendar pa je osnovna ideja ta, da izberemo bolj učinkovit element, npr. procesno pravilo v proceduralnem spominu, z višjo stopnjo učinkovitosti, saj bo od tega celoten kognitivni sistem imel največ koristi. Modul za primerjavo vzorca izvaja neke vrste analizo *strošek—korist* (ob uporabi ustreznih enačb – simbolov), in samo na osnovi tega neodvisno izbere produkcijska pravila, brez kakršnegakoli stika s »centralnim sistemom«. Avtorji ACT-R so to računanje poimenovali *podsimbolno računanje*. Uporabljeni princip je izjemno pomemben princip, ki se standardno uporablja za opis tega, kako delujejo umetne nevronske mreže. Vsako produkcijsko pravilo je samo zase popolnoma simbolno, narejeno po vseh pravilih hipoteze fizikalnih simbolnih sistemov. Proces njihovega delovanja je vnaprej popolnoma določen, vendar pa v njih ni določeno, ali in kdaj se bodo izvajali. To odločitev izvajajo na podsimbolnem nivoju podsimbolna števila, ki predstavljajo učinkovitost tega posameznega produkcijskega pravila. Če se naslonimo na nevroznanost, govorimo o akcijskem potencialu posamičnega nevrona, ki se sproži, ko doseže ali preseže prag tega potenciala. Ta števila so podsimbolna prav in predvsem zato, ker v nobenem segmentu ne odražajo simbole strukture produkcijskega pravila in delujejo popolnoma neodvisno od njega.

⁷⁷ Teorija racionalne izbire, znana tudi kot teorija izbire ali racionalna teorija ukrepanja, je okvir za razumevanje in mnogokrat tudi osnova za formalno modeliranje socialnih in ekonomskih obnašanj. Racionalnost je pogosto uporabljena predpostavka o obnašanju posameznikov v mikroekonomskih modelih, v sodobni politični znanosti, pa tudi pri različnih drugih vedah, kot je npr. sociologija in filozofija. Enačimo jo lahko z instrumentalno racionalnostjo, ki upošteva iskanje najbolj učinkovite poti za dosego določenega cilja, kar je primer tudi v ACT-R.

ACT-R pa ima še druge podsimbolne dimenzije. Dostop do deklarativnega spomina in njegov priklic sta prav tako modelirana kot podsimbolna sistema v skladu s sodobnimi izsledki nevroznanosti in kognitivne psihologije. Osnovni element deklarativnega spomina je *chunk* (sliki 6.7 in 6.8), v nasprotju s proceduralnim spominom, kjer je osnovni element prenosa informacij *procesno pravilo*. Vsak gradnik je povezan s svojim aktivacijskim nivojem, ki je lahko predstavljen kar z določeno številko. Čim višji je aktivacijski nivo nekega gradnika, tem lažje ga prikličemo iz deklarativnega spomina. Aktivacijski nivoji gradnika v deklarativnem spominu so določeni s posebnimi enačbami, podobnimi tistim v produkcijskih pravilih. Pri tem je celoten gradnikov aktivacijski nivo odvisen od dveh osnovnih komponent, in to:

- od tega, koliko je bil gradnik uporaben v preteklosti, kar pomeni, koliko je prispeval k realizaciji skupnih ciljev celotnega sistema, in
- od tega, kako je gradnik relevanten v sedanji situaciji in v sedanjem kontekstu.

6.6.3 ACT-R: teorija

V ACT-R obstajata dve dolgotrajni spominski skladišči: *deklarativni* in *procesni spomin*. (Anderson, Lebiere, 1998) Znanje v deklarativnem spominu, to so dejstva in cilji, predstavljajo termine *gradnikov*. Na simboličnem nivoju so gradniki strukturirani kot semantične mreže. Nasprotno temu pa je znanje v proceduralnem spominu predstavljeno kot procesno pravilo v obliki parov *če—potem*, pri katerih kontrolni tok teče od ene aktivnosti k drugi, ko ene aktivnosti ustvarjajo potrebne pogoje, da se prične naslednja aktivnost. To je produkcijski sistem, ki zagotavlja osnovo za enovito kognitivno teorijo. Izbran proces in ustrezni cilji bodo učinkovali skupno, tako da se bodo priklicane informacije iz deklarativnega spomina povezovale v nove celote, novo znanje. Končno, vsebine priklicanih vozlišč iz deklarativnega spomina bodo uporabljene za zagotavljanje zastavljenih ciljev v odvisnosti od procesnih aktivnosti. Takšne hierarhično organizirane ciljne strukture se uporabljajo tako za predstavljanje načrtov aktivnosti, kot tudi nadzor poteka kognitivnega procesiranja.

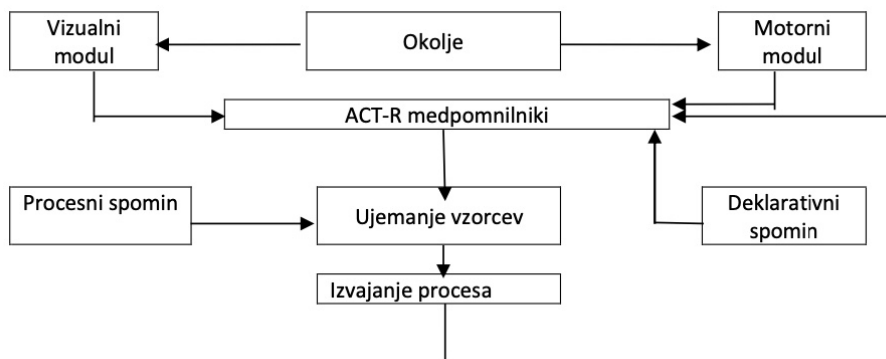
Osnovne komponente ACT-R so:

- moduli,
- medpomnilniki (angl. buffers),

- usklajevalci vzorcev (angl. pattern matcher).

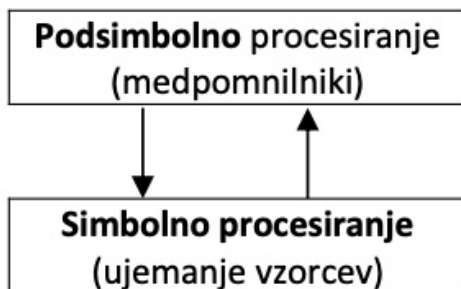
Moduli: obstajata dva tipa modulov:

1. *Zaznavno-pogonski moduli*, ki skrbijo za stik z realnim svetom (okoljem) (tj. za simulacijo realnega sveta). Najbolj dodelana takšna modula v ACT-R sta vizualni in motorni modul.
2. *Spominski moduli*, pri katerih obstajata v ACT-R dva tipa:
 - *Deklarativni spomin*, kjer so spravljene informacije v gradnikih, kot so Ljubljana je glavno mesto Slovenije, Francija je država v Evropi ali $2+3=5$.
 - *Procesni spomin*, ki ustvarja *proces*. Termin procesiranje predstavlja znanja o tem, kako delamo izdelke: npr., znanje, kako pritisnemo črko “Q” na tipkovnici ali kako vozimo avto.



Slika 6.5: Zgradba ACT-R.

Medpomnilniki: ACT-R je povezan s svojimi moduli, razen s proceduralnim spominskim modulom, preko medpomnilnikov. Pri vsakem modulu ustrezen medpomnilnik služi kot vmesnik za ta modul. Vsebina medpomnilnika v danem trenutku predstavlja stanje ACT-R tega trenutka, s čimer ACT-R »simulira« zavest.



Slika 6.6: Simbolno podsimbolne zveze.

Usklajevalci vzorcev: Iščejo procese, ki bodo uskladili trenutno stanje medpomnilnikov. V danem trenutku se lahko izvede samo eno takšno procesiranje. Ko je izvršeno to procesiranje, se lahko modificira medpomnilnike in s tem spremeni stanje sistema. Tako se v ACT-R kognicija razvija z zaporednostjo procesnih iniciacij, izmenjevanjem simbolnih in podsimbolnih procesov (slika 6.6).

ACT-R je *hibridna kognitivna arhitektura*. Njena *simbolna* struktura je produkcijski sistem; *podsimbolna* struktura pa je predstavljena kot niz vzporednih procesov, ki jih lahko povzamemo s številnimi matematičnimi enačbami. Podsimbolne enačbe nadzorujejo mnoge simbolne procese. Na primer, če nekaj produkcij uskladi stanje medpomnilnikov, potem podsimbolne enačbe določijo relativne prednosti in slabosti in izmed njih izberejo posamezno produkcijo z najvišjo stopnjo koristnosti. Podobno, če je določeno dejstvo potrebno priklicati iz deklarativnega spomina, je to odvisno od podsimbolnih obnovitvenih enačb, ki upoštevajo zveze in zgodovino uporabe tega dejstva. Podsimbolni mehanizmi so prav tako odgovorni za večino procesov učenja v ACT-R.

6.6.4 Primer: algebra

Pa si koncept modulske gradnje v ACT-R oglejmo na majhnem fragmentu iz algebre. V tem procesu sodeluje pet modulov (sliki 6.10 in 6.11):

1. vizualni modul je tisti, ki predstavi enačbo kot npr.: $3x - 5 = 7$;

2. problemski modul (angl. problem state modul, včasih poimenovan tudi angl. imaginal modul) ustvarja trenutno mentalno predstavo o problemu, npr. študent mora spremeniti izvirno enačbo v mentalno predstavo: $3x = 12$;
3. kontrolni modul (tudi modul ciljev) ohranja eno primarno namero reševanja problema, npr. izvedbo algebraične transformacije;
4. deklarativni modul priključuje kritične informacije iz deklarativnega spomina, kot npr. $7 + 5 = 12$;
5. manualni modul programira rezultat, npr. $x = 4$.

Vsak od teh modulov ponazarja določen del možganov; za ACT-R je zapisana tudi teorija, kako potekajo interni procesi v teh modulih. Šesti, centralni procesni modul razpoznavlja vzorce informacij v medpomnilnikih in reagira tako, da pošlje zahteve posameznim modulom. Njegovo delovanje temelji na procesnih pravilih. Na primeru zgornje enačbe algebraičnega izraza: $3x - 5 = 7$ se glasi:

Če je cilj rešiti enačbo,

In ima enačba obliko »izraz – število 1 = število 2«

Zapiši »izraz = število 2 + število 1«

Tako se prva vrsta izvaja v kontrolnem (ciljnem) medpomnilniku, druga v vizualnem medpomnilniku in tretja kot manualna aktivnost v manualnem modulu. Na ta način se s ponavljanjem naloge skrajšuje potreben čas za izvedbo naloge.

6.6.5 Stopnje usvajanja veščin

Poleg kognitivnega področja je ACT-R namenjen tudi psihomotoričnemu področju, torej usvajanju veščin. Te so razdeljene na tri stopnje:

- kognitivna stopnja,
- asociativna stopnja
- avtonomna stopnja (rutina).

Usvajanje kognitivnih veščin je progresiven proces od kognitivne do avtonomne stopnje, ki, po ACT-R teoriji, predstavlja transformacijo iz deklarativnega do proceduralnega znanja skozi proces posploševanja (od posebnosti k splošnosti).

Proces se pričinja z interpretativno aplikacijo deklarativnega znanja pri kognitivni stopnji. Nato se proces nadaljuje s pretvarjanjem deklarativnega znanja v procesna pravila med asociativno stopnjo. Na koncu procesa postane niz pravil pogoj, da bo aktivnost vse bolj natančno uglasena. Med avtonomno stopnjo se nato potrebna količina truda in napora, potrebna za izpolnjevanje naloge – aktivnosti, neprestano zmanjšuje. Oglejmo si omenjeno nekoliko bolj podrobno.

1. Kognitivna stopnja

Na začetku procesa usvajanja veščin potrebujemo nov vnos v deklarativni obliki (vedno na začetku potrebujemo teoretično razlago, preden se lotimo dela). Na tej stopnji se učenci učijo o nizu dejstev, ki so pomembne za posamično veščino, kot npr. opis postopka. Znanje o tem, kako izvesti postopek je deklarativen. Gre za predstavitev celotnega postopka po korakih. Na tej točki učenec ustvarja aktivnosti z interpretacijo verbalnih izrazov in pazljivo spremlja rezultate aktivnosti, medtem ko izvaja posamične korake celotnega procesa. Proces je v tej fazi zavesten, previden, počasen in zahteva popolno zbranost.

2. Asociativna stopnja

Glavni razvoj na tej stopnji je združevanje deklarativnih znanj. Ta proces razumemo kot proces ustvarjanja uspešnih (novih) postopkov, katerih cilj je hitrejšo izvajanje postopkov, zmanjševanje ponavljanja in odpravljanje predhodnih slabih ali napačnih postopkov. Med asociativno stopnjo se izvaja proces združevanja in procesiranja pomena pretvorbe deklarativnih dejstev v proceduralno obliko (iz dejstev – informacij ustvarjamo pravila). Združevanje je postopek organiziranja vrste aktivnosti v poenoten postopek, kar povzroča povečevanje hitrosti, tako da dele posamičnih korakov sestavljamo v eno samo aktivnost (npr. vožnja avtomobila je sestavljena iz posamičnih faz, ko jih združimo v proces, ne razmišljamo več, kako moramo pritisniti sklopko ipd.). Posledica tega je, da takoj, ko so veščine pretvorjene v proces, ti novi postopki ne zahtevajo več priklica specifičnih deklarativnih informacij (npr. pritiska sklopke) v delovni spomin. Pomembna posledica takšnega procesiranja je, da zmanjša obremenitev delovnega spomina in s tem povečuje storilnost (postopki postanejo rutina).

3. *Autonomna stopnja (rutina)*

Ko so posamične veščine združene v proces specifičnih nalog, učni proces vpeljuje iskanje izboljšav (faza racionalizacije) pravnega postopka. Na tej stopnji postane postopek dela vse bolj avtomatičen, rutinski in s tem tudi hitrejši. Proces, ki posebej to stopnjo je usklajevanje. Za stopnjo usklajevanja rabijo trije učni mehanizmi: *generalizacija* (posploševanje), *diskriminacija* (razlikovanje) in *utrjevanje*.

Osnovna funkcija *procesa generalizacije* je iskanje in izločevanje skupnega v različnih postopkih. Proces generalizacije ustvarja bolj splošna produkcijska pravila na nekem področju uporabe. Usmerja proces obstoječega znanja v nove situacije. V nasprotju s tem pa *proces diskriminacije* omejuje področja uporabe postopkov in zožuje področje produkcijskih pravil. Proces diskriminacije omejuje področja uporabe pravila na ustrezne, primerne okoliščine. Pomaga identificirati posebne pogoje in množi variante nadzora pogojev pri isti aktivnosti. Ta proces usmerja razvoj močnega, za posamična področja specifičnega dela, kar je predvsem v veliko pomoč pri reševanju konfliktnih situacij. Proces *utrjevanja* je povezan s ponavljanjem istih opravil na nekoliko modificirane načine, s čimer povečujemo verjetnost nezavedne izbire določene oblike obnašanja (podzavestno izvajanje neke naloge, npr. pritiska sklopke med vožnjo avtomobila).

Na tej stopnji učenci postajajo uspešnejši pri izbiri najbolj primernih oblik dela v konkretni situaciji. Kriterij izbire je stopnja pogostnosti uporabe nekega pravila in primernost pravila za želeni postopek (rešitev nekega problema).

6.6.6 Računsko procesiranje v ACT-R

ACT-R ni le teorija, povezana s pridobivanjem znanja. Razvit je bil tudi zato, da bi z njim razložili vprašanje, kako ljudje izberejo ustrezno znanje (način ukrepanja) v posamični situaciji. (Anderson, 1995) Ob uporabi *racionalne analize* ACT-R določi, katero znanje je dosegljivo v zvezi z njegovo različnostjo uporabe v konkretni situaciji. Pravzaprav proces implicitno izvaja *Bayesijansko sklepanje*, s čimer izračunava različne možne kombinacije v odvisnosti od primernosti (praga) neke miselne povezave. (Anderson, 1990) Osnovna enačba se tako glasi:

$$\text{Aktivacijski nivo} = \text{osnovni aktivacijski nivo} + \text{asociacijski prag}$$

Osnovno izhodišče pri uporabi te enačbe je, da je dostopnost do določenih informacij določena z dvema faktorjema, tj. uporabo informacije v preteklosti, kar določa osnovni aktivacijski nivo, in njenega pomena za dani cilj, ki ga določa trenutni za to informacijo pomemben asociacijski prag.

USTVARJANJE PRAVIL

V ACT-R se pravila ustvarijo neposredno na osnovi potreb, zato tudi ni potrebe po ustvarjanju deklarativnih informacij. Ustvarjanje pravil se tako izvede po naslednjem zaporedju:

- obstajajo gradniki tipa »*odvisnost*«, ki predstavljajo človeško razumevanje konkretnih korakov v primerih reševanja problemov;
- *odvisnost* se ustvari, ko človek ustvari niz ciljev za razumevanje dela primera ali navodila;
- ko je odvisen cilj ustvarjen, je vpeljano produkcijsko pravilo iz odvisnosti ter dodano k produkcijskemu sistemu;
- štirje posebni *sloti*⁷⁸ odvisnih struktur so: *slot cilja* (ki vsebuje cilj, preden se je pričel problem reševati), *omejitveni slot* (ki podaja vsebino gradnika in ustvarja most med pogoji in aktivnostmi in nadomešča vzorce v združenih pravilih), *prilagoditveni slot* (ki predstavlja spremenjen cilj) in *skladiščni slot* (angl. *stock slot*) (ki označuje kakršnokoli spremembo ciljev).

6.6.7 Simbolne in podsimbolne zveze

Simbolni nivo v ACT-R je abstraktna karakterizacija tega, kako možganska struktura dekodira znanje. *Podsimbolni nivo* pa je v ACT-R abstraktna karakterizacija vloge nevronskega preračunavanja zato, da je to znanje lahko dostopno.

Povzemimo Newellovo razlago simbolov in z njo opišimo v ACT-R neposredno simbolni, posredno pa tudi podsimbolni nivo. Simboli zagotavljajo dostop do struktur znanja, ki so fizično locirane kjerkoli znotraj sistema. Zahteva po dostopu je omejitev računalniškega sistema. Ta izhaja iz tega, da morajo biti aktivnosti vedno fizično lokalne, povezane le z omejeno količino znanja. Le-to je kodirano znotraj

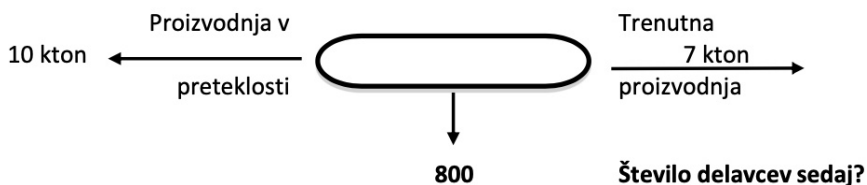
⁷⁸ Slot je ena od osnovnih struktur v računalniški arhitekturi.

končnega volumna prostora ter povezano s človeškim duhom. Količina potrebnega znanja pa je ogromna. Zaradi tega mora obstajati način prenosa znanja iz prostora, kjer je shranjeno, na mesto, kjer ga procesiranje potrebuje (npr. s trdega diska v RAM). Simboli nam zagotavljajo ta dostop. (Newell, 1980: 135 –182)

Simbolno in podsimbolno razlikovanje je posebej intenzivno razvito v dveh moduli ACT-R, v deklarativnem in proceduralnem modulu.

6.6.8 Simbolne in podsimbolne strukture v deklarativnem modulu

Ob upoštevanju deklarativnega modula na simbolnem nivoju ima ACT-R mrežo znanja kodirano v t. i. *gradnikih*. Slika 6.7 prikazuje deklarativni gradnik, v katerem je zapisano neko dejstvo. (Za primer glej Anderson, 2007)



Slika 6.7: Predstavitev deklarativnega gradnika.

Odgovor, skrit v tem gradniku, se nanaša na konkretno vprašanje: *Če je bila proizvodnja v preteklosti 10.000 ton in smo imeli 800 zaposlenih, kakšno je število zaposlenih pri trenutni proizvodnji?* Skratka, na konkretno vprašanje imamo le en konkreten odgovor, ki mora biti nekje zapisan (govorimo o bazah podatkov, o ekspertnem znanju, ki ga imamo ali pa ne). Kaj pa, če ni nobenega gradnika s tem odgovorom? Definirati moramo nevralni proces, v katerem bo izbran ustrezen (konkreten) gradnik kot odgovor. Gradnik ima torej *aktivacijo* na podsimbolnem nivoju. Najbolj aktiven bo tisti gradnik, ki bo priklican (se bomo spomnili), in bo njegova aktivacijska vrednost določila, kako je bil priklican. Aktivacijska vrednost gradnika je določena z izračunom, ki temelji na poskusu abstrakcije vpliva nevronskega učenja, podobnega npr. *Hebbianovemu učenju*⁷⁹ in širjenju aktivacije po nevronskih povezavah.

⁷⁹ Hebbovo učenje je nevronskega učenja, ki je povezano z mnogimi različnimi konekcionističnimi učnimi algoritmi.

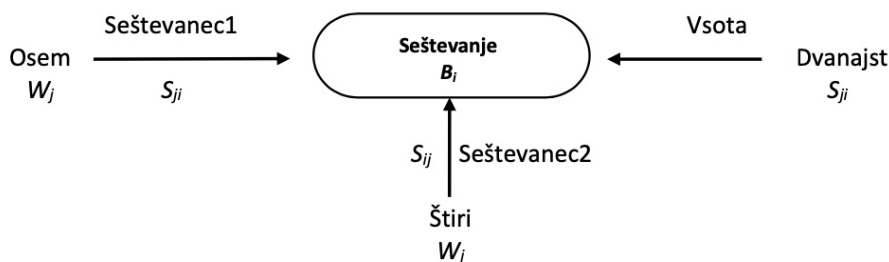
DEKLARATIVNI SPOMIN V KOGNITIVNI ARHITEKTURI

Skoraj vsaka naloga potrebuje deklarativni spomin, ki se v ACT-R kognitivni arhitekturi nahaja v deklarativnem modulu (glej sliko 6.9). Skoraj vsaka naloga zahteva, da deklarativni modul na nek kompleksen način sodeluje z ostalimi moduli. Osnovni elementi deklarativnega spomina, ki deluje na podsimbolnem nivoju, so gradniki, ki imajo aktivacijsko vrednost, ki določa tako hitrost kot tudi uspešnost priklica delov spomina. Aktivacija chunka izkazuje tako osnovno vrednost, poimenovano tudi osnovni aktivacijski nivo, in asociacijsko vrednost, ki je odvisna od trenutne soodvisnosti. To lahko v formalni simbolni obliki zapišemo z aktivacijsko enačbo:

$$A_i = B_i + \sum_{j \in C} W_j S_{ij} \quad (6.1)$$

Kjer je A_i aktivacijski nivo gradnika i , C je soodvisnost, ki je določena kot niz elementov j , ki so trenutno v medpomnilniku, B_i je osnovna aktivacijska vrednost gradnika i , W_j je ustrezna utež za element j v sovisnosti in S_{ij} je trdnost zveze med elementom j in gradnikom i .

Predstavljena enačba je standardna aktivacijska enačba nevrona. Interpretacija enačbe je na primeru enačbe $8 + 4 = 12$ prikazana na spodnji sliki:



Slika 6.8: Predstavitev gradnika s podsimbolno aktivacijsko enačbo na primeru aritmetike.

Ta aktivacijska enačba pa ni edina, v tabeli 6.3 so predstavljene tudi ostale osnovne podsimbolne enačbe v ACT-R, npr. podsimbolna enačba osnovnega nivoja učenja, ki predstavlja, kako se aktivacijski nivo spreminja v odvisnosti od uporabe gradnika v preteklosti ipd. (Anderson, 2007)

Tabela 6.3: ACT-R enačbe za aktivacijo

Aktivnost	Enačba
Aktivacijska enačba	$A_i = B_i + \sum_{j \in C} W_j S_{ij}$
Enačba osnovnega nivoja učenja	$B_i = \ln \left(\sum_{k=1}^n t_k^{-d} \right)$
Enačba za uteženje pozornosti	$W_i = W/n$
Enačba trdnosti asociacije	$S_{ji} = \ln(\text{prob}(i \rightarrow j) / \text{prob}(i))$
Enačba časa priklica v spomin	$\text{Čas} = F e^{-A_i}$
Enačba verjetnosti priklica v spomin	$\text{Prob} = 1 / (1 + e^{-(A_i - \tau)/s})$

6.6.9 Simbolne in podsimbolne strukture v procesnem modulu

Kot smo že omenili, procesni modul sestavljajo procesna pravila – lahko bi jih imenovali tudi izbiranje aktivnosti – kar bi morda bolje opisalo, kaj pravila, preslikave (angl. mapping) počnejo. Slika 6.9 ilustrira procesno pravilo, ki se lahko uporabi pri reševanju enačbe $3 + x = 8$. Procesno pravilo se odziva na vzorce, ki se pojavijo v nizu modulov, in to v prikazanem primeru tako, da dekodira enačbo v vizualnem modulu in nastavlja kontrolno stanja za reševanje enačbe v modulu ciljev.

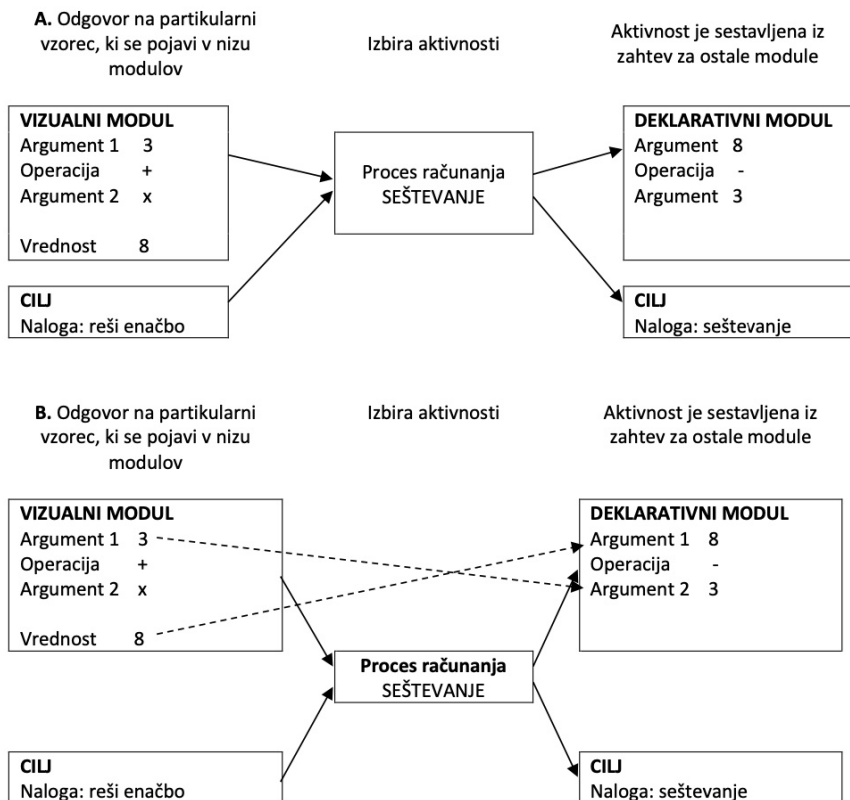
Slika 6.9 B prikazuje splošno pravilo za sliko 6.9 A. To pravilo ni partikularno in ne velja le za ta konkreten primer, skratka za števili 3 in 8, temveč se npr. na mestu *Argument 1* v vizualnem medpomnilniku lahko pojavi katerokoli število, ki bo preslikano na mesto *Argument 2* v deklarativni spomin.

Prikazani proces je primer, ki določa, kako neko informacijo premaknemo iz vmesne lokacije na drugo – končno lokacijo – in to simbolično natančno tako, kot je predlagal Newell. Uporabljajo pa se lahko tudi večprocesna pravila. Pri teh večprocesnih pravilih se odločamo, katero pravilo uporabiti na podsimbolnem nivoju, kjer imajo procesna pravila svoj rang. Izberejo se pravila z najvišjim rangom, za kar uporabljamo posebne izračune, ki upoštevajo posebne aspekte nevronskega utrjevalnega učenja (angl. neural reinforcement learning⁸⁰ ali tudi Q – learning).

⁸⁰ Utrjevalno učenje (angl. reinforcement learning) je področje strojnega učenja v računalništvu, povezano predvsem s tem, kako nek agent kadarkoli izvaja aktivnosti v okolju tako, da maksimira neko stališče skupne nagrade. Pravila, kako to narediti, so pogosto stohastična.

A.: vsebina medpomnilnika, ki se lahko izvede v posebnem primeru.

B.: splošen vzorec dekodiranja v pravilu, uporabljen v tem primeru.



Slika 6.9: Ilustracija procesnih pravil v ACT-R.

Vir: Anderson, 2007

6.6.10 Končna refleksija o simbolnih podsimbolnih razlikah

Vedno se nam lahko dozdeva, da obstajajo razlike med simbolnimi zapisi najrazličnejših računalniških simulacij, konekcioniščni zapisi in ACT-R specifikacijami. Te razlike nedvoumno so, predvsem na nivojih abstrakcij določanja nekih teorij. Najboljši nivoji abstraktnosti za razvoj določene teorije v znanosti pa so najvišji, strateški nivoji. V primeru konekcioniščnih elementov ali v primeru simbolne strukture v ACT-R se zastavi vprašanje, na katerem nivoju bomo dosegli najboljšo poenostavitev zvez in vzajemnega delovanja med možgani in duhom. V

obeh primerih so predlagane rešitve, ki so pomembne za prehod od nevronov do realnih procesov v možganih. Podobno so v obeh primerih pomembne razdalje od funkcij človeških možganov do funkcije simulacij možganov. Verjetno je ta razdalja prav v primeru ACT-R manjša, ACT-R bolj realno simulira dejanske procese in delovanje možganov kot kateri koli drugi sistem. V obeh primerih so bile definirane neke enote in predlagane predvsem zato, da bi lažje merili napredek simulacije povezav med možgani in duhom. Konekcionistični modeli so zelo uspešni pri opisovanju zaznavnih procesov, medtem ko je ACT-R predvsem nadvse uspešen pri opisovanju višje rednih kognitivnih procesov, kot sta npr. algebra in reševanje enačb.

Če se vrnemo k izhodiščnemu vprašanju: Kako se lahko človeški duh dogodi v fizičnem vesolju? Potem je vloga kognitivne arhitekture poiskati in specificirati strukturo možganov na abstraktnem nivoju, ki bo razložila, kako delujejo možgani in kako se duh počuti v njih. Verjamemo, da ACT-R omogoča to najbolje, da omogoča najboljše ločevanje človeškega duha od duhov ostalih bitij. V nadaljevanju bo to podrobneje razloženo in dokazovano, ter upam, tudi potrjeno, za širok krog »vernikov«. (Anderson, 2007)

PRENOS TEORIJE V PRAKSO

Oglejmo si vpliv teoretičnih izsledkov na konstruiranje inštrukcijskega sistema KIS (angl. ISD – Instructional System Design) na primeru teorije procesiranja informacij pri KIS. Dve ključni predpostavki pri teoriji procesiranja informacij, ki imata velik vpliv pri formuliranju inštrukcijskih principov, sta:

1. *Sistem pomnjenja*, ki je aktivno organiziran procesor informacij:
 - Raziskave o pozornosti in zaznavi, ki so filtri za prepoznavanje vzorcev modelov pozornosti, in dualna teorija kodiranja imajo velik vpliv na konstruiranje navodil za oboje; za besedilna, kakor tudi za vizualna sporočila, z namenom maksimirati pozornost in zaznavo učencev.
 - Študije s področja kratkotrajnega spomina, kot so omejen prostor ali kratkotrajnost, vse bolj poudarjajo pomen mnemotehničnih naprav (naprave, ki urijo spomin) za zmanjšanje obremenitev kratkotrajnega spomina, organiziranja informacij v gradnike ali manjše komponente, s čimer povečujemo zmogljivost spomina.

- Teoretične razlage o ohranjanju informacij v dolgotrajnem spominu poudarjajo vpliv različnih pogojev na raven procesiranja. Pomensko dekodiranje pospešuje kasnejše spominjanje. Grafična predstavitev je še posebej učinkovita pri pospeševanju dekodiranja in shranjevanja informacij.

2. *Predhodna znanja* igrajo pomembno vlogo pri učenju:

- Vpliv je dokazljiv z uporabo naprednih organizatorjev in kakršnihkoli inštrukcijskih strategij za utrjevanje aktivacije obstoječih spominskih struktur. Izpopolnjevalne strategije in Ausubelovo vsepomensko učenje (Ausubel, Novak, Hanesian, 1978) obsega pri konstruiranju navodil sistem, ki priporoča uporabo povezanih pomenov novih informacij za vsakega posameznega učenca, prav tako pa tudi, uporabo metafor in analogij, s čimer povečuje uspešnost dajanja navodil;
- Poudariti je potrebno pomen samoregulacijskih veščin pri učenju: zavestno sklepanje in mišljenje.

Z razvojem informacijsko procesnega modela učenja je potrebno naloge zdaj preučevati s perspektive človeškega miselnega procesa. Kognitivne operacije, kot je učenje, potrebuje KIS za celovito reševanje problema, kar postaja osnovni cilj te analize. Informacijsko procesna analiza ciljev uporablja diagrame poteka, s katerimi predstavlja kognitivne operacije korak za korakom in s tem označuje proces odločanja. (Myers, 1995.)

Pri ustvarjanju konceptov aktivnih sistemov pomnjenja je eden od najpomembnejših delov izvajanje operacij z informacijami. Težišče je na kaj in kako je ta sistem povezan z učenjem in kognicijo. V okviru tega je bila predlagana vrsta hipotez za razlaganje različnih tipov spominskih sistemov, predstavitev/struktur znanja v spominu in kako te predstave vplivajo in medsebojno učinkujejo na vstopne informacije. Te hipoteze zagotavljajo implikacije o tem, kako nadzorovati pogoje dajanja navodil. Predpostavlja se, da korespondenca med pogoji za navodila in notranjimi pogoji tega aktivnega sistema spomina maksimira učinkovitost navodil.

6.6.11 ACT-R in KIS

Konstruiranje inštrukcijskega sistema (KIS) se rabi kot računalniško simulacijsko orodje. Učitelji lahko z njim ustvarjajo inteligentne tutorje, računalniško podprte sisteme navodil in učijo kognitivne veščine, ki temeljijo na osnovi produkcijskega sistema.

Analiza načina učenja ljudi je uporabna v praksi implicitno ali eksplicitno v obliki navodil na osnovi pravil. Tudi ACT-R teorija omogoča dva osnovna mehanizma nadzora delovnega spomina:

- širjenje aktivacijskega vira (angl. spreading source activating);
- propadanje osnovnonivojskih aktivnosti (angl. decaying base-level activation).

Te teoretične predpostavke poudarjajo strategijo pretresanja učnih dejstev in strategije učenja ob delu za pridobivanju proceduralnega znanja. Pri tem se lahko uporabljajo različne metode, najpomembnejše so:

- *Analiza naloge*

ACT-R teorija (podobno kot analiza izvrševanja kognitivnih veščin ali tudi analiza njihovega usvajanja) zagotavlja okvir za analizo znanja in veščin, ki si jih mora učenec razviti zato, da lahko izvaja ustrezne naloge. Prikazuje pomen ciljev in namen razvoja tega, kar se učimo. Model ciljev, operatorjev, metod in pravil za izbiro – COMP⁸¹ (angl. Goal-Operated-Method-Selection – GOMS) – ustreza karakteristikam mehanizmov ciljno vodenega učenja in učenja na osnovi pravil in je del ACT-R teorije. Model analizira cilje, operatorje, metode in izbira pravila za določeno nalogo tako, da nalogo razbije na pomembne cilje in podcilje.

- *Proces reševanja problemov*

ACT-R teorija zahteva, da je osnovna kontrolna arhitektura med različnimi učnimi situacijami hierarhična, ciljno strukturirana in organizirana za reševanje problemov. (Anderson, 1982) Modelira bistveni proces reševanja problemov: identificiranje

⁸¹ COMP – cilji, operatorji, metode in pravila za izbiro (angl. GOMS – Goals, Operators, Methods, and Selection rules) – je oblika posebnega človeškega informacijsko procesnega modela za opazovanje povezav človek–računalnik.

ciljne strukture problemskega prostora in analiziranje soodvisnosti informacij. Vključitev tega modela upošteva učenje z reševanjem problemov in posredovanjem takojšnjih povratnih informacij učencu.

– *Analogija*

Mehanizem analogije vsebuje naslednje korake:

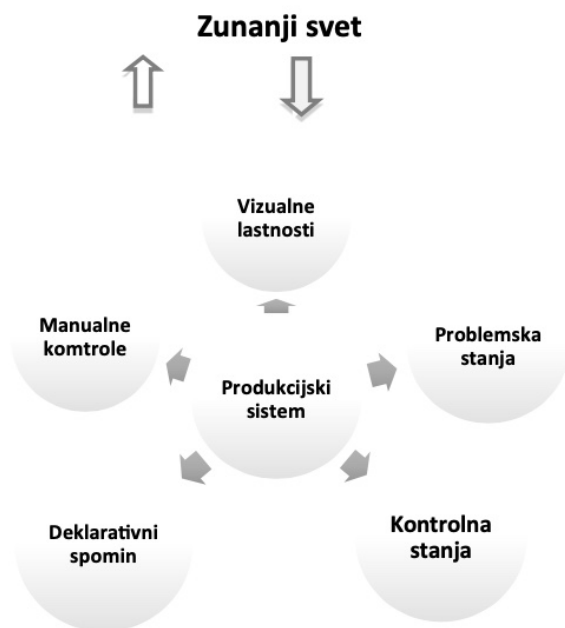
1. V nekem trenutku se ustvari deklarativna struktura znanja, tako da predstavi razumevanje posameznih korakov pri reševanju problema.
2. V naslednjem trenutku lahko, če se vzpostavi podobna problemska situacija, to deklarativno strukturo priključimo in uporabimo kot osnovo analogije. Dve iskanji se priključeta v spomin: vir analogije in povezava med predhodnim in sedanjim primerom.
3. Analogija je arhitekturno enostavna in ustvari produkcijsko pravilo s katerim prikaže ACT-R razumevanje, kako je primer uporabljen v trenutni situaciji. To produkcijsko pravilo je nato lahko uporabno tudi za morebitno kasnejšo uporabo brez dodatne analogije.
4. Če produkcijsko pravilo ni dovolj trdno, da bi se sprožilo, se mora reanalogizirati. Postalo bo trdnejše in po potrebi bolj uporabno.

6.6.12 Modeliranje možganov v ACT-R

Funkcijo in strukturo človeških možganov bi v grobem lahko razdelili na:

- *percepcijsko* – zaznave pomembnih informacij;
- *motorično* – izvajanje določenih aktivnosti;
- *centralno* – koordinacijo aktivnosti in misli za doseg želenih ciljev.

Za predstavitev zgradbe in delovanja človeških možganov Anderson (2007) predlaga, da uporabimo klasična pravila arhitekture, ki smo jih podrobneje obravnavali v predhodnih poglavjih. Skratka, govorimo o kognitivni arhitekturi, ki je po Andersonu: »*Specificirana struktura možganov na abstraktnem nivoju, ki razlaga, kako doseči funkcioniranje duha*«. (Anderson, 2007: 7) Sliki 6.10 in 6.11 prikazujeta takšno simbolno arhitekturo človeških možganov – ACT-R.



Slika 6.10: Povezava posameznih modulov v ACT-R.

V tej arhitekturi se kognicija pojavi kot interakcija števila neodvisnih modulov. Vloga kognitivne arhitekture je poiskati specifikacijo strukture možganov na ravni abstrakcije, ki bo razložila, kako možgani v resnici delujejo. Anderson je v svoji arhitekturi predvidel 8 modulov, vsakemu od teh modulov pa ustreza prav poseben center v človeških možganih. Lokalizacija posameznih kognitivnih funkcij v človeških možganih (skratka lokacija posameznih modulov po Andersonu) pa je prikazana na spodnji sliki:

Tabela 6.4: Prikaz lokacij 8 področij (modulov) v človeških možganih.

Št.	MODULI	PODROČJA V MOŽGANIH
1	Vidni	Fusiform
2	Slušni	Auditory
3	Motorični	Motor
4	Glasovni	Motor
5	Predstave	Parietal
6	Deklarativni (spomin)	Prefrontal
7	Cilji	ACC
8	Proceduralni (spomin)	Caudate

Vir: Anderson, 2007: 75

Teh 8 modulov bi lahko razdelili v dve strukturi modulov:

1. *Periferni moduli (zažnavno motorični moduli):*

- vizualni,
- slušni,
- ročni,
- govorni.

2. *Centralni moduli (kognitivni moduli):*

- predstavn (ujemanje vzorcev),
- deklarativni,
- ciljni,
- procesni moduli.

Prvi ključni modul za delovanje celotnih možganov je *deklarativni modul*, ki predstavlja deklarativni spomin, torej vse naše spomine, izkušnje, vso našo preteklost, vse to, kar nam daje identiteto. Njegova funkcija je narediti znanje dostopno v obliki, uporabi za različne namene. Pri tem obstaja količinska omejitev informacij, ki jih lahko vzdržujemo in prikličemo. V deklarativnem spominu je zato predpisan postopek, kako priklicati posamezen spomin in kako ga obravnavati. Deklarativni spomin je kritičen za naše prilagajanje spremembam v svetu. Vendar samo to ni dovolj. Za njegovo uporabo se zahteva kompleksen, preudaren proces, v katerem sklepamo, delamo povzetke in napovedi na osnovi teh znanj.

Drugi ključni modul je *procesni modul*, ki skrbi za pravilno ravnotežje med ukrepanjem, aktivnostmi in premišljanjem.

Kakšna je koordinacija med posameznimi moduli in kako v vse to vključiti zavest? Če prenehamo verjeti v Chalmersov »trdi problem« zavesti, si lahko zastavimo naslednji vprašanj:

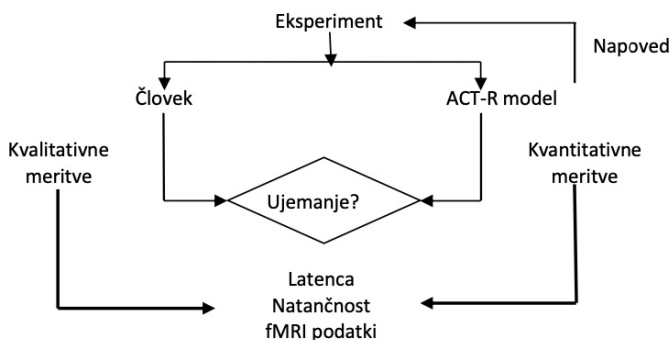
- Kako je lahko zavest rešitev temeljnega problema – procesa razmišljanja v možganih;

- Kako lahko z zavestjo rešimo problem procesa mišljenja v možganih?

Delovanje možganov zahteva komunikacijo med vsemi moduli (glej sliko 6.11) in zaradi tega morajo biti nekatere informacije globalno dostopne. Zato poleg teh osnovnih modulov v možganih potrebujemo še *medpomnilnike* (angl. buffer), ki ustvarjajo informacijske poti. Adaptivna kognicija velikokrat potrebuje te informacijske poti. Tako se zavest kaže kot rešitev potreb po globalni koordinaciji med moduli in bi jo verjetno še najlažje ponazorili prav s temi medpomnilniki oz. če govorimo v jeziku strojev, s hitrim spominom (RAM-om).

6.6.13 Kognitivno modeliranje

Kognitivno modeliranje bomo podrobneje predstavili na modelu ACT-R 6.0, ki predstavlja teorijo o tem, kako deluje človeška kognicija. ACT-R lahko uporabimo kot izhodišče, da na nek problem pogledamo tudi iz drugega zornega kota, iz zornega kota ACT-R. Omogoča nam, da določene predpostavke preverimo in s primerjavo človeških odzivov pri reševanju enake naloge potrjujemo (ali pa tudi ne) pravilnost dobljenih rezultatov.



Slika 6.11: Proces potrjevanja napovedi.

S pojmom rezultati mislimo tradicionalne meritve kognitivne psihologije, kot so:

- latenca, tj. čas (od dražljaja do reakcije) za izvedbo naloge;
- natančnost pri izvajanju naloge in

- (po novejših raziskavah) nevrološke podatke, tj. tiste, ki jih dobimo z fMRI.

Bistvena prednost ACT-R pred ostalimi podobnimi arhitekturami je njegova kompleksnost, ki omogoča raziskovalcem zbiranje kvantitativnih meritev, ki so direktno primerljive s kvantitativnimi podatki dobljenimi z meritvami na ljudeh.

ACT-R je bil uspešno uporabljen za ustvarjanje modelov s področij, kot so:

- učenje in pomnjenje,
- reševanje problemov in odločanje,
- jezik in sporazumevanje,
- zaznava in pozornost,
- kognitivni razvoj in
- posamezno razlikovanje.

Poleg omenjenih aplikacij s področja kognitivne psihologije, pa je bil ACT-R uporabljen tudi za:

- povezave človek – računalnik, izdelavo uporabniških modelov, ki lahko ocenjujejo različne računalniške učinke (interference);
- izobraževanje (kognitivni tutorski sistemi) za »napovedovanje« težav, ki bi jih študenti lahko imeli in na podlagi tega izdelovanje ustrezne pomoči;
- generiranje računalniških slik, s katerimi ustvarjamo kognitivne akterje za ustvarjanje okolja za usposabljanje;
- nevropsihologijo za interpretacijo fMRI podatkov.

Ena od najbolj uspešnih aplikacij je bil verjetno *Kognitivni tutor za matematiko*, ki je bil uporabljen v tisočih šolah ZDA. Ta kognitivni tutor je bil uporabljen kot platforma za raziskovanje učenja in kognitivnega modeliranja, ki se je izvajal v Pittsburgovem centru za znanost o učenju. (Anderson, 2007)

Spodnja slika 6.12 povzema nekatere potencialne aplikacije:



Slika 6.12: Področja uporabe ACT-R.

6.7 Sklep

Šesto poglavje je namenjeno kombinaciji konekcionističnih in simbolnih sistemov, ki jih imenujemo tudi hibridni sistemi. Fizikalni simbolni sistemi so bili v zgodovini prvi, ki se jih je prijel vzdevek inteligentni sistemi. Ogledamo si nekaj osnovnih poskusov in na temelju teh poskušamo razložiti filozofske principe delovanja teh sistemov ter nekatere trenutno najbolj delujoče simbolne arhitekture.

Na kratko smo opisali trenutno najbolj aktualne kognitivne sisteme, njihovo delovanje in na primeru SOAR arhitekture pokazali, kako bi se lahko naloge, ki zahtevajo verjetnostna znanja, reševale s kognitivnimi simbolnimi arhitekturami. Na začetku je SOAR uporabil preudarno sklepanje, s katerim je izračunal verjetnost pričakovane vrednosti, ki jo nato lahko pri odločanju kombinira s hevrističnim

znanjem in modeli nasprotovanja. Pokazali smo, da so za takšne naloge gradniki (chunki) izredno primerni, saj lahko s pomočjo njih sistem kombinira različne tipe znanj in ustvarja nova pravila. Nato pa s pomočjo utrjevalnega učenja ta pravila prilagaja tako dolgo, da z njimi doseže znatne izboljšave želenih lastnosti. Za prikaz delovanja je bila igra s kockami izjemno primerna in je dokazala, da so simbolni kognitivni sistemi zelo primerni za reševanje takšnih in podobnih problemov.

Če na koncu izpostavimo nekatere osnovne kontraste med simbolnim in podsimbolnim, konekcionističnim nivojem, lahko ugotovimo, da je *gradnik* sam po sebi simbolni način posredovanja informacij. Zgrajen je iz osnovnih simbolov in tudi njegov način delovanja znotraj arhitekture je prav tako določen s simbolno strukturo. Vendar pa ne more narediti nič sam od sebe, saj je shranjen znotraj deklarativnega spomina oz. deklarativnega modula. Da bi lahko deloval znotraj sistema, mora biti priklican iz spomina in nameščen v medpomnilnik. Ta sistem, ta priklic, usmerja podsimbolna enačba, ki določa aktivacijski nivo vsakega gradnika kot funkcijo njegove uporabnosti v preteklosti in koristnosti za trenutno situacijo. In to je tudi osnovni problem, zakaj simbolni sistemi samostojno ne morejo izkazovati inteligentnega delovanja. Ta enačba priklica je podsimbolna zato, ker je popolnoma neodvisna od gradnikove notranje simbolne strukture. Iz napisanega izhaja, da vsaj bližnja prihodnost zagotovo ni ne črna in ne bela, oz. ni ne mrežna, konekcionistična, ne fizikalno simbolna, temveč je nujna njuna kombinacija. To hipotezo dokazujemo na osnovi trenutno verjetno najbolj kompleksne in tudi najbolj delujoče kognitivne hibridne arhitekture, torej na primeru ACT-R, ki nam omogoča simuliranje vrste kognitivnih procesov in je po svojem principu in po svojem delovanju danes še najbližje delovanju človeških možganov, kar potrjujejo tudi različne empirične raziskave v nevroznanosti. Spodnja tabela na kratko povzema zveze med simbolnimi in podsimbolnimi dimenzijami v ACT-R, ki so podrobneje opisane v tem poglavju.

Kar ACT-R tako omogoča, je možnost razmišljanja o modelu celovitega organiziranja duha, ki kombinira elemente dveh popolnoma različnih konceptov procesiranja informacij, skratka simbolnega, ki je neposredno povezan s hipotezo fizikalnega simbolnega sistema na eni strani, in podsimbolnega, ki je povezan z modelom umetnih nevronske mreže na drugi.

Preglednica 6.5: Primerjava simbolnih in podsimbolnih dimenzij predstavljanja znanja v ACT-R arhitekturi.

	MEHANIZEM PREDSTAVE		MEHANIZEM UČENJA	
	SIMBOLNO	PODSIMBOLNO	SIMBOLNO	PODSIMBOLNO
DEKLARATIVNI GRADNIKI	Znanje (splošna dejstva), ki ga lahko neposredno ubesedimo.	Relativna aktivacija deklarativnega gradnika, ki vpliva na priklic v spomin.	Dodajanje novih deklarativnih gradnikov v zbirko.	Spreminjanje aktivacije deklarativnega gradnika in spreminjanje trdnosti vezi med gradniki.
PROCESNA PRAVILA	Znanje, potrebno za izvajanje posamičnih aktivnosti v posamičnih situacijah.	Relativna korist produkcijskih pravil, ki vpliva na izbiro.	Dodajanje novih produkcijskih pravil v zbirko.	Spreminjanje koristi produkcijskih pravil.

Res je, da avtorji ACT-R na ta tip podsimbolnega procesiranja v resnici niso (vsaj ne v začetku) gledali na opisani način, tj. kot na implementiranje koncepta umetnih nevronske mreže oz. kot na prevlado umetnih nevronske mreže nad *simbolnim* procesiranjem informacij, zato tudi te arhitekture ne moremo neposredno preslikati na razliko med fizikalnim simbolnim sistemom in umetnimi nevronskimi mrežami. Vendar pa se kljub temu iz tega lahko naučimo vsaj dvoje:

- *Prvič* pravilno razmišljanje o modularni organiziranosti duha zahteva razmišljanje tudi o tem, kako bi različni moduli lahko izvajali naloge procesiranja informacij.
- *Drugič*, različni deli arhitekture duha bi lahko izkoriščali različne modele procesiranja informacij. Nekatere naloge že po naravi težijo bolj k simbolnemu načinu reševanja, medtem ko druge bolj k podsimbolnemu, seveda če na problem ne gledamo iz drugega zornega kota, iz zornega kota teorije dinamičnih sistemov. In znanstvene razprave o teh modelih procesiranja informacij zagotovo niso in ne morejo biti le po načelu vse ali nič, samo eno ali samo drugo.

7 Dinamični sistemi

Kaj so dinamični sistemi (DS)? Najširše bi lahko opisali dinamične sisteme kot katerekoli sisteme, ki se razvijajo po določenih zakonih skozi čas. Tako je tudi naš sončni sistem dinamični sistem. Iz tega izhaja, da smo dinamični sistemi tudi mi vsi, saj smo del našega sončnega sistema, s tem pa tudi npr. Turingov stroj in umetne nevronske mreže. (Bermudez, 2010)

Splošna teorija sistemov je abstraktna in interdisciplinarna teorija, ki skuša najti lastnosti sistemov, ki pri transformaciji ne spremenijo ne svojih vrednosti, ne zakonov upravljanja, ki so značilni za sistem. To je metodološka veda o sistemih. Njen cilj je najti splošne lastnosti različnih sistemov. Sistem je celota, sestavljena iz manjših delov, in je lahko vse, kar nas obkroža. Pojem sistemskega pristopa je eden od načinov za iskanje najboljših ali vsaj dovolj dobrih rešitev problemov, ob upoštevanju celovitosti in omejitev sistema. Splošna teorija sistemov je neke vrste okvir metodologij za reševanje kompleksnih problemov. Cilj proučevanja teorije sistemov je, da se naučimo sistematskega pristopa pri ugotavljanju in reševanju problemov in spoznamo zakone obnašanja sistema, ki izvirajo iz različnih povezav med elementi sistema in vplivov iz okolja. Sistemski pristop je iskanje najboljših rešitev iz množice možnih rešitev ob upoštevanju celovitosti in omejitev problema. Poznati je potrebno metode za proučevanje problema, glede na njegove lastnosti (zapletenost, odprtost, dinamičnost, ...). Lastnosti, ki so skupne vsem sistemom so

cilji, stanje, omejitve, funkcioniranje, komuniciranje in upravljanje, stabilnost ter gibanje. (Von Bertalanffy, 1950)

Osredotočimo se na del splošne teorije sistemov, na *teorijo dinamičnih sistemov* – TDS, ki je področje uporabne matematike in se uporablja za opisovanje obnašanja kompleksnih dinamičnih sistemov. Je močno matematično orodje za povečevanje razumevanja razvoja določenih tipov naravnih pojavov. Newtonova mehanika je morda eden od najbolj znanih primerov dinamičnega modeliranja, vendar pa je za vse dinamične sisteme značilno, da imajo določene poteze enake. Dinamični modeli opisujejo razvoj zvez med relativno majhnim številom vrednosti, ki so časovne spremenljivke.

Po navadi jih ponazarjamo s sistemom *diferencialnih*⁸² ali *diferenčnih enačb*.⁸³ Če uporabljamo diferencialne enačbe, potem to teorijo imenujmo *kontinuiran dinamični sistem*, če pa uporabljamo diferenčne enačbe, potem govorimo o *diskretnih dinamičnih sistemih*. Pri teh preračunih v današnjih raziskavah vse pogosteje uporabljamo tudi *teorijo kaosa*.⁸⁴ (Alligood, Sauer in Yorke, 1997)

TDS in teorija kaosa se ubadeta z dolgoročnim obnašanjem dinamičnih sistemov. Pri tem te raziskave niso osredotočene na iskanje natančne rešitve (kar je mnogokrat tudi brezupno), temveč na iskanje odgovorov na nekatera vprašanja, kot npr.:

- Ali se bo sistem stabiliziral v daljšem časovnem intervalu, in če se bo, kakšno vrednost bo zavzel?
- Ali je njegovo obnašanje v daljšem časovnem intervalu odvisno od njegovih začetnih pogojev?

Pomemben cilj je tudi opis fiksnih točk (*atraktorjev*) ali stabilnega stanja danega dinamičnega sistema, kar pomeni vrednosti spremenljivk, ki se ne bodo več spreminjale s časom. Celo preprost dinamični sistem lahko pogosto izkazuje

⁸² Diferencialna enačba je enačba, ki omogoča spremljanje zveznih sprememb, npr. pospeševanje padajočega predmeta.

⁸³ Diferenčna enačba je enačba, ki omogoča modeliranje razvoja danega sistema in so pri tem spremembe diskretne, npr. spreminjanje biološke populacije, pri kateri je časovna spremenljivka npr. eno leto.

⁸⁴ Teorija kaosa opisuje obnašanje nelinearnih, dinamičnih, nepredvidljivih sistemov v matematiki in fiziki, ki se pod določenimi pogoji obnašajo na prividno nepredvidljiv način (pojem kaosa po navadi označuje nered in slučajnosti).

naključnost, popolnoma nepredvidljiva/nenapovedljiva obnašanja, ki jih imenujemo *kaos*.

7.1 Definicija stanja sistema in prostor stanja

Če slika pove več kot 1000 besed, potem je v primeru dinamičnih sistemov teorija vsake slike vredna desetkrat toliko kot matematični zapis. Za smotrno uporabo *teorije dinamičnih sistemov* se v praksi uporabljajo geometrijske predstave kot pomoč za prikaz, kako se sistem konceptualno spreminja. Preprosta slika je slika stanj, ki je prečkana v sistemu skozi čas, to je sistemska *trajektorija*⁸⁵ *skozi prostor stanja*. Vsaka dimenzija v prostoru stanja ustreza eni spremenljivki sistema in vsaka točka v prostoru ustreza eni od možnih stanj sistema. Za sisteme z eno zvezno spremenljivko je prostor stanja vrsta vrednosti v eni dimenziji (npr. frekvenca tona, višina osebe, sprožilno razmerje nevrona, velikost populacije vrste v njenem okolju). Trajektorija čistega tona je regularna oscilacija med dvema vrednostnima (če dodamo k sliki še dimenzijo časa: drugo dimenzijo – dobimo sinusoido).

Če se sedaj pomaknemo v dvodimenzionalen prostor stanja, to najbolj očitno naredimo tako, da upoštevamo še eno dodatno spremenljivko. Vizualizirajmo si na primer odraščanje otroka z grafikonom zveze teže in višine. Vsak teden vrišemo točko (na abscisi odmerimo težo in na ordinati višino). Ko te točke povežemo med seboj skozi prostor teža–višina, dobimo trajektorijo odraščanja tega otroka v zveznem času. Drugi način gibanja po dvodimenzionalnem prostoru stanja pa je proučevanje iste spremenljivke za dva različna posameznika (ali nevrona, populaciji, enoti mreže, ipd.). Z dodajanjem dimenzij prehajamo v višjeredne dimenzionalne prostore, katerih dimenzionalnost teoretično navzgor ni omejena.

7.1.1 Prostor stanja

Ena od temeljnih teoretičnih idej modeliranja dinamičnih sistemov je ideja *prostora stanja* (angl. state space), ki je prostor, v katerem se ponazarja gibanje dinamičnega sistema in je splošno večdimenzionalen. Model prostora stanja opisuje fizikalni sistem kot n diferencialnih (ali diferenčnih) enačb prvega reda, namesto *ene* enačbe n -tega reda. Prostor stanja dinamičnih sistemov je geometričen način razmišljanja o

⁸⁵ Trajektorija je zvezna krivulja, če je sistem definiran v realnem zveznem času ali zaporedje točk v diskretnem času.

možnih stanjih, v katerih se sistem lahko nahaja. Prostor stanja, ki ima mnogo različnih dimenzij, ima tudi prav toliko prostostnih stopenj sistema. Vsako stanje dinamičnega sistema bo povzročilo, da bo sistem imel partikularno vrednost za vsako dimenzijo, in tako bomo lahko enoznačno definirali stanje sistema na osnovi partikularnih nizov koordinat systemskega prostora stanja. Stanje sistema je množica spremenljivk, ki vsebujejo vse tiste informacije o preteklosti sistema, ki so potrebne, da lahko določimo njegovo obnašanje v prihodnosti. Metode, ki temeljijo na opisu sistema s spremenljivkami stanja, se uporabljajo predvsem pri obravnavanju kompleksnih dinamičnih sistemov. Gibanje pomeni sleherno spremembo vrednosti spremenljivk sistema v času. Pravil za izbor spremenljivk stanja kompleksnih sistemov ni. Izbrati jih moramo na osnovi izkušenj in intuicije, upoštevajoč cilje in smotre. Če hočemo na nek sistem vplivati ali oceniti prihodnje obnašanje, moramo poznati vrednosti njegovih spremenljivk in zakon njihove povezanosti. Informacije o stanju sistema so nujne za določene ukrepe, vendar pod pogojem, da nam zares nudijo potrebno količino informacij. Pri kompleksnih sistemih težko govorimo o potrebnih in zadostnih spremenljivkah, ki določajo obnašanje sistema. Po navadi se zadovoljimo z izkušnjami in intuicijo.

Dimenzija prostora stanja je enaka številu neodvisnih spremenljivk, ki definirajo sam prostor stanja. Neodvisne spremenljivke stanja imenujemo koordinate stanja sistemov. Za želeno stanje pravimo, da je tisto, v katerem sistem doseže v svojem delovanju najboljše rezultate.

Spremenljivke stanja so množica med seboj neodvisnih količin, potrebnih za opis obnašanja sistema:

- stanje sistema je vrednost spremenljivk stanja v določenem času;
- prostor stanja je prostor, v katerem se ponazarja gibanje dinamičnega sistema in je splošno večdimenzionalen;
- enačbe stanja podajajo zvezo med vhodnimi vrednostmi, stanjem sistema in izhodnimi vrednostmi sistema in
- dinamični sistem je matematični koncept, definiran s šestimi parametri:

$$\Sigma = (T, U, X, Y, F, G) \quad (7.1)$$

$T \Rightarrow$ časovno urejena množica realnih števil;

$U \Rightarrow$ množica vhodnih vrednosti;

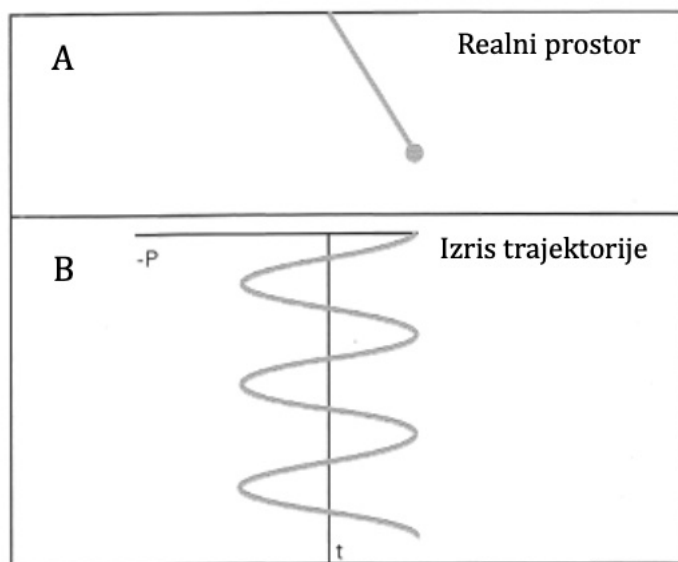
$X \Rightarrow$ množica stanj;

$Y \Rightarrow$ množica izhodnih vrednosti;

$F \Rightarrow$ funkcija prehajanja stanj, ki preslika dovoljeno množico vhodov v množico stanj, definirana kot $F: T \times T \times U \times X \Rightarrow X$ in jo imenujemo enačba stanja;

$G \Rightarrow$ funkcija izhodne preslikave, ki preslika stanje sistema v izhodne vrednosti: $G: T \times X \Rightarrow Y$ in predstavlja enačbo izhoda.

Ponazorimo to na primeru idealiziranega (matematičnega) nihala, kjer ne upoštevamo ne mase in ne izgub energije. Takšno nihalo ima dve dimenziji, *kot odmika* in kotno *bitrost gibanja*. Tako lahko katerokoli stanje nihala zapišemo s parom števil, ki predstavljata točko v dvodimenzionalnem prostoru. Glej sliko 7.1 A.



Slika 7.1 A in 7.1 B: Trajektorija skozi prostor stanja idealiziranega nihajočega nihala. Položaj nihala določa njegov kotni pomik iz vertikalne linije (pozitiven – desno, negativen – levo).

Časovna os je vertikalna os navzdol.

Vir: Povzeto po Bermudez, 2010

Če sedaj dodamo še eno dimenzijo k prostoru stanja, ki predstavlja čas, pričnemo razmišljati o razvoju nihanja v smislu trajektorije skozi prostor čas, ki si jo lahko preprosto ponazorimo kot niz točk v večdimenzionalnem prostoru, kar prikazuje slika 7.1 B.

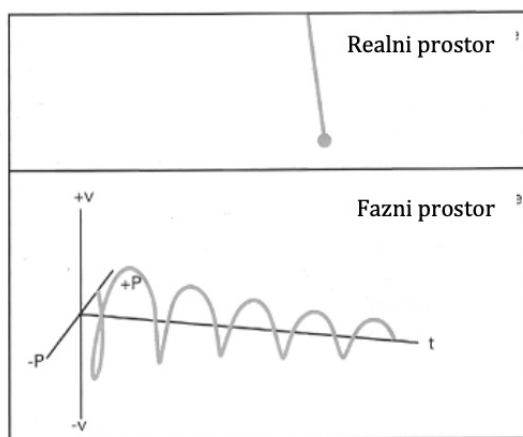
Ena od osnovnih nalog modeliranja dinamičnih sistemov je tako zapisati enačbo, ki bo opisovala razvoj sistema; to pomeni, da bo opisovala različne možne trajektorije, ki jih sistem lahko zavzame v prostoru stanja, odvisno od začetnih pogojev. V primeru zgoraj opisanega idealiziranega nihala (matematično nihalo) lahko zapišemo diferencialno enačbo, ki opisuje nihanje matematičnega nihala v obliki:

$$\frac{d^2\varphi}{dt^2} + \frac{g}{l} \sin \varphi = 0 \quad (7.2)$$

Če to diferencialno enačbo poenostavimo in rešimo, pa dobimo enostavno rešitev:

$$p = A \sin(t) \quad (7.3)$$

Trajektorija in odgovarjajoče enačbe postanejo neprimerno bolj zamotane, če odpravimo izhodiščno idealizacijo in vpeljemo še več spremenljivk, npr. upoštevamo trenje oz. izgube zaradi trenja in vpeljemo hitrost. V tem primeru lahko prostor stanja ponazorimo s sliko 7.2.



Slika 7.2: Prostor stanja nihajočega nihala v tridimenzionalnem prostoru stanja.

Vir: Povzeto po Bermudez, 2010

Če nihalo upoštevamo kot fizikalno nihalo, skratka upoštevamo vse spremenljivke, lahko zapišemo enačbo v obliki splošne diferencialne enačbe nihala:

$$J \frac{d^2 \varphi}{dt^2} + mgl \sin \varphi = 0 \quad (7.4)$$

m je masa telesa, l dolžina nihala, $J = m l^2$ vztrajnostni moment, φ kot zasuka od ravnovesne lege, g težni pospešek.

7.1.2 Osnove teorije nihanj

Osnova študij dinamičnih sistemov je splošna teorija nihanj. Elemente dinamičnih sistemov lahko obravnavamo kot homogene ali kot nehomogene, kar vpliva predvsem na njihove nihajne odzive. Ni pa dovolj, da dinamično obravnavamo samo sistem sam, temveč moramo obravnavati tako tudi okolico, kot npr. zunanje obremenitve, ki so po navadi časovno odvisen proces in se lahko spreminjajo tako po smeri delovanja, velikosti, kot tudi po mestu vprijetja. To splošno teorijo izražamo v obliki gibalnih enačb, po navadi nelinearnih in nehomogenih diferencialnih enačb višjega reda.

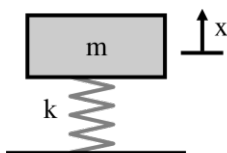
Vsak element sistema, kakor tudi sistem kot celota, predstavlja nihalni sistem, ki je sestavljen iz delov, ki upoštevajo potencialno energijo (simuliramo jih kot vzmeti), delov, ki upoštevajo kinetično energijo (masa ali vztrajnost), in delov, ki povzročajo izgubo energije (dušilci). Nihanje sistema povzroča prehod med oblikami energije, tj. potencialno in kinetično. Če upoštevamo dušenje, pride do izgub energije, ki jo moramo nadomestiti od zunaj. Pri translacijskem gibanju predstavljajo pomiki linearne razdalje, pri rotacijskem gibanju pa so pomiki definirani kot kotna gibanja. (Aberšek, 1993)

Osnove analize nihanj najlažje obravnavamo, če celoten sistem ponazorimo z enostavnim modelom *masa – vzmet – dušilec*, oz. s sistemom diferencialnih enačb, ki jih za najbolj splošen model vsiljenega nihanja z dušenjem zapišemo v matrični obliki:

$$M \ddot{x} + C \dot{x} + K x = F(t) \quad (7.5)$$

kjer je M – matrika mas, C – matrika dušenja in K – matrika vzmetnih konstant.

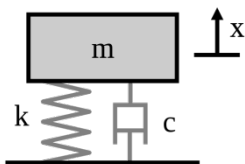
Ločimo nevsiljeno in vsiljeno nihanje. Za začetek si oglejmo osnovno teorijo na primeru ene prostostne stopnje. Nevsiljeno nihanje brez dušenja kot najbolj enostaven primer lahko ponazorimo s spodnjim modelom oz. zapišemo z navadno diferencialno enačbo



Slika 7.3: Model masa – vzmet.

$$m \ddot{x} + k x = 0 \quad (7.6)$$

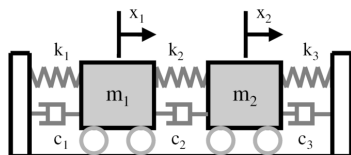
Če modelu dodamo še dušenje, kar kaže spodnja slika, se glasi diferencialna enačba za nevsiljeno dušeno nihanje takole:



Slika 7.4: Model masa – vzmet – dušilec.

$$m \ddot{x} + c \dot{x} + k x = 0 \quad (7.7)$$

Kadar moramo obravnavati več prostostnih stopenj, dobimo sistem diferencialnih enačb (za vsako prostostno stopnjo eno). Na sliki 7.5 je prikazan model z dvema prostostnima stopnjama, ki ga zapišemo z matrično obliko diferencialnih enačb:



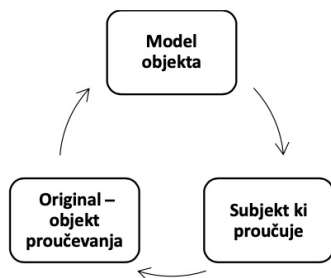
Slika 7.5: Model za dve prostostni stopnji.

$$\begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{Bmatrix} + \begin{bmatrix} C_1 + C_2 & -C_2 \\ -C_2 & C_2 + C_3 \end{bmatrix} \begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} + \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 + k_3 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix} \quad (7.8)$$

Za primere sistemov z več prostostnimi stopnjami pa zapišemo sistem diferencialnih enačb podobno kot v zgornjem primeru. (Aberšek, 1993)

7.2 Modeliranje dinamičnih sistemov

Miselne modele lahko uporabljamo le, kadar imamo opravka s preprostimi stvarmi oz. sistemi, pri kompleksnih sistemih pa nismo več sposobni razumeti dinamike medsebojnega delovanja in spreminjanja vhodno-izhodnih parametrov (razlog: omejena zmogljivost predelave informacij na mentalni ravni). Najpreprostejše modeliranje je dejanje, ko naše izkušnje z določenim procesom opišemo v nekem od obstoječih jezikov v okviru določene teorije (naše izkušnje postanejo dostopne tudi drugim). Shematski prikaz subjekta v procesu modeliranja je prikazan na spodnji sliki 7.6.



Slika 7.6: Subjekt v procesu modeliranja.

Odnos opazovalca (tistega, ki opazuje) do objekta proučevanja je bistvenega pomena v spoznavni metodi. Model je predstavitev ali reprezentacija dejstev o realnem pojavu v okviru neke teorije. V principu ločimo dve vrsti primerjanja modela in originala:

- *funkcionalnega*, značilnega za teorijo črne škatle oziroma vhodno-izhodno obnašanje;
- *strukturnega*, kjer želimo doumeti globlje razloge za določeno obnašanja sistema.

Subjekt s svojimi izkušnjami z originalom S gradi na koncu (poenostavljen) sistem elementov in njihovih zvez kot mentalni model, ki jih želi realizirati s pomočjo formalnih metod v model. Poskuša torej vzpostaviti določeno povezavo med originalom in modelom. Relacija med modelom, sistemom in opazovalcem je proces učenja subjekta v procesu preverjanja veljavnosti rezultata modela glede na obnašanje originala.

Naloga modeliranja je, da s pomočjo modela razloži znana dejstva, ter predvidi ali spoznavno napove in interpretira spoznavne dogodke nekega problema v okviru določene teorije. Model je najpogosteje poenostavljena in idealizirana podoba, ki zajema vse pomembne karakteristike in funkcionalne odvisnosti realnega sistema.

Proces modeliranja nekega sistema bi lahko opisali kot postopek iskanja poenostavljenega sistema, ki pa opisuje še vedno dovolj natančno obnašanje sistema v nekem formalnem jeziku. Modeliranje je širok spekter metod in postopkov poenostavljanja in iskanja analogij pri opisovanju sistema.

7.2.1 Odnos med modelom in originalom

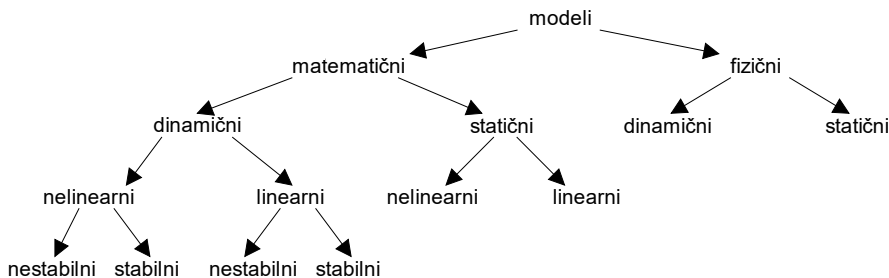
Model vedno odraža določene lastnosti originala. Odnos modela in originala je lahko homomorfen⁸⁶ ali izomorfen⁸⁷. *Homomorfni modeli* predstavljajo poenostavljanje v odnosu do originala, saj model opisuje le bistvene lastnosti originala. *Izomorfnost* med

⁸⁶ Homomorfnost predstavlja delno podobnost med modelom in originalom (strukturna podobnost).

⁸⁷ Izomorfnost predstavlja, da imajo sistemi iste vhodne in izhodne veličine in se ob tem tudi obnašajo enako. Ni pa nujno, da imajo isto fizično naravo elementov ali procesov (funkcionalna podobnost, npr. živosrebrni in digitalni termometer).

modelom in originalom pa pomeni, da se model in original enako funkcionalno obnašata na relaciji vhod—izhod ne glede na njuno bistvo.

Pri modeliranju pogosto naletimo na ovire, npr. nepoznavanje notranjosti procesa (dostopno nam je le opazovanje njegovega obnašanja). Ali nam sam sistem ne dovoli, da bi videli njegovo notranjost, ali pa je število spremenljivk in njihovih povezav ogromno, in jih ne moremo vseh opisati. Tam, kjer uporabljamo metodo črne škatle, prevladujejo predvsem kvalitativne verbalne teorije (predvsem v družboslovju). Kjer pa so procesi, ki jih opisujemo na podlagi naravnih zakonov, bela škatla, prevladujejo kvantitativne formalne teorije (npr. na področju naravoslovja in tehnike). Na sliki 7.7 je prikazana klasifikacija modelov.



Slika 7.7: Klasifikacija modelov.

7.2.2 Matematični model

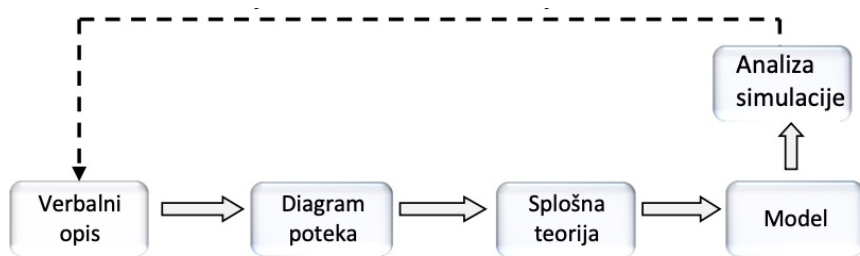
Osredotočil se bomo na matematične modele, ker so najbolj uporabni za predstavljeni primer. Matematični model je abstraktna ponazoritev določenega sistema in se uporablja za njegovo proučevanje. V večji meri so homomorfni realnemu sistemu. S pomočjo matematičnega modela najpogosteje analiziramo obnašanje sistema, da se lahko odločamo o načinih ukrepanja, ob tem pa se izognemo najpogosteje dragim in včasih tudi neizvedljivim eksperimentom. Matematični model uporabimo, če za opis realnega sistema uporabimo kar formalni jezik. Na njegovi osnovi lahko izvedemo zaključke o karakteristikah in obnašanju sistema.

Nivo modeliranja je lahko kvalitativen in pripada *teoriji množic*. Modeli tega tipa določajo dopustne vhodno–izhodne relacije in zakone njunih preslikav. Z njimi lahko ponazorimo strukturo sistema, tj. razpored elementov in njihov odnos v sistemu. So natančni, a ponazarjajo le najpomembnejše lastnosti. Prav zato so abstraktni in prikazujejo le trenutno pomembna stanja. Na nižjem nivoju kvantitativne narave pa najdemo *modeliranje v prostoru stanj* (dinamika sistemov). Temu sledijo na analitičnem nivoju *diferencialne enačbe*, ki imajo velik pomen pri študiju dinamičnih lastnosti sistemov. Uporabljamo jih, ko želimo dokazati natančno optimalnost upravljanja sistema in njegovo stabilnost. Na linearnem nivoju pa so »najožji« modeli – modeli, ki uporabljajo *linearno algebro*. Z njimi lahko proučujemo sisteme le v določenem nominalnem območju. Analiziramo stacionarne in nedinamične lastnosti sistemov, njihova najbolj tipična metoda pa je linearno programiranje.

Pri proučevanju naravnih zakonitosti nekega procesa ločimo dve fazi:

- *analiza* $\Rightarrow$ vzpostavljajo se osnovne povezave s fizikalnimi lastnosti, določajo se posamezne lastnosti procesa;
- *sinтеza* $\Rightarrow$ na podlagi poznavanja posameznih lastnosti procesa in intuicije raziskovalca se vzpostavlja sistemski pristop.

METODOLOGIJA SISTEMSKE DINAMIKE: RAČUNALNIŠKA SIMULACIJA



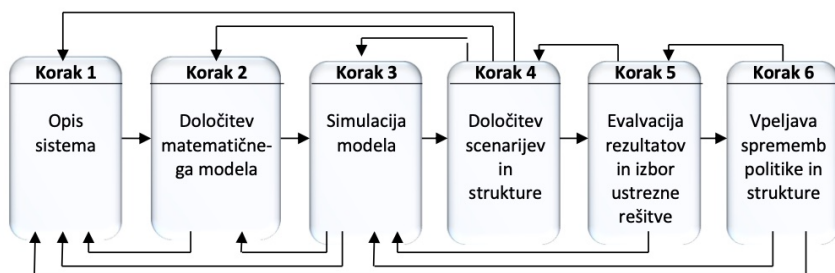
Slika 7.8: Potek modeliranja računalniških simulacij.

Analitično rešitev diferencialnih enačb, ki opisujejo nek dinamičen sistem, lahko poiščemo le v najbolj enostavnih primerih. V bolj kompleksnih sistemih (zapleteni sistemi diferencialnih enačb) pa moramo za njihovo reševanje uporabiti numerične metode. Teh metod je veliko, ena od njih je tudi računalniška simulacija.

Celoten proces modeliranja računalniških simulacij na sliki 7.8 bi lahko razdelili na:

1. *definiranje problema* – opredelimo raven in cilj modeliranja, obsega obravnavanega sistema in interakcije z okoljem;
2. *določitev spremenljivke, povratno znančnih zvez, interakcij med spremenljivkami in med deli obravnavanega sistema* – te zveze lahko ilustriramo z diagramom poteka (algoritmom), ki definira dinamiko in vzročno posledične odnose sistema;
3. *upoštevanje splošne teorije – analiza problema v splošnejšem okviru* – vzpostavljanje zveze med konkretnim problemom in konkretnimi rešitvami s splošnega vidika in
4. *matematično detajliranje posameznih delov (zapisov) z ustreznimi enačbami modela, primernimi za izbrani simulacijski jezik.*

Verbalnemu opisu problema sledi algoritem z jasnimi in razumljivimi enačbami. Nato pa zapis enačb v enem od simulacijskih jezikov. (Kljajič, 2002)



Slika 7.9: Postopek reševanja problemov.

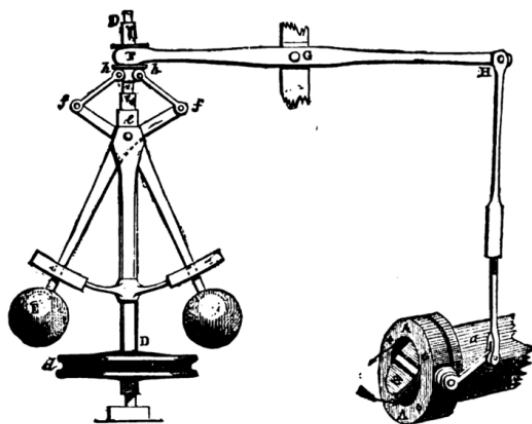
7.3 Teorija dinamičnih sistemov in kognitivna znanost

Osnovno izhodišče teorije dinamičnih sistemov (TDS) v kognitivni znanosti je, da za razlago kognicije niso najboljše ne predstave informacij, kakor tudi ne računanje in algoritmi. Skratka, kako naj razmišljamo o duhu, možganih in obnašanju, če ne smemo uporabljati ne računanja in ne predstav (ne simbolnih in ne mrežnih

sistemov). Mnogi bi rekli, da je to nemogoče, v bistvu pa ima TDS ogromno zelo uspešnih orodij, ki nam pri tem lahko pomagajo. V nadaljevanju si oglejmo nekatere od njih.

Kako in zakaj bi bili dinamični sistemi pomembni za kognitivno znanost, lahko hitro ugotovimo, če si ogledamo stališče filozofa Toma van Gelderja, enega od najzgodnejših predlagateljev hipoteze dinamičnih sistemov. (van Gelder, 1991, van Gelder, Port, 1995). Van Gelder je predstavil dva načina razmišljanja na primeru inženirskega problema, ki je bil izredno pomemben za začetek industrijske revolucije. Eden od teh dveh načinov reševanja zastavljenega problema je bil strukturno zelo podoben informacijsko procesnemu modelu razmišljanja o tem, kako duh rešuje zastavljene probleme. Drugi način pa je podoben realnemu (inženirskemu) reševanju problema, ki razkriva moč modela dinamičnih sistemov.

Van Gelderjevo osnovno izhodišče je, da kognitivni znanstvenik z metodo povratnega inženirstva duha ugotavlja, kako je duh konfiguriran za reševanje nekega konkretnega problema. Kognitivni znanstvenik teži k uporabi orodij povratnega inženirstva na poseben način in to tako, da predpostavlja, da je duh stroj za procesiranje informacij. Van Gelder pa je vzporedno pokazal tudi, da to ni niti edini niti najboljši način reševanja takšnih problemov, kar je nazorno pokazal na enostavnem primeru Wattovega centrifugalnega regulatorja, ki je sprožil industrijsko revolucijo.



Slika 7.10: Wattov centrifugalni regulator in dušilni ventil.

Centrifugalni regulator je regulator pare, katerega bistveni del je vztrajnik z dvema krogla, ki se s povečevanjem vrtenja vretena dvigujeta in s tem zapirata regulacijski ventil tako, da se dotok pare, posledično pa tudi hitrost vrtenja vretena, zmanjšuje. Problem centrifugalnega regulatorja, to je kako regulirati dovod pare za delovanje batnega stroja, je jasen. Toda kako ga rešiti? Van Gelder je pokazal, da obstaja več možnih rešitev. Oglejmo si dve. (van Gelder, 1991, 1995, 1999)

7.3.1 Prva rešitev – simbolno računski centrifugalni regulator

Nalogo reguliranja hitrosti vztrajnika je van Gelder razdelil v vrsto podnalog in predpostavil, da se vsaka od teh podnalog izvaja ločeno. Na osnovi tega je izdelal algoritem za reševanje problema regulacije kot niz korakov reševanja podnalog. Ta algoritem ima okvirno naslednjo obliko:

1. Meri hitrost vztrajnika.
2. Primerjaj dejansko hitrost S_1 z želeno hitrostjo S_2 .
3. Če je $S_1 = S_2$ se vrni na začetek – korak 1.
4. Če $S_1 \neq S_2$ potem;
 - a izmeri trenutni tlak pare;
 - b izračunaj zahtevano spremembo tlaka pare;
 - c izračunaj nastavitev dušilke tako, da bo dosežena zelena sprememba.
5. Izvedi nastavitev dušilke.
6. Vrni se na začetek – korak 1.

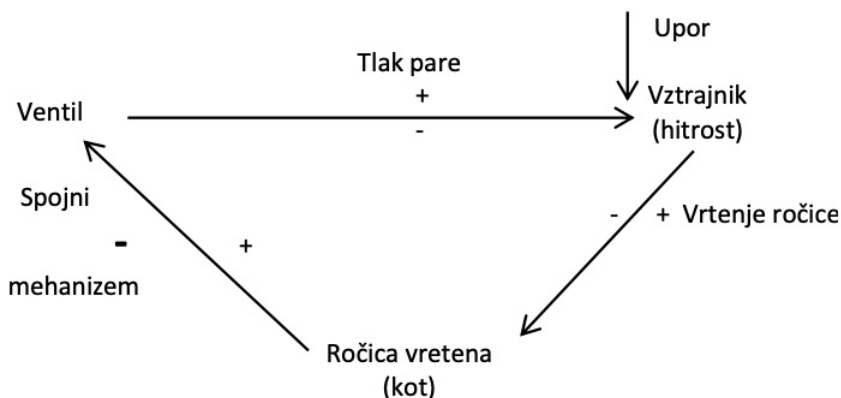
Takšen simbolni računski regulator ima nekatere znane značilnosti. Te so naslednje:

- *predstavljenost* (angl. representation) – model ne more delovati, če si na nek način ne predstavljamo hitrosti vztrajnika, tlaka pare in stanja dušilnega ventila;
- *simbolna izračunljivost* – za algoritem je pomembna primerjava, transformacija in manipulacija hitrosti, tlaka pare ipd.;
- *sekvencialnost* – model deluje diskretno (zaporedno), korak za korakom;
- *razčlenljivost* – to pomeni, da izvajajo operacije polavtomatsko posamični podsistemi in da je vsak podsistem odgovoren za določeno podnalog, npr. podsistem merjenja hitrosti za merjenje hitrosti, podsistem merjenja tlaka za merjenje tlaka itd.

Ko vse te značilnosti združimo, vidimo, da je simbolni računski regulator aplikacija nekaterih osnovnih principov kognitivne znanosti. Principov, ki jih kognitivni znanstveniki uporabljajo za to, da razumejo delovanje duha. In ne na koncu, vse skupaj deluje tudi kot zelo naraven način reševanja zastavljenega problema. Osnovno dejstvo pri vsem tem pa je, da je Watt o problemu razmišljal na popolnoma drugačen način.

7.3.2 Druga rešitev – fizikalni dinamični centrifugalni regulator

James Watt je pri razvoju svojega regulatorja seveda uporabljal popolnoma druge rešitve, torej znane principe, ki so jih pred njim uporabljali že pri mlinih na veter, in prav nobena od obstoječih rešitev ni imela nobene značilnosti simbolnega računskega regulatorja. Kot je pokazal van Gelder, so bile te rešitve izvedene kot dinamični (neračunski, nesimbolni) sistemi ob uporabi orodij teorije dinamičnih sistemov. Shematsko je princip delovanja prikazan na sliki 7.11.



Slika 7.11: Povratna zanka delovanja centrifugalnega regulatorja.

Vir: Povzeto po Bermudez, 2010

Vztrajnik na sliki 7.10, ki je sestavljen iz dveh krogel, je v spodnjem položaju. Dviguje se tako, da vztrajnik rotira vreteno s hitrostjo, ki jo določa hitrost vztrajnika. Ko hitrost rotacije vretena narašča, centrifugalna sila povzroča dvigovanje krogel. Če se hitrost zmanjšuje, se krogli spuščata. Wattova ključna ideja je bila, da je neposredno povezal ročice, na katerih so bile krogle obešene, z dušilnim ventilom parnega stroja. Ventil se odpira, ko se ročici spuščata, in se zapira, ko se ročici dvigujeta. To je

splošna rešitev, ki dopušča regulatorju, da regulira hitrost s skoraj trenutnim kompenziranjem prevelike ali premajhne hitrosti. Slika 7.11 prikazuje povratno zanko. Van Gelder je na osnovi te sheme poudaril štiri zelo pomembne značilnosti centrifugalnega regulatorja:

1. *Dinamični (neračunski) sistem*: Najboljši način za razumevanje Wattovega regulatorja dosežemo z uporabo orodij TDS. Relativno enostavno je napisati diferencialno enačbo, ki bo določala spremembo kota ročice kot funkcije hitrosti stroja. Sistem je tipičen dinamični sistem, zato ima tudi enačba majhno število spremenljivk. Zapis odgovarjajoče diferencialne enačbe lahko izpeljemo iz drugega Newtonovega zakona gibanja in se glasi:

$$\begin{aligned}\frac{d\varphi}{d\tau} &= \psi \\ \frac{d\varphi}{d\tau} &= c^2 \Omega^2 \sin \varphi \cos \varphi - \frac{g}{l} \sin \varphi - \frac{b}{m} \psi \\ \frac{d\Omega}{dt} &= \frac{1}{l} (\mu \cos \varphi - F)\end{aligned}\tag{7.9}$$

$\varphi \in (0, \pi/2)$ pomeni kot ročice regulatorja (z vertikalno osjo), $\Omega \in [0, \infty)$ je kotna hitrost vztrajnika, τ je čas, $\psi = d\varphi/d\tau$, g je težnostni pospešek, l je dolžina ročice, m je masa vsake krogle vztrajnika, $c > 0$ je prestavno razmerje, $b > 0$ je konstantno torni razmerje, F je ekvivalentna torna sila in $\mu > 0$ je torni koeficient. (Pontryagin, 1962) Z vpeljavo novih spremenljivk in poenostavitvijo pa lahko enačbe 7.9 zapišemo v poenostavljeni obliki (Sotomayor, et al., 2007) takole:

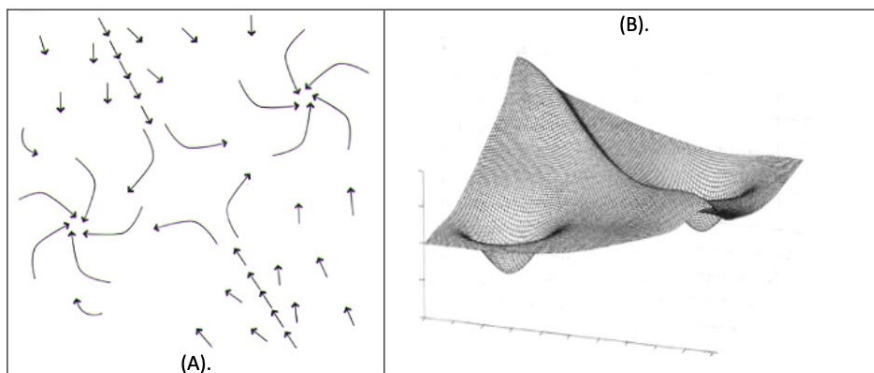
$$f(x, \mu) = (y, z^x \sin x) \cos x - \sin x - \varepsilon y, \alpha(\cos x - \beta) \tag{7.10}$$

Iz zapisa je razvidno, da smo dinamični sistem zapisali z enostavno diferencialno enačbo prvega reda z majhnim številom spremenljivk.

2. *Časovna sekvencnost*: Regulator je odvisen od časovne usklajenosti. Deluje zato, ker sprememba hitrosti vztrajnika skoraj istočasno sledi spremembam kota ročice. Diferencialna enačba, ki krmili razvoj sistema, sledi časovnim povezavam med hitrostjo vztrajnika in kotom ročice.
3. *Sklopjenost*: Regulator deluje zaradi neodvisnosti med kotom ročice regulatorja, dušilnega ventila in hitrostjo vztrajnika. Kot ročice je parameter, ki ureja hitrost

vztrajnika. Vendar pa istočasno, po istem zakonu, je hitrost vztrajnika tudi parameter, ki ureja kot ročice. Sistem kot celota je to, kar teoretiki dinamičnih sistemov (DS) imenujejo *sklopljeni sistem*, definiran s povratno zanko.

4. *Dinamika atraktorja*: Za katerokoli hitrost parnega stroja vedno obstaja ravnovesni kot ročice, kot, ki omogoča, da stroj nemoteno deluje pri dani hitrosti. O tem ravnovesju kota ročice kot atraktorju lahko razmišljamo kot o točki v prostoru stanja, v katero lahko različne trajektorije konvergirajo, kot je prikazano na sliki 7.12.



Slika 7.12: Prikaz konvergenčnih točk (stabilnih stanj sistema).

Mnogi dinamični sistemi lahko imajo vrsto konvergenčnih točk (stabilnih stanj sistema), v katere se stekajo atraktorji, in vrsto različnih trajektorij, označenih s puščicami (A). To lahko prikažemo tudi na drugačen način, kot energijsko pokrajino (B). Takšni so nelinearni dinamični sistemi. (Bermudez, 2010)

Kot smo omenili, lahko Wattov centrifugalni regulator opisujemo tudi z orodji TDS. Centrifugalni regulator je sklopljeni sistem, ki izkazuje enostavno verzijo dinamike atraktorja. V nasprotju z računskim modelom regulatorja ne zahteva nobenih predstav, računanja ali dekompozicije na podsisteme. In na koncu, lahko poudarimo še to, da regulator deluje v realnem času. Nastavitve se izvajajo skoraj v trenutku, natančno tako, kot se zahteva. Vse to bi težko trdili za računski regulator iz prvega primera. (van Gelder, 1995)

Vse lepo in prav, pokazali in dokazali smo vlogo in pomen dinamičnih sistemov, ugotovili, da dinamični sistemi na neprimerno bolj jasen in natančnejši način razložijo, kaj se dogaja med regulacijo in kakšne so posledice te regulacije. Vendar pa, kaj ima vse to za opraviti z duhom? Saj je popolnoma očitno, da Wattov parni stroj zagotovo ni kognitivni sistem. Ali pač?

Ključna točka, ki so jo izpostavili tako van Gelder kot tudi zagovorniki DS, je, da pri razlagi delovanja duha lahko uporabimo ista osnovna orodja kot pri razlagi Wattovega regulatorja. Vendar pa ni vse samo v razlagi. Teoretiki DS menijo, da razlaga deluje prav zato, ker sledi osnovnim konstrukcijskim principom duha. Ne menijo zgolj, da je duh dinamičen sistem, temveč tudi, da ko opazujemo zveze med organizmom in okolico, vidimo, da je to sklopljen sistem. Kompleksnost sklopljenega sistema organizem – okolje pa je takšna, da lahko njegovo vedenje opisujemo kot funkcijo malega števila spremenljivk.

7.3.3 Kritika prve rešitve

Predstavljena hipoteza dinamičnih sistemov je v letih svojega nastanka doživljala tudi vrsto kritik, omenimo le eno. Chris Eliasmith je kritiko hipoteze predstavil v svojem članku *Computational and Dynamical Model of Mind* (1997) in jo podkrepil s trditvijo, da Wattov centrifugalni regulator ne zadovoljuje zahtevanih pogojev. Na kratko povzemimo njegove argumente: kritični pregled kognitivne dinamične hipoteze kaže na vrsto težav v zvezi s konceptom kognicije. Najpomembnejši teoretični problem, ki bi ga morali rešiti, bi lahko povzeli z enostavnim vprašanjem: Kakšna povezava obstaja med centrifugalnim regulatorjem in naravo kognicije? Z drugimi besedami, kako lahko predpostavljamo, da so možgani podobni regulatorju? Ne obstaja nikakršna očitna razlaga, da lahko centrifugalni regulator ali njegove lastnosti na kakršenkoli način ponazarjajo kompleksnost človeških možganov. Teoretično morda obstajajo trije očitni načini, s katerimi bi lahko regulator povezali s človeško kognicijo, in to so:

- regulator bi lahko imel vlogo, primerljivo s simbolnim *Turingovim strojem*;
- lahko bi bil temeljni element za modeliranje kognicije, kot je npr. *nevron*;
- regulator bi lahko odigral *analogno vlogo* pri motivaciji možnega novega koncepta kognicije.

Tabela 7.1: Primerjava Turingovega stroja in centrifugalnega regulatorja.

Turingov stroj	Centrifugalni regulator
matematični objekt	fizični objekt
neskončen	končen
formalno dokazano, da je striktno enak drugim članom istega razreda	analogija ali primer
predstavitveni	nepredstavitveni

Očitno je, da je van Gelder dajal prednost prvi možnosti. Vendar pa centrifugalni regulator zelo verjetno ni Turingov stroj, kar je lepo razvidno iz tabele 7.1. (Eliasmith, 1997)

Kakorkoli, raziskave so se nadaljevale. Potrdile so hipotezo, da je TDS dobro in primerno orodje za modeliranje kognitivnih procesov in vsega, kar je s tem povezano.

7.4 Hipoteze dinamičnih sistemov v kognitivni znanosti

Teorija dinamičnih sistemov je v kognitivno znanost ponovno vnesla idejo o poenoteni teoriji, s katero bi lahko razložili kognicijo. Ta teorija opisuje spreminjanje kompleksnih sistemov skozi čas in tako vsak sistem, ki se časovno spreminja, razume kot dinamični sistem. Kar odlikuje pristop TDS je hipoteza, da je mogoče kognitivne sisteme preučevati z orodji dinamičnega modeliranja. Vhodni podatki ne določajo enoznačno notranjega stanja, s pomočjo katerega opisujemo zunanje stanje, ampak je TDS povezana s pristopom utelešene kognicije. Namesto kognicije, ki se ukvarja z abstraktnimi nalogami, je poudarek na telesu, ki potrebuje kognicijo za izvrševanje svojih elementarnih funkcij. Biološko telo je umeščeno v okolje in se v njem giba in deluje. Kognicija se rodi iz medsebojnega delovanja telesa in sveta. S tega vidika je kognicija odvisna od izkustev, ki izvirajo iz tega, da imamo telo z določenimi zaznavnimi in gibalnimi sposobnostmi. Te sposobnosti so povezane z vsemi ostalimi vidiki življenja.

Beer je optimističen glede vloge TDS v kognitivni znanosti in takole podpira svoj optimizem:

»Glede na to, da dinamični pristop priskrbi običajen jezik za kognicijo, za nevrofiziološke procese, ki kognicijo podpirajo, za nekognitivno človeško vedenje in za prilagoditveno vedenje preprostejših živali, se zdi obetaven za vzpostavitev poenotenega teoretičnega okvirja tako za

kognitivno znanost kot tudi za razumevanje pojava kognicije v razvoju in evoluciji.« (Beer, 2000: 97)

TDS se v dveh pomembnih vidikih razlikuje od simbolnega in konekcionističnega pristopa. Za razliko od obeh, kjer pri razlagah iščemo mehanizme, se TDS ponovno vrača k deduktivno-nomološkemu modelu razlage. V okviru tega modela je nek pojav razložen, če pokažemo, kako je mogoče opis izpeljati iz množice zakonov in začetnih pogojev. Radikalna različica TDS in utelešene kognicije zavrača reprezentacijo. Vendar se postavlja vprašanje, ali taka teorija še lahko razloži višje kognitivne procese kot npr. logično sklepanje. Clark (1997) se zato zavzema za pristop utelešene kognicije, ki ohranja reprezentacije pri višjih kognitivnih procesih.

Dinamične hipoteze v kogniciji identificirajo različne raziskovalne paradigme, ki uporabljajo matematične opise dinamičnih sistemov za razumevanje kognitivne funkcije. Vsi ti modeli so navdihnjeni z raziskavami v nevroznanosti zadnjih šestdeset let, ko je bila objavljena dinamična enačba, na osnovi katere je bil izdelan model delovanja posamičnega nevrona, imenovan Hodgkins in Huxleyev model. (Hodgkins, Huxley, 1952) Istega leta je Ashby predlagal, da bi lahko za celoten kognitivni sistem izdelali ustrezen dinamični model, vendar je bila ta ideja precej pred časom, saj tedaj ni bilo niti primernih matematičnih metod niti ustreznih računalnikov. (Ashby, 1952) V zadnjih desetletjih je konekcionistično gibanje (Rumelhart, McClelland, 1986) to idejo ponovno obudilo. Danes, šestdeset let kasneje, razmišljamo o projektu Modri možgani, ko bi po predvidevanju lahko z ogromnimi vlaganji v naslednjih 10 letih izdelali računalnik (ali sistem računalnikov), s katerimi bi bilo moč simulirati delovanje umetnih možganov.⁸⁸

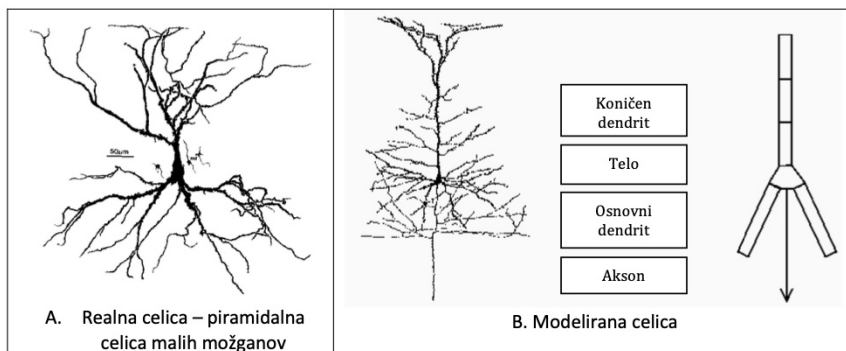
Dinamični prikaz kognicije je prav tako ozko povezan z idejo o utelešenju duha in v okolje umeščeni človekovi kogniciji, vse odkar se poudarja povezanost med obnašanjem in nevroprocesi ter kognitivnimi procesi na eni strani in psihologijo in dogajanje v okolju na drugi strani. Najpomembnejša sorodnost, skupna vsem tem področjem, pa je dimenzija časa, saj v realnem svetu dinamika enega sistema vpliva na časovne odzive drugega, oba pa morata biti koordinirana v realnem času (npr. ples na glasbo – zvoke iz okolice s sluhom zaznavamo, temu pa moramo z ustrežno motoriko telesa slediti v realnem času). Zaradi te povezave med svetom, telesom in

⁸⁸ Glej poglavje 7.7 oz. <http://bluebrain.epfl.ch>.

duhom, je metoda diferencialnih enačb uporabna za dogodke na vseh nivojih analize skozi celotno časovno skalo.

7.4.1 Matematične zveze

Kako modeliramo realni nevron je prikazano na sliki 7.13.



Slika 7.13: Primerjava realne (slika A) in modelirane celice – nevrona (slika B).

Kot lahko vidimo na sliki 7.13 A dejanska celica ni ravno podobna tistim, ki jih srečamo v različnih umetnih nevronskih mrežah. S slike 7.13 A je razvidno, da obstaja ekstenzivno drevo dendritov s piramidalnimi dendriti na koncih (živčni vršički) in osnovnimi dendriti ob dnu. Ostali dendriti ustvarjajo sinaptične povezave v različnih točkah strukture in omogočajo nevrottransmisijo, ki odpira kanale, ki omogočajo ionom pretok iz celice v celico in s tem ustvarjajo postsinaptični potencial. Ti dendriti igrajo izjemno pomembno vlogo v procesiranju informacij med celicami piramidalne oblike. Telesa celice vsebujejo napetost, ki jo povzročajo kalijevi in natrijevi kanali, podobni velikim *nevrinom* (*aksonom*), kot sta jih raziskovala Hodgkins in Huxley (1952).

Sedaj si lahko zastavimo naslednje vprašanje: *Kako bi lahko realistično modelirali tako zahtevne strukture?* Splošen odgovor bi lahko bil zelo enostaven: Potrpežljivo, po delih, košček za košček. Simbolično je to prikazano na sliki 7.13 B. Sodobni simulacijski programi nam omogočajo, da modeliramo neprimerno bolj kompleksne strukture, npr. Purkinjevo celico malih možganov, ki ima 4550 komponent in 8021 aktivnih vodnikov – kanalov. (Beeman, 2005, Bower, Beeman, 1998, 2007).

Matematični modeli na osnovi dinamičnih sistemov so plod različnih bioloških in fizikalnih odkritij. Vsebujejo ogromno biokemije, molekularne biologije in drugih temeljnih naravoslovnih ved, zato se ne bomo spuščali v podrobnosti, poskušali bomo problem dojeti na splošni ravni. Zato izpostavimo le dve osnovni izhodišči:

- *Prvo* je ideja nevronske mreže, katere temelj je bilo odkritje Hodgkins in Huxleyeve enačbe. Lahko bi ugotovili, da je vsaka celica ali celična skupina mreža, ki jo lahko zapišemo z naslednjo enačbo:

$$C_m \frac{dV_m}{dt} = (\sum_i (E_i - V_m) G_i). \quad (7.11)$$

Podrobneje jo bomo analizirali v nadaljevanju.

- *Drugo izhodišče* pa je klasična enačba enostavnega oscilatorja, kot je npr. nihalo. Očitno je, da imajo naše okončine mnoge značilnosti nihala. Ugotovljeno je bilo, da je lahko gibanje nihala dober približek za mnoge gibe človeških okončin. Skoraj identičen sistem (njegova pomanjkljivost je le gibanje po neki krivulji) lahko zapišemo z enačbo sistema masa vzmet:⁸⁹

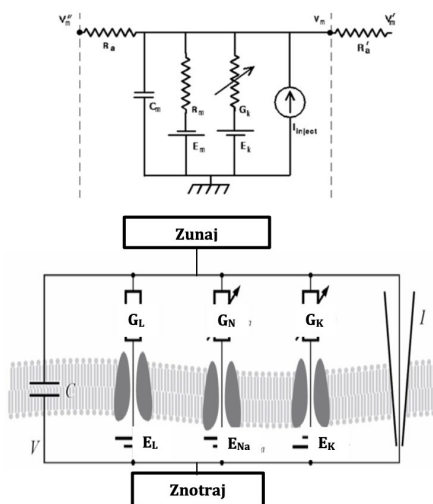
$$m(d^2x)dt^2 + d(dx/dt) + k(x-x_0) = 0 \quad (7.12)$$

Zgornja enačba določa enostavno harmonično nihanje mase m , časovnega pospeška $(d^2x)/dt^2$, dušenja d , hitrosti (dx/dt) in vzmetne konstante k , proporcionalno od nevtralne pozicije mase x_0 .

7.4.2 Matematični model nevrona: Hodgkin-Huxleyjev model

Znanstveni model je nekaj, kar obravnava ciljni pojav T in ga predstavi v obliki algoritma ali funkcije S , ki je lahko implementirana v fizikalnem sistemu, računalniškem programu, matematični enačbi ipd. Prvi model nevrona sta v matematični obliki zapisala Hodgkin in Huxley leta 1952. Od takrat je znan kot Hodgkin-Huxleyev model in je prinesel revolucijo na področje kognitivne znanosti in umetne inteligence.

⁸⁹ Za podrobnosti glej poglavje 7.1.2.



Slika 7.14: Shematski prikaz Hodgkin-Huxleyjevega modela – krožni diagram generičnega predelka.

Hodgkin-Huxleyjev model je matematični model, ki opisuje, kako se akcijski potencial v nevronih spodbudi in širi. Pri tem upošteva akcijski potencial kot svoj ciljni fenomen in ga predstavi v obliki matematične enačbe. Osnovne komponente so simbolično prikazane na sliki 7.14.

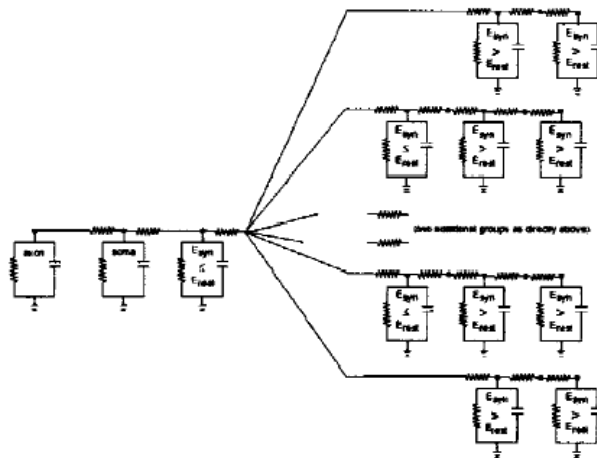
Pri tem je:

- V_m – membranski potencial v točki znotraj predelka;
- V'_m in V''_m – potenciali v različnih predelkih;
- R_a in R'_a – razlika v potencialu vzdolž aksialnega upora, ki lahko proizvede tok za vstop ali izstop iz predelka;
- C_m – kapacitivnost lipidne dvoslojne celične stene;
- R_m – nelinearni električni prevodnik – ionski kanal (angl. voltage-gated ion channel);
- G_k – linearni prevodni, pretočni kanal (angl. leak channel);
- I_{inject} – vir energije, ionska črpalka;

- E_m in E_k – baterije, elektrokemični gradienti, ki vodijo tok ionov. Njihovo vrednost določamo iz *Nernstovega potenciala*⁹⁰ ionov. Ione poganja gradient elektrokemijskega potenciala, zaradi dveh učinkov:
 - po eni strani zaradi njihovega naboja na ione deluje električno polje oz. gradient električnega potenciala;
 - po drugi pa zaradi difuzije ioni potujejo iz območja, kjer je njihova koncentracija višja, v območje, kjer je ta nižja. Neto tok ionov se ustavi, ko se elektrokemijska potenciala na obeh straneh membrane izenačita. To se zgodi pri neki ravnovesni vrednosti električnega potenciala, ki jo imenujemo *Nernstov potencial* in je podan z izrazom:

$$\phi = \frac{RT}{ZF} \ln \frac{c_1}{c_2} \quad (7.13)$$

Pri tem je R splošna plinska konstanta, T absolutna temperatura, Z valenca iona, $F = N_A e_0$ Faradayeva konstanta, c_1 in c_2 pa koncentraciji ionov na eni in drugi strani membrane.



Slika 7.15: Model na sliki je sestavljen iz 12 sinaptičnih predelkov.

⁹⁰ Nernstov potencial je električni potencial, ki se vzpostavi zaradi različne koncentracije ionov na eni in drugi strani membrane (npr. celične membrane).

Model na sliki 7.14 pa lahko povezujemo v bolj kompleksne strukture. Takšen primer je prikazan na sliki 7.15.

Hodgkin-Huxleyjev model je torej sistem diferencialnih enačb, ki simulira električne karakteristike vzbujenih celic, kot so nevroni in srčne mišice. Če model na sliki 7.15 upoštevajoč teorijo električnih tokokrogov zapišemo v matematični obliki, dobimo diferencialno enačbo (Hodgkin, Huxley, 1952, De Schutter, Bower, 1994, Beeman, 2005):

$$\frac{dV_m}{dt} = - \frac{1}{C_m} (\sum_i I_i) \quad (7.14)$$

Če enačbo nekoliko preuredimo, dobimo:

$$C_m \frac{dV_m}{dt} = - (\sum_i I_i) \quad (7.15)$$

In upoštevajmo da je posamičen ionski tok:

$$I_i(V_m, t) = (V_m - E_i) G_k \quad (7.16)$$

Potem dobimo:

$$C_m \frac{dV_m}{dt} = (\sum_i (E_i - V_m) G_k) \quad (7.17)$$

ozi.

$$C_m \frac{dV_m}{dt} = \frac{(E_m - V_m)}{R_m} + (\sum_k (E_k - V_m) G_k) + \frac{(V_m' - V_m)}{R_a'} + \frac{(V_m'' - V_m)}{R_a''} + I_{inject} \quad (7.18)$$

V tej enačbi lahko možnost prevajanja od aksona do nevrotansmitorja sledi linearnemu porastu in eksponentialnemu upadu, zato jih mnogokrat modeliramo s t. i. a funkcijo s posamično časovno konstanto:

$$G_k(t) = g_{max} \frac{t}{\tau} e^{1-t/\tau} \quad (7.19)$$

Včasih pa lahko uporabimo tudi dvojno eksponentialno funkcijo:

$$G_k(t) = \frac{Ag_{max}}{\tau_1 - \tau_2} \left(e^{-t/\tau_1} - e^{-t/\tau_2} \right) \quad (7.20)$$

Prevodnost kalijevega kanala je proporcionalna četrti potenci aktivacijske spremenljivke n :

$$G_k = \bar{g}_k n^4 \quad (7.21)$$

Prevodnost natrijevega kanala pa je proporcionalna produktu tretje potence svoje aktivacijske spremenljivke m in deaktivacijske spremenljivke h , ki zmanjšuje porast membranskega potenciala:

$$G_{Na} = \bar{g}_{Na} m^3 h \quad (7.22)$$

Razmerje spremembe aktivacijske in deaktivacijske spremenljivke lahko zapišemo z diferencialno enačbo:

$$\frac{dn}{dt} = \alpha_n(V) (1 - n) - \beta_n(V) n \quad (7.23)$$

Napetostno odvisne parametre v zgornji enačbi pa zapišemo:

$$\alpha_n(V) = \frac{n_\infty(V)}{\tau_n(V)} \quad ; \quad \beta_n(V) = \frac{1 - n_\infty(V)}{\tau_n(V)} \quad (7.24)$$

Podobno lahko zapišemo tudi ostale spremenljivke.

Brez poglobljenega matematičnega znanja lahko razložimo bistvo te enačbe. Enačba prikazuje, da je celotna količina toka, ki poteka skozi membrano, sestavljena iz štirih ločenih tokov (kapacitivnega, kalijevega, natrijevega in propustnega toka). Ti tokovi pa so odvisni od številnih različnih spremenljivk, kot je prikazano zgoraj. Celotna enačba je izpeljana iz teorije električnih tokokrogov, upoštevajoč različne zakone, še posebej Coulombov in Ohmov zakon.

Zanimivo vprašanje je: Ali je ta predstavljeni model *razlagalen* ali ni? Zanimivo pri tem je, da sta avtorja sama menila, da je model *nerazlagalen*, saj prikazuje, kako se opazovane spremembe v membranskem potencialu na sploh dogodijo. Nista pa

razmišljala o tem, da bi natančno razložila, kaj se v resnici dogaja v nevronu. Vedela sta, da Na, K in ostali ioni prehajajo skozi celično membrano, zato sta to vključila v svoj model in to upoštevala s tremi faktorji (m , n in h) v enačbi. Kaj pa ti trije faktorji predstavljajo, jima je ostala skrivnost, verjetno prav zato, ker nista poznala in razumela delovanja ionskih kanalov. Šele kasneje so ugotovili, da so ti kanali skupek proteinov, ki so vstavljeni v celično membrano in jih lahko določamo z ustreznimi eksperimenti (angl. voltage clamp experiment). Pod določenimi pogoji ti odprejo pore v membrani, skozi katere lahko prehajajo ioni. Različni proteini, ki ustvarjajo te kanale, so ti »skrivnostni« faktorji, katerih Hodgkin in Huxley nista poznala. Eksperimenti na strukturah teh kanalov in pogojev, pod katerimi se odprejo, pa so omogočili, da Hodgkin-Huxleyjev model preide iz nerazlagalnega v razlagalni model oz. iz modela, *kako bi bilo možno*, v model, *kako je zares*. (Bower, Beeman, 1998)

7.4.3 Matematični model nevrone: Fitzhugh-Nagumov model

Obstaja pa še vrsta drugih, najpogostejše poenostavljenih modelov. Eden od teh je tudi *FitzHugh-Nagumov model*, poimenovan po FitzHughu, ki je predlagal leta 1961 prikazani sistem (FitzHugh, 1961) in Nagumu ki je ustvaril prikazani sistem, prototip nevrone. (Nagumo, Arimoto, Yoshizawa, 1962)

Ko zunanji stimulus I_{ext} preseže neko vrednost praga, prične sistem izkazovati karakteristiko premika v faznem prostoru, preden se spremenljivki v in w vrne v svoje mirujoče stanje. Takšno vedenje je tipično za konične generatorje (kratko povišanje membranske napetosti v) v nevronu po stimuliranju s pomočjo zunanjega vhodnega toka. Tako lahko enačbo za ta dinamični sistem zapišemo:

$$\begin{aligned}\dot{v} &= v - v^3 - w + I_{ext} \\ \tau \dot{w} &= v - a - bw\end{aligned}\tag{7.25}$$

Dinamiko tega sistema lahko enostavno opišemo z nenehnim menjavanjem med levo in desno stranjo *kubične izokline*.⁹¹

⁹¹ Izočrte (tudi izolnije, izokline) so črte, ki na zemljevidih ali diagramih povezujejo točke enakih vrednosti fizikalnih, meteoroloških ali jezikoslovnih količin. Izoklina lahko predstavlja magnetno inklinacijo ali pa tudi dinamiko populacij, kjer se izoklina nanaša na velikost populacije. Izokline se uporabljajo za reševanje navadnih diferencialnih enačb.

7.5 Primer dinamičnega modeliranja kognitivnih zmožnosti

Seveda pa nobena teorija ne obstoji brez dokazov, brez takšnega ali drugačnega potrjevanja. Poskusimo teorijo dinamičnih sistemov potrditi na konkretnih primerih. Verjetnost hipoteze kognitivnih dinamičnih sistemov seveda ne more temeljiti na podobnosti med duhom in parnim strojem, kljub temu da je ta analogija zelo sugestivna. Izvedenih je bilo kar nekaj pomembnih raziskav, tiste najbolj zanimive in najbližje zastavljenemu problemu so bile predvsem s področja motoričnih veščin in motoričnega učenja. Na osnovi teh raziskav je bilo dokazano, da je TDS izjemno uspešno orodje za razumevanje npr. kako se otrok nauči hoditi.

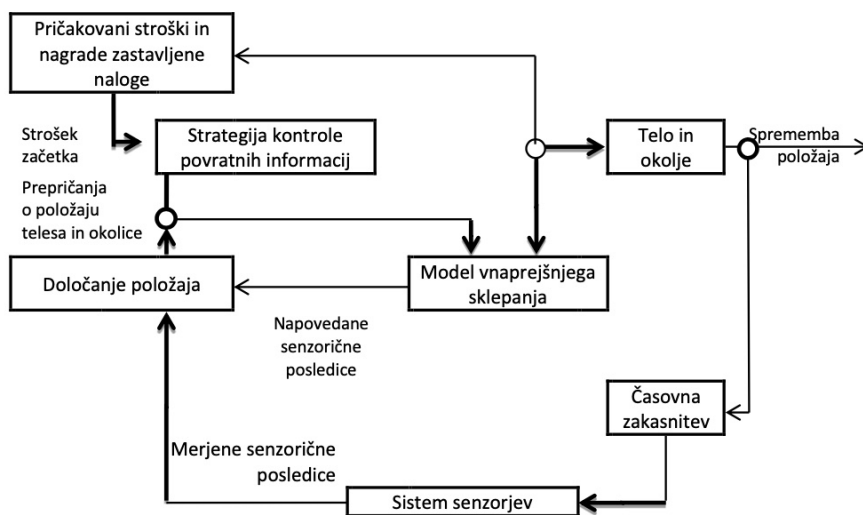
V nadaljevanju se bomo osredotočili na kognitivne dinamične modele. Pri tem moramo rešiti dve vrsti problemov:

- motorične – ponazarjanje delovanja telesa;
- kognitivne – ponazarjanje delovanja duha in mentalnih procesov.

7.5.1 Motorična kontrola: osnovni model

Ena od ključnih značilnosti modelov DS (dinamičnih sistemov) je časovna odvisnost. Dinamični model lahko zasleduje razvoj sistema skozi čas zelo podrobno, kar omogoča študije, kako se otrok uči novih veščin in si pridobiva nove sposobnosti, ali kako se gibajo prsti pri izvajanju neke motorične naloge. Oba primera sta s področja kontrole motorike in si jih bomo ogledali v nadaljevanju.

Razlaga, kako se otrok nauči hoditi, je zelo podobna primeru Wattovega centrifugalnega regulatorja. (Bermudez, 2010) Dominantni model, ki se uporablja za razumevanje, kako so gibi načrtovani in izvedeni, je računski model kontrole gibanja, ki je nekakšen ekvivalent računskemu centrifugalnemu regulatorju. Razlaga pa je lahko tudi drugačna. Model po TDS ponuja alternativo, kako so gibi organizirani in kako se gibalne veščine razvijajo.



Slika 7.16: Primer računskega modela motorične kontrole.

Računski model kontrole gibanja lahko ponazorimo s preprostim primerom. Zamislimo si gibanje pri tem, ko segamo po predmetu. Pri računskem modelu se prične načrtovanje te aktivnosti v centralnem živčnem sistemu, kjer se računa tako položaj ciljnega objekta kot tudi položaj roke. To računanje zahteva vnos podatkov, tako vizualnih kot tudi vseh drugih čutnih zaznav (npr. senzorja, ki v roki zaznava napetost mišic). Načrtovanje giba zahteva izračun trajektorije giba, od začetnega do končnega položaja. Prav tako je potrebno računati delne gibe posamičnih mišic, ki bodo vodile roko po predvideni (izračunani) trajektoriji. In na koncu: Izvedba giba zahteva računanje sprememb v gibih mišic, tako da lahko usklajujemo vidne in čutne povratne informacije. Tako dobimo večpodsistemsko računanje, ki se zdi, kot da je pisano na kožo algoritmičnemu informacijskemu procesu, ki ga prikazuje slika 7.16. Slika ponazarja računski model kontrole gibanja in predstavlja standardni informacijsko-procesni diagram, ki ustreza zgornjemu splošnemu opisu. Model vsebuje mehanizem vnaprejšnjega sklepanja, ki proizvaja predvidevanja o zaznavnih posledicah posamičnih gibov in primerjalni mehanizem, ki primerja predvidene in dejanske povratne informacije. (Shadmehr, Krakauer, 2008)

Psihologinji Esther Thelen in Linda Smith (1994, 2003) sta pokazali, da to ni niti edini, verjetno pa tudi ne najboljši način razlage. Izdelali sta model, kako prične dojenček hoditi. Izhajali sta iz tega, da hoja ni načrtovana aktivnost. Model ne uvaja

posebnega niza gibalnih ukazov. Njun model je zasnovan na ideji, da se aktivnost hoje pojavi iz kompleksnih interakcij med mišicami, telesnimi okončinami in različnimi značilnostmi okolja. Obstaja veliko povratnih zank, ki nadzirajo gibanje okončin, kot funkcija sprememb na telesu ali v okolju.

Kako se dojenčki učijo hoditi? Večina normalnih dojenčkov se prične učiti hoditi proti koncu svojega prvega leta. Najprej so sposobni stopicljanja, sposobni so narediti nekaj posameznih korakov – koračno gibanje. Tak način gibanja se konča med t. i. nekoračno fazo (angl. Non-stepping window). Gibanje se popravi, ko otrok shodi. Tradicionalna razlaga te U-trajektorije razvoja je, da je otrokovo začetno gibanje, stopicljanje, popolnoma refleksno in izgine med nekoračno fazo, ker so možgani že dovolj razviti, da ustavijo refleksne odzive, vendar pa še niso dovolj razviti, da bi stopicljajoče gibanje lahko svobodno nadzorovali. Thelen in Smith sta ta model potrdili z vrsto eksperimentalnih dokazov. (Thelen, Smith, 1994) Odkrili sta, da bi stopicljajoče gibanje lahko umetno ustvarili v nekoračni fazi tudi tako, da bi manipulirali značilnosti okolja, npr. tako, da bi dojenček stopil v toplo vodo. Ta možnost manipuliranja z dojenčkovimi stopicajočimi gibi predstavlja težave za možganski model – vse od takrat, ko se stopicanje spreminja neodvisno od tega, kako se možgani razvijajo. Rešitev leži v TDS, v definiranju ključnih parametrov pri razvoju hoje dojenčkov, parametrov kot so masa nog, trdnost mišic, gravitacija in vztrajnost. Možgani skupaj s centralnim živčnim sistemom pri tem nimajo nikakršnega privilegirane položaja. Torej dobimo obnašanje, ki bi ga v principu lahko modelirali z enačbami malega števila neodvisnih spremenljivk. Thelen in Smith sta to idejo razvili, podrobno dodelali in potrdili, vključno z vrsto eksperimentalnih študij in analiz.

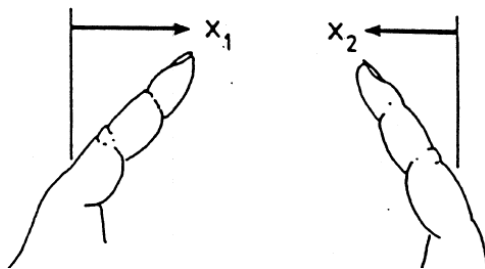
7.5.2 Motorična kontrola: Haken–Kelso–Bunz model

Podobno kot zgornji primer, je tudi naslednji s področja motorike. Oglejmo si nekoliko podrobneje model, ki je povezan z gibanjem, z motoriko prstov. Haken, Kelso in Bunz so leta 1985 izdelali matematični dinamični (HKB) model, s katerim so lahko simulirali gibanje prstov. Za izhodišče so uporabili gibalne enačbe, s katerim so ponazarjali gibanje posameznih prstov in preučevali povezave med njimi. Te gibalne enačbe imajo naslednjo obliko:

$$\ddot{x}_1 + f_1(x_1, \dot{x}_1) = I_{12}(x_1, x_2) + F_1(t) \quad (7.26)$$

$$\ddot{x}_2 + f_2(x_2, \dot{x}_2) = I_{21}(x_1, x_2) + F_2(t) \quad (7.27)$$

Leva roka na sliki 7.17 opisuje gibanje posameznega prsta z amplitudo x_1 , medtem ko desna roka opisuje gibanje x_2 , njuno za združevanje, ki se izvaja s pomočjo živčnega sistema.



Slika 7.17: Simetrični pomiki konic prstov leve in desne roke.

Za bolj jasno predstavo je za gib prstov na sliki 7.17 vpeljan poenostavljen model enačb:

$$\begin{aligned} x_1 &= r_1 \cos(\omega t + \phi_1) \\ x_2 &= r_2 \cos(\omega t + \phi_2) \end{aligned} \quad (7.28)$$

kjer je:

ω – osnovna frekvenca gibov roke;

r_1 in r_2 – amplitudi;

ϕ_1 in ϕ_2 – faze, ki so časovno odvisne.

Enačbi 7.26 in 7.27 opisujeta kompleksni sistem mehaničnega gibanja rok, ki ga povzročajo nevro mišične aktivnosti. Ker je to gibanje pretežno oscilacijsko, morata biti enačbi vsaj drugega reda, in ker ima gibanje rok več ali manj stabilno amplitudo, mora biti enačba nelinearna. Temu splošnemu zapisu ustreza kar nekaj posebnih enačb – če zanemarimo vpliv sil, je to lahko npr. van der Polova enačba:

$$\ddot{x} + \varepsilon(x^2 - r_0^2)x + a\dot{x} = 0 \quad (7.29)$$

Pri tem so:

ε in r_0 – nastavljiva, vendar skalarna parametra in

a – kontrolni parameter.

Za lažje reševanje izberemo:

$$x = A e^{i\omega t} + A^* e^{-i\omega t} \quad \text{in} \quad \omega^2 = a \quad (7.30)$$

Pri tem je A kompleksna in časovno odvisna amplituda. Če želimo najti porast amplitude s frekvenco, lahko zapišemo enačbo v nekoliko drugačni obliki:

$$\ddot{x} + \varepsilon (\dot{x}^2 - \omega_0^2 r_0^2) \dot{x} + a x = 0 \quad (7.31)$$

Sedaj pridemo do osrednjega problema, izpeljave primerne povezave med dvema makroskopskima vrednostma, t. j. amplitudama x_1 in x_2 . Najenostavnejša povezava bi bila linearna, vendar ta ne izpolni osnovnih zahtev, zato uporabimo nelinearno zvezo, ki pa še vedno ni popolnoma ustrezna. Za rešitev problema imamo dve možnosti, *zvezo s časovno zakasnitvijo* ali *časovno izpeljavo*.

Če se ne spuščamo v matematične podrobnosti izpeljave, ker to tudi ni naš namen, potem zgoraj prikazan model lahko poenostavimo, dobimo (poenostavljen) HKB model, s katerim lahko pri kognitivnem modeliranju simuliramo dinamiko človeškega telesa na primeru ritmičnega gibanja prstov.

Omenjeni avtorji so pri svojih raziskavah ugotovili, da obstaja relativna fazna zveza, fazna povezava $\emptyset$ med gibanjem prstov, ki bi bila primerna koordinacijska spremenljivka. Razlogi za to so naslednji:

- $\emptyset$ določa opazovani vzorec vedenja;
- $\emptyset$ spreminja odsek pri gibu in je samo šibko odvisen od parametrov zunaj tega giba (ni občutnega vpliva okolice nanj);
- $\emptyset$ ima zelo enostavno dinamično gibanje, pri katerem vedenjski vzorec lahko označimo kot atraktor.

Če vemo, da frekvenca oscilacije ozko sledi eksperimentom in ni odvisna od sistema samega, je lahko frekvenca kontrolni parameter. Dinamiko $\emptyset$ lahko tako določimo iz nekaj osnovnih zahtev:

1. Opazovani vzorci vedenj pri $\emptyset = 0$ in $\emptyset = \pm 180^\circ$ so modelirane fiksne točke atraktorja. Dinamika je zato samo čista sprostitvev. To je minimalna strategija, pri kateri se le opazovani tip atraktorja pojavi v modelu.
2. Model mora proizvesti opazovani vzorec relativnega faznega vedenja: bistabilnost pri nižjih frekvencah in monostabilnost pod kritično frekvenco.
3. Pojavijo se lahko samo točkovni atraktoji relativne fazne dinamike in
4. Relativne faze ciklične spremenljivke so stabilne, to pomeni, da mora, če dodamo ali odvezamemo mnogokratnik 2π , sistem ostati nespremenjen. Vsaka gibalna enačba mora biti zapisana v obliki periodične funkcije, torej sinusne ali kosinusne funkcije. Splošna gibalna enačba z vsemi temi lastnostmi se tako glasi:

$$\emptyset = a_0 + \sum_{k=1}^{\infty} \{a_k \cos(k \emptyset) + b_k \sin(k \emptyset)\} \quad (7.32)$$

Predstavljeni model ni idealen, saj izkazuje prekinitve simetrije, vendar pa je dovolj dober za prikaz in študijo osnov dinamičnih sistemov v kognitivni znanosti, predvsem s filozofskega stališča. Enačba 7.32 je splošna enačba, ki jo lahko zapišemo poenostavljeno takole:

$$\emptyset = -a \sin \emptyset - 2b \sin 2\emptyset \quad (7.33)$$

To bo za naš namen prav tako primerna oblika. Minusi v enačbi 7.33 in faktor 2 pred b zelo poenostavita celoten izračun, saj relevantno področje prostora parametra ta podaja le za pozitivna faktorja a in b , in faktor 2 dovoljuje, da je potencial $V(\phi)$ definiran brez ulomka:

$$V(\emptyset) = -a \cos \emptyset - b \cos 2\emptyset \quad (7.34)$$

Gibalno enačbo lahko sedaj še poenostavimo z uporabo normiranja, ki dodatno omejuje prostor parametra na samo en pozitiven parameter, brez spreminjanja katerihkoli drugih dinamičnih lastnosti. Tako dobimo:

$$\ddot{\theta} = -\sin \theta - 2k \sin 2\theta \quad (7.35)$$

V tej enačbi je parameter $k = a/b$ in ustreza cikličnemu periodičnemu gibanju prstov, ki so inverzni gibalnemu razmerju oz. oscilacijski frekvenci v eksperimentu. (Haken et al., 1985, Jirsa, Kelso, 2005)

Če želimo določiti, ali ta enačba predstavlja veljaven teoretični model eksperimentalnih raziskav, moramo najti fiksno točko in raziskati njeno stabilnost, kar z drugimi besedami pomeni rešiti to enačbo za pogoj $\dot{\theta} = 0$.

Če na kratko povzamemo, enačba 7.35 prikazuje, da je razmerje sprememb relativne faze enako periodični funkciji trenutne relativne faze in frekvence oscilacije. Prikazuje, kako ima, ko k narašča, sistem dve stabilni stanji (ki ustrežata faznemu in antifaznemu gibanju), kar se v nadaljevanju pretvori le v eno stabilno stanje (fazno gibanje). Pomembno je poudariti dve psihološko pomembni lastnosti tega modela. *Prva lastnost* je, da model prikazuje podatke, ne da bi kakorkoli predpostavljal notranji mehanizem preklapljanja (angl. switching mechanism). HKB model izkazuje, kako se izvaja preklop med dvema stabilnima stanjema pri nizkih frekvencah v eno stabilno stanje pri višjih frekvencah kot naravni produkt normalnih, samoorganiziranih evolucij sistemov brez povezave s kakršnokoli obliko »centralne izvedbe«. *Druga lastnost*, ki jo je Kelso izpostavil, pa je, da je model sposoben izvajati predvidevanja obnašanja, ki še niso bila opazovana pri ljudeh, v času, ko je bil model razvit, pa so bila v nadaljnjih raziskavah potrjena. Uporabljeni HKB model je, med drugim, sposoben napovedovati rezultate tudi pri simuliranju različnih motenj (npr. kaj se dogodi, če uporabimo nizek električni pulz na človeški roki in s tem začasno prekinemo koordinacijo).

7.5.3 Model možganskih aktivnosti

Kljub temu pa, tudi če je gibanje prstov ali hoja zagotovo zelo kompleksna aktivnost, vseeno ni čista kognitivna aktivnost. Ali bi lahko bil model dinamičnih sistemov bolj kognitiven? Nekaj primerov dokazuje prav to, da je bil model dinamičnih sistemov uspešno uporabljen pri študijah, kako se človek odloča. Teorija odločanja (angl. Decision Field Theory), ki sta jo razvila Busemeyer in Townsend, kaže na to, da je razlaga določenih eksperimentalnih rezultatov v vedenjski ekonomiji in psihologiji sklepanja odvisna od medsebojnega delovanja sedmih parametrov. (Busemeyer,

Townsend, 1993) Primer, ki smo si ga že ogledali izhaja iz dela Thelen in Smith ter se ponovno nanaša na dojenčkov razvoj. Predstavlja dinamični model tega, kako dojenček razume predmet. V nadaljevanju sedmega poglavja se bomo osredotočili le na najnovejši in najbolj kompleksen dinamični model za modeliranje možganskih aktivnosti.

Sodobne diagnostične tehnike nevroznanosti omogočajo tako vpogled v dinamiko človeških možganov kakor tudi študij človekovega vedenja in njegovega duha. Meritve pa vsebujejo ogromno šuma, tako da je na njihovi podlagi skoraj nemogoče določiti enačbo, ki bi dovolj natančno opisovala dinamiko živčnih aktivnosti. Nekaj fizikov je formuliralo zvezne modele za napovedovanje živčne aktivnosti z uporabo anatomije možganov. Jirsa in Haken (1996) pa sta te modele posplošila in problem zapisala v obliki:

$$u_{tt} - \alpha \Delta u = a(u, p) u_t + b(u, p, p_t), \quad \text{na } [0, +\infty) \times \Omega \quad (7.36)$$

$$u(0, \cdot) = u_0 \quad u_t(0, \cdot) = u_1. \quad (7.37)$$

$$u = 0, \quad \text{na } [0, +\infty) \partial\Omega \quad (7.38)$$

Tukaj je Ω odprto mejno področje v R^n ($n \leq 4$) z zadostno gladko mejo $\partial\Omega$. Simbol u_t označuje prvi in u_{tt} drugi parcialni odvod u kot funkcijo časovne spremenljivke t . Δ je Laplasov operator⁹² in p zvezna funkcija, s katero simuliramo zunanja vzbujanja (vhodna vzbujanja). Zaradi poenostavitve predpostavimo, da je $a = 1$.

Čeprav ima enačba 7.36 dokaj enostavno obliko, pa sta obstoj in enkratnost te rešitve še vedno odprt problem, razen če ne predpostavimo dodatnih pogojev za a , b ali u_0 in u_1 (enačbi 7.37 in 7.38). Dodatno pa obstoj posplošene rešitve ni vedno zagotovljen. (Jradeh, Bergounioux, 2008)

Enačba 7.36 spada med nelinearne dušene valovne enačbe, ki so jih uporabljali mnogi matematiki in inženirji. Osnovna valovna enačba ali tudi d'Alembertova enačba je pomembna linearna parcialna diferencialna enačba drugega reda, ki v

⁹² Laplasov operator je v vektorskem računu skalarni diferencialni operator skalarne funkcije. Je enak vsoti vseh drugih parcialnih odvodov odvisne spremenljivke.

splošnem opisuje različna valovanja, kot so zvočno valovanje, svetlobno valovanje in vodno valovanje. Valovna enačba podaja odmik $s = s(x, y, z, t)$ v valovanju kot funkcijo kraja in časa, $s = s(x, t)$ je *amplituda*, mera jakosti valovanja v določeni točki x v času t .

$$\frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} - \nabla^2 s = 0 \quad (7.39)$$

Enorazsežno obliko lahko izpeljemo, če si zamislimo upogljivo struno, napeto med dvema točkama na osi x . Velja

$$\frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} - \frac{\partial^2 s}{\partial x^2} = 0 \quad (7.40)$$

zapisano v dveh razsežnostih:

$$\frac{1}{c^2} \frac{\partial^2 s}{\partial t^2} - \left(\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2} \right) = 0 \quad (7.41)$$

V še bolj zapletenih in resničnih oblikah valovne enačbe se lahko konstanta c spreminja s frekvenco ali amplitudo valovanja. Takšne enačbe so nelinearne:

$$\frac{1}{c(s)^2} \frac{\partial^2 s}{\partial t^2} - \nabla^2 s = 0 \quad (7.42)$$

Valovna enačba se pojavlja na mnogih različnih področjih, kot so akustika, elektrika, magnetizem in dinamika tekočin in v našem primeru na področju kognitivnega modeliranja. Različne oblike valovnih enačb najdemo tudi v kvantni mehaniki in splošni teoriji relativnosti.

Obstoj in izjemnost pogojev v enačbah 7.37 in 7.38 lahko upoštevamo z uporabo različnih metod, kot so npr. teorije *polgrupe* ali tudi *semigrupe*⁹³, metode negibnih točk (metoda navadne iteracije) ali pa tudi Galerkinove metode.⁹⁴

⁹³ Polgrupa ali tudi semigrupa $S = \{a, b, \dots\}$ je v matematiki par $(S, *)$, kjer je S množica in $*$ asociativna dvočlena operacija na S : $S \times S \rightarrow S$, ki vsakemu urejenemu paru $(a, b) \in S$ priredi natanko en element $a * b \in S$.

⁹⁴ Galerkinova metoda je v matematiki področje numeričnih analiz za spreminjanje zveznih problemov, kot npr. diferencialne enačbe v diskretne probleme.

Podobno enačbo je zapisal tudi Zhijian (2003):

$$u_{tt} - \Delta u - \Delta u_t = \sum_{i=1}^n \frac{\partial}{\partial x_i} \left(\sigma_i(u_{xi}) + \beta_i(u_{xit}) \right) + F(u, u_t, \nabla u, \nabla u_t) \quad (7.43)$$

na $[0, \infty) \times \Omega$

Enačba ima enake začetne in robne pogoje kot enačbe 7.36 do 7.38. Zhijian je dokazal, da je tu problem edinstvena klasična rešitev, ki predpostavlja, da so začetni podatki takšni:

$$\left\| \sum_{i=1}^n \frac{\partial}{\partial x_i} \left(\sigma_i(u_{xi}(0)) + \beta_i(v_{xi}(0)) \right) + F(u(0), v(0), \nabla u(0), \nabla v(0)) \right\|_{H^k} < \delta$$

$$\|v(0)\|_{H^{k+1}}^2 + \|v(0)\|_{H^k} + \frac{3}{2} \|u(0)\|_{H^k} + \langle u(0), v(0) \rangle_{H^k} < \gamma \quad (7.44)$$

Velja, kjer sta δ in γ dve realni števili. V tem delu je H^k označba za Sobolev prostor in $\|\cdot\|_{H^k}$ označuje normiranje H^k .

PRIMER UPORABE MODELA MOŽGANSKIH AKTIVNOSTI

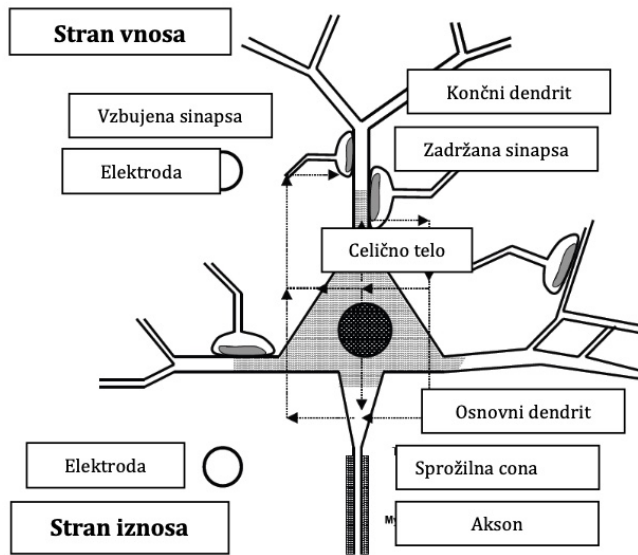
Predhodno predstavljeni model Jirsa in Hakena je uporaben tudi pri modeliranju cerebralnih aktivnosti. Kot vemo, je osnovna enota živčnega sistema nevron (glej poglavje 3). Informacije, ki se prenašajo med dvema nevronoma A in B, bi lahko opisali kot:

- Impulz (električni signal) prispe v dendrit nevrona A (kot potencialna aktivnost) in spodbudi nevrottransmitorje (kemične substance) tako, da se transformirajo v val. Po Freymanu (1982) je ta sprememba linearna operacija med nevronskimi stenami.
- Val nato doseže t. i. »sprožilno cono« (na sliki 7.18) in se ponovno spremeni nazaj v impulz. Ta pretvorbeni zakon je *sigmoidalen*.⁹⁵

⁹⁵ Sigmoidalna krivulja nastane z uporabo matematične funkcije in je zanjo značilno, da ima »S« obliko. Pogosto sigmoidalno funkcijo označujemo kot poseben primer logistične funkcije.

Iz analize pretvorbenih operacij pri sinapsah in sprožilni coni nevronov ugotovimo, da lahko izpeljemo nelinearno parcialno diferencialno enačbo, ki opisuje časovno-prostorsko obnašanje možganskih aktivnosti. Pulzne in valovne spremenljivke so klasificirane kot podspremenljivke v odvisnosti od spodbujevalnih ali zaviralnih karakterjev; spodbujevalni pulz $E(x, t)$ in zaviralni pulz $I(x, t)$, spodbujevalni val $e(x, t)$ in zaviralni val $b_i(x, t)$. Vale lahko zapišemo z naslednjo enačbo:

$$\psi_k(x, t) = \int_{\Omega} f_k(x, X) H_k(x, X, t) dX. \quad (7.45)$$



Slika 7.18: Dendritsko drevo sestavljeno iz sinaps.

Funkcije $H_k(x, X, t)$, $k = i, e$ so pretvorbeni operacijski rezultati, funkcije $f_i(x, X)$ pa odgovarjajoče prostorske zvezne porazdelitve. Ω označuje nevronske niz plošč (na površini možganov, za katere predpostavljamo, da so zvezne). Iz tega sledi, da je $\Omega \subset \mathbb{R}$. Izbrali smo 1D model geometrije možganov, kot je prikazano v Jirsa, 2004.

Predpostavimo, da ima funkcija naslednjo obliko:

$$f_k(x, X) = \frac{1}{2\sigma_k} e^{-\frac{|x-X|}{\sigma_k}} \quad (7.46)$$

Tukaj sta $k = e, i$ in σ_e in σ_i sta vzbujevalna in zaviralna vezna parametra. Če je $\sigma_i \ll 1$, so kortikularne povezave (med dvema oddaljenima nevronoma po kortikularnih vlaknih) samo vzbujevalne.

Nasprotno pa širjenje informacije vzdolž kortikularnih vlaken povzroča zakasnitve:

$$t = \frac{|x-X|}{v_j} \quad (7.47)$$

Tukaj sta v_e in v_i hitrost širjenja vzbujanja in zaviranja, tako da dobimo naslednje:

$$H_k(x, X, t) = \alpha_k E\left(X, t - \frac{|x-X|}{v_k}\right) \quad (7.48)$$

Tukaj so $k = e, i$ in a_e in a_i realne konstante. Če vstavimo enačbo 7.48 v enačbo 7.45 in predpostavimo, da je $v = v_e$ dobimo:

$$\psi_e(x, t) = a_e \int_{\Omega} f_k(x, X) E\left(X, t - \frac{|x-X|}{v}\right) dX \quad (7.49)$$

$$\begin{aligned} \psi_i(x, t) &= \int_{\Omega} f_i(x, X) H_i(x, X, t) dX \approx f_i(x, x) H_i(x, x, t) = a_i I(x, t) \\ &\quad (\text{ker } \sigma_i \ll 1) \end{aligned} \quad (7.50)$$

Vrednost pulza izračunamo iz vrednosti vala in zunanega vnosa v nevronske membrano (slika 7.24); ta dobimo tako sigmoidalno obliko:

$$\begin{aligned} E(x, t) &= S_e[\psi_e(x, t) - \psi_i(x, t) + p_e(x, t)], \\ I(x, t) &= S_i[\psi_e(x, t) - \psi_i(x, t) + p_i(x, t)]. \end{aligned} \quad (7.51)$$

Tukaj je $p_k(x, t)$ zunanji vnos. Sigmoidalna funkcija S_j se tako glasi:

$$S_j(n_j) = \frac{1}{1 + \exp(-v_j n_j)}. \quad (7.52)$$

In če je n_j dovolj majhen, lahko zgornjo enačbo preoblikujemo takole:

$$S_j(n_j) \approx n_j - \frac{v_j^3 n_j^3}{48} \quad (7.53)$$

Ker je $v_e/v_i \approx 100$, upoštevamo v zgornji enačbi 7.53 le linearni del in tako dobimo:

$$\psi_e(x, t) \approx \frac{a_i v_i}{4 + a_i v_i} \left(\psi_e(x, t) + p_i(x, t) \right). \quad (7.54)$$

Končno dobimo enačbo za dinamiko vzbujanja sinaptičnih aktivnosti:

$$\psi_k(x, t) = a_e \int_{\Omega} f_e(x, X) S_e \left[\tilde{\rho} \psi_e \left(X, t - \frac{|x-X|}{v} \right) + p \left(X, t - \frac{|x-X|}{v} \right) \right] dX. \quad (7.55)$$

Tukaj je

$$p(x, t) = p_e(x, t) - \frac{a_i v_i}{4 + a_i v_i} p_i(x, t), \quad (7.56)$$

In

$$\tilde{\rho} = 1 - \frac{a_i v_i}{4 + a_i v_i}. \quad (7.57)$$

Predpostavimo, da je nevronski niz plošč $\Omega := (0, L)$ enodimenziionalen, kar pomeni, da povežemo vse nevrone plošče z linijo. Linija je enačba, ki sledi $\psi_e(0, t) = \psi_e(L, t) = 0$, kot v (Kiguradze, 2001), $f_e(x, X)$ pa je eksponentialna funkcija, ki je odvisna samo od razdalje med dvema nevronskima ploščama. Zato velja enačba 7.55.

Ker lahko, oz. moramo podvomiti v vsako trditev, ki ni na nek način dokazana, poskusimo *matematično dokazati* pravilnost zgornje enačbe. Vzemimo, da je:

$$\rho(X, T) = a_e S_e \left(\tilde{\rho} \psi(X, T) + p(X, T) \right) \quad (7.58)$$

in

$$\mu: L^1(\mathbb{R} \times \mathbb{R}) \rightarrow \mathbb{R}$$

$$\varphi \rightarrow \langle \mu, \varphi \rangle = \int_{\mathbb{R}} h(x) \varphi\left(x, \frac{|x|}{v}\right) dx \quad (7.59)$$

z

$$h(x) = 1_{(0,L)} e^{-\frac{|x|}{\sigma}}. \quad (7.60)$$

Ker je $1_{(0,L)}$ je indicatrix⁹⁶ funkcija $(0, T)$. Tako dobimo:

$$\mu * \rho(x, t) = \int_{\mathbb{R}} h(X) g\left(X, \frac{|X|}{v}\right) dX$$

$$z \quad g(X, T) = \rho(x - X, t - T)$$

$$\mu * \rho(x, t) = \int_{\mathbb{R}} h(X) \rho(x - X, t - T) dX = \int_{\Omega} f_e(x, X) \rho\left(X, t - \frac{|x-X|}{v}\right) dX \quad (7.61)$$

in

$$\mu * \rho(x, t) = \psi_e(x, t) \quad (7.62)$$

Razširjena Fourierova transformacija pravi:

$$\widehat{\psi_e} = \widehat{\mu} \widehat{\rho} \quad (7.63)$$

Tako lahko dokažemo, da je

$$\widehat{\mu} = T_f, f \in L^1_{loc}(\mathbb{R} \times \mathbb{R}). \quad \text{Naj bo } \varphi \in D(\mathbb{R} \times \mathbb{R}). \quad (7.64)$$

$$\langle \widehat{\mu}, \varphi \rangle = \langle \mu, \widehat{\varphi} \rangle = \int_{\mathbb{R}} h(X) \widehat{\varphi}\left(X, \frac{|X|}{v}\right) dX$$

$$= \int_{\mathbb{R}} h(X) \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x, t) e^{-2i\pi\left(xX + t\frac{|X|}{v}\right)} dx dt \right) dX$$

⁹⁶ Indicatrix je sferična oblika krivulje. Najbolj splošni indicatrix je Dupin's indicatrix.

$$\begin{aligned}
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(X) e^{-2i\pi\left(xX+t\frac{|X|}{v}\right)} dX \right) \varphi(x, t) dx dt \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, t) \varphi(x, t) dx dt.
\end{aligned} \tag{7.65}$$

Pri tem je

$$f(x, t) = \frac{1}{2\sigma_e} \int_{\Omega} e^{-\frac{|X|}{\sigma}} e^{-2i\pi\left(xX+t\frac{|X|}{v}\right)} dX \tag{7.66}$$

Tako dobimo naslednje:

$$\widehat{\psi}_e(x, t) = f(x, t) \widehat{\rho}(x, t) \quad \forall (x, t) \in \mathbb{R} \times \mathbb{R} \tag{7.67}$$

kot

$$\int_0^{+\infty} e^{-(a+ia)\zeta} d\zeta = \frac{1}{a+ia}. \tag{7.68}$$

Predpostavimo, da je $\omega_0 = v/\sigma$, in dobimo:

$$\begin{aligned}
f(x, t) &= \frac{1}{2} \left[\frac{1}{\frac{1}{\sigma_e} + i\left(\frac{t}{v} - x\right)} + \frac{1}{\frac{1}{\sigma_e} + i\left(\frac{t}{v} + x\right)} \right] \\
&= \frac{1}{2} \left[\frac{\omega_0}{\omega_0 + i(t - xv)} + \frac{\omega_0}{\omega_0 + i(t + xv)} \right] \\
&= \left[\frac{\omega_0^2 + i\omega_0 t}{(\omega_0 + it)^2 + x^2 v^2} \right] \in L_{loc}^1(\mathbb{R} \times \mathbb{R})
\end{aligned} \tag{7.69}$$

Enačba 7.67 tako dobi naslednjo obliko:

$$\widehat{\psi}_e(x, t) = \left[\frac{\omega_0^2 + i\omega_0 t}{(\omega_0 + it)^2 + x^2 v^2} \right] \widehat{\rho}(x, t) \tag{7.70}$$

Če razvijemo enačbo 7.67 dobimo:

$$\begin{aligned}
\omega_0^2 \widehat{\psi}_e(x, t) + 2\omega_0 \frac{\partial \widehat{\psi}_e}{\partial t}(x, t) + \frac{\partial^2 \widehat{\psi}_e}{\partial t^2}(x, t) - v^2 \frac{\partial^2 \widehat{\psi}_e}{\partial x^2}(x, t) &= \omega_0^2 + \\
\omega_0 \frac{\partial}{\partial t} \rho(x, t).
\end{aligned} \tag{7.71}$$

Ob uporabi inverzne Fourierove transformacije formalno dobimo naslednjo izpeljano parcialno enačbo:

$$\frac{\partial^2 \psi_e}{\partial t^2} + \left(\omega_0^2 - v^2 \frac{\partial^2}{\partial x^2} \right) \psi_e + 2\omega_0 \frac{\partial \psi_e}{\partial t} = \left(\omega_0^2 + \omega_0 \frac{\partial}{\partial t} \right) \rho \quad (7.72)$$

Tukaj je:

$$\rho(x, t) = a_e S_e \left(\tilde{\rho} \psi_e(x, t) + p(x, t) \right). \quad (7.73)$$

Če razvijemo enačbo 7.72 in seštejemo enake izraze dobimo:

$$\begin{aligned} \frac{\partial^2 \psi_e}{\partial t^2} - v^2 \frac{\partial^2 \psi_e}{\partial x^2} = & \left(a_e \omega_0 \tilde{\rho} S'_e (\tilde{\rho} \psi_e + p) - 2\omega_0 \right) \frac{\partial \psi_e}{\partial t} - \\ & \omega_0^2 \psi_e + a_e \omega_0^2 S_e (\tilde{\rho} \psi_e + p) + a_e \omega_0 S'_e (\tilde{\rho} \psi_e + p) p_t. \end{aligned} \quad (7.74)$$

Če v enačbo vstavimo $\psi_e(t, 0) = \psi_e(t, L) = 0$ in Couchijeve začetne pogoje, lahko na podlagi predhodnih izpeljav dobimo naslednje:

Teorem 1: Predpostavka $p \in H^2([0, T], L^\infty(0, L))$. Enačba 7.74 $\approx \psi_e(t, 0) = \psi_e(t, L) = 0$ in Cauchijevim začetnim pogojem $(\psi_e, 0 \ \psi_e, 1) \in H^1_0(0, L) \times L^2(0, L)$ ima eno šibko rešitev, in to je $H^2([0, T], L^2(0, L)) \cap H^1([0, T], H^1_0(0, L))$.

Dokaz: Enačba 7.74 je partikularni primer enačbe 7.36 z

$$a(u, p) = a_e \omega_0 \tilde{\rho} S'_e (\tilde{\rho} u - p) - 2\omega_0, \quad (7.75)$$

in

$$b(u, p, p_t) = -\omega_0^2 S_e (\tilde{\rho} u + p) + a_e \omega_0 S'_e (\tilde{\rho} u + p). \quad (7.76)$$

Če predpostavimo, da je $H^2([0, T], L^\infty(0, L))$, je očitno, da a in b funkcije potrjujejo teorem o šibki rešitvi. S tem pa je podan tudi matematični dokaz izračuna. (Jradeh, Bergounioux, 2008)

7.6 Ocenjevanje modelov dinamičnih sistemov

Iz predstavljenih eksperimentalnih rezultatov in modelov je jasno razvidno, da je TDS močno orodje tako za preučevanje duha in obnašanja kot tudi za njun razvoj. Razlage modelov teorije dinamičnih sistemov delujejo na prvi pogled tako bolj kompleksno, a istočasno tudi bolj enostavno od standardnih informacijsko procesnih razlag, ki so v kognitivni znanosti še vedno prevladujoče. Delujejo bolj kompleksno predvsem zato, ker moramo pri njih upoštevati mnogo širši spekter faktorjev, ki jih pred tem kognitivni znanstveniki niso upoštevali. Ob tem se oddaljujemo od razlag o enem samem informacijsko procesnem mehanizmu (kar je bilo za kognitivne znanstvenike zelo udobno) in ga zamenjujemo s časovno odvisnim kompleksnim sistemom s komaj opaznimi medsebojnimi soodvisnostmi. Ob tem vpeljujemo še dodatno splošno časovno odvisnost. Istočasno pa ideja deluje tudi kot bolj enostavna, predvsem zato, ker ne zahteva ne predstav in ne »računanja«.

Modeli dinamičnih sistemov bi lahko bili radikalna alternativa nekaterim osnovnim idejam kognitivne znanosti, še posebej ideji, da kognicija nujno potrebuje računanje in informacijsko procesiranje. Nekateri začetniki, omenimo le van Gelderja, so pred dinamične sisteme zato postavili zelo stroge zahteve, s katerimi so jih poskušali ločiti od vseh ostalih možnosti. Te zahteve mnogokrat ignorirajo najbolj osnovne in pomembne značilnosti kognitivne znanosti. Kot smo že večkrat izpostavili, nihče, tudi teoretiki DS, ne morejo zanikati, da je kognitivna znanost tako *interdisciplinarna* kot tudi *mnogonivojska*. (Bermudez, 2010) Duh je preveč kompleksen, že če ga opazujemo samo kot pojav, da bi ga lahko popolnoma razumeli s pomočjo ene same znanstvene discipline ali na posameznem nivoju, kar seveda velja tudi za hipotezo dinamičnih sistemov. Vse teorije in modeli nam zagotovo podajajo poglobljeno, vendar le delno in enostransko sliko celotne problematike. To je tako, kot da bi na svet gledali npr. skozi rdeč ali zelen filter svetlobe, vemo pa, da je svetloba (bela svetloba, ki jo seva sonce) sestavljena iz mavričnega spektra, ki ustvarja svojo celovito predstavo o resničnem svetu. Resnično delo kognitivne znanosti je zagotovo, da vključi vse te delne poglede v poenoteno in celovito sliko duha.

Prepad, ki ga je van Gelder ustvaril med računskim regulatorjem in Wattovim (dinamičnim – neračunskim) regulatorjem je izrazit in provokativen, vendar pa zagotovo ne more biti neposredno prenesen iz inženirske prakse v kognitivno znanost, kar smo dokazovali že v predhodnem poglavju na primeru Eliasmithove

argumentacije kritike hipoteze dinamičnih sistemov. (Eliasmith, 1997) Pri reševanju zastavljenega inženirskega problema centrifugalnega regulatorja lahko uporabimo en model za drugim, *nikakor pa ne obeh hkrati*. Nič takega pa ne velja, ko preidemo na kognitivno področje, saj so modeli dinamičnih sistemov popolnoma združljivi z informacijsko procesnimi modeli kognicije. Modeli dinamičnih sistemov obratujejo na visokem nivoju abstraktnosti. Kognitivnim znanstvenikom omogočajo, da se oddaljijo od detajlov informacijsko procesnih mehanizmov, zato da lahko študirajo, kako se sistem kot celota razvija skozi čas. Vendar pa, tudi če imamo model razvoja kognitivnega sistema skozi čas, še vedno potrebujemo znanje in razumevanje o tem, zakaj se sistem razvija tako, kot se pač razvija.

S podrobnejšo analizo stanja bi ta prastari vzorec med starim in novim, med uveljavljenim in tistim, ki šele išče svoj prostor in svoje priznanje, lahko zaznali tudi na področju kognitivne znanosti. Staro, je uveljavljeno, vsesplošno priznано, ima svoj položaj, ne želi sprejemati novosti in je sebi samozadostno. Nasproti temu se pojavi nekaj novega. To novo je (in seveda tudi mora biti) brezkompromisno, brezobzirno, samozaverovano in hoče pokazati to, da je bilo vse pred njimi narobe, hoče počistiti s preteklostjo. Vendar, tako kot svet ni črno-bel, tudi ni črno-bele znanosti. Prej ko slej mora vsako »novo« odrasti in *konstruktivno* iskati svoj prostor pod soncem. Tako kot sta se »moralci« zblížati simbolna in mrežna kognitivna znanost, da so nastali hibridni modeli, npr. ACT-R, tako bo v prihodnosti potreben trezen in zdrav razmislek, kje vključiti modele dinamičnih sistemov, izkoristiti njihove prednosti in dobre strani. Tako bomo razvili modele, s katerimi bomo lahko pričeli svet zaznavati globalno, sistemsko, in ne le kot do sedaj, ko smo se ubadali z omejenimi, parcialnimi, »laboratorijskimi« problemi. Ne bomo sposobni modelirati le majhne nevronske mreže, temveč se bomo lahko lotili tudi modeliranja celotnih možganov, kar je danes, zaradi tehnoloških nezmožnosti (računalniška moč in spominske kapacitete ali celo napačnega pristopa) – še nemogoče. Morda pa to ni le tehnološki problem, morda pa smo se problema lotili napačno, z napačnim orodjem in na napačni strani.

Tako kot je narava (ali njen stvaritelj, odvisno od tega, v kaj pač verjamemo) uporabila evolucijo, da je iz preprostih sistemov ustvarila človeka, tako se bo tudi človek moral naučiti (če se bo želel postaviti v vlogo stvaritelja), da vsak razvoj, vsaka evolucija, poleg vsega ostalega, potrebuje predvsem čas in z njim povezano zrelost.

7.7 Možnosti realizacije umetnih možganov

7.7.1 Izhodišča

Človeški možgani so brezmejno mogočen, energijsko učinkovit, samoučljiv in samopopravljiv računalnik. Če bi razumeli in bi lahko posnemali načine, kako delujejo, bi lahko naredili revolucijo v informacijski tehnologiji, medicini in družbi. Da bi nam uspelo ustvariti *umetne možgane*, bi bilo treba združiti vse naše dosedanje znanje in vse, kar smo se še sposobni naučiti o notranjem življenju možganskih molekul, celic in krogotokov. Umetni možgani so splošno uporabljen termin, s katerim se opisujejo raziskave s področja razvoja tako programske kot tudi strojne opreme, ki bi imela podobne kognitivne lastnosti, kot jih imajo živalski ali še posebej človeški možgani. Raziskave s področja *umetnih možganov* imajo v znanosti tri pomembne cilje:

1. s poskusi nevroznanstvenikov razumeti, kako človeški možgani delujejo, kar je bolj znano kot *kognitivna nevroznanost*;
2. v *filozofiji UI* z eksperimenti dokazati, da je teoretično mogoče ustvariti stroj, ki bi imel človekove kognitivne sposobnosti;
3. z vrsto dolgoročnih projektov znotraj UI ustvariti stroj z *univerzalno inteligenco* oz. ustvariti *univerzalno umetno inteligenco*. To idejo je populariziral Ray Kurzweil (2005) in jo imenoval *močna UI*, kar pomeni, da mora biti stroj enako inteligenčen kot človek.

V okviru *prvega cilja* znanstveniki uporabljajo biološke celice za to, da iz njih ustvarjajo »*nevrosfere*« (majhne skupine nevronov). S tem želijo razviti nove načine zdravljenja različnih bolezni, kot so npr. Alzheimerjeva in Parkinsonova bolezen.

Raziskave v sklopu *drugega cilja* so vezane na splošno znane argumente, kot so kitajska soba Johna Searla,⁹⁷ kritika Huberta Dreyfusa⁹⁸ ali argument Rogerja Penrosa Vladarjev novi duh.⁹⁹ Te kritike so v glavnem povezane s trditvijo, da človeške zavesti

⁹⁷ Za podrobnosti glej (Bechtel, Abrahamsen, 2002: 303–304)

⁹⁸ Dreyfusova kritika UI se nanaša na štiri osnovne predpostavke v raziskovanju UI, na biološko in psihološko ter iz nje izpeljani epistemološko in ontološko predpostavko. Dreyfus je v svoji kritiki pokazal, da ne moremo razumeti svojega lastnega obnašanja na enak način, kot razumemo objekte, npr. v fiziki ali kemiji. (Dreyfus, 1979)

⁹⁹ V knjigi Vladarjev novi duh Roger Penrose dokazuje, da so zakoni fizike neprimerni za razlago fenomena zavesti. Zato je v njej predlagal »*ново fiziko*«, ki bi zgradila most med klasično in kvantno mehaniko. Imenoval jo je *pravilna kvantna gravitacija* (angl. *correct quantum gravity*). (Penrose, 1989)

ali drugih emocionalnih »veščin« stroji pač niso sposobni producirati. Eden od odzivov na njihove argumente je, da je to absurd, saj bi lahko biološke procese v možganih simulirali s poljubno natančnostjo, tako da ne bi prišlo do nikakršnih odstopanj od naravnih procesov. Ta odziv je v svojem bistvu že zelo star, saj izhaja iz leta 1950 in ga je zapisal že Alan Turing v svojem klasičnem delu *Computing Machinery and Intelligence*.

Tretji cilj se imenuje *univerzalna umetna inteligenca* (Voss, 2006) ali *močna umetna inteligenca*. (Kurzweil, 2005) Raziskave na tem področju so se osredotočile na implementacijo umetnih možganov v konvencionalne (digitalne) računalnike ter na tej osnovi analiziranju simuliranja delovanja celotnih možganov. Pri tem je Kurzweil trdil, da bo vse to predvidoma izvedljivo do leta 2025. Nekaj podobnega trdi tudi Henry Markam, vodja *projekta modri možgani*.¹⁰⁰ Vse pogosteje pa se zastavlja vprašanje, ali so digitalni računalniki res primerni za simuliranje zveznih možganskih procesov. V nadaljevanju bomo poskušali poiskati nekatere bolj ali manj verjetne odgovore na to vprašanje, ki preizkušajo meje znanosti.

7.7.2 Modeli posnemanja bioloških možganov

Čeprav trenutno modeli neposrednega posnemanja možganov uporabljajo umetne nevronske mreže na visoko sposobnih digitalnih računalnikih (angl. high-performance computing engine), kar si bomo podrobneje ogledali v nadaljevanju, pa obstajajo tudi drugačni računalniški sistemi. (Kurzweil, 2005):

- *Bioračunalniki*, ki uporabljajo sistem bioloških molekul, kot so DNA in proteini, da z njimi izvajajo računalniške operacije, kot so ustvarjanje, procesiranje in shranjevanje podatkov. Razvoj teh računalnikov je omogočila nova interdisciplinarna znanost, imenovana *nanobiotehnologija*. To novo vedo bi lahko definirali na različne načine, v najbolj splošnem pomenu pa je to katerikoli tip tehnologije, ki uporablja biološke materiale in materiale nano velikosti. Bolj podrobna definicija določa *nanobiotehnologijo* kot ustvarjanje in proizvodnjo proteinov, ki se jih da združevati v večje funkcionalne strukture, ki so funkcionalna osnova bioračunalnikov. Na osnovi te definicije obstajajo tri vrste bioračunalnikov, ki se med seboj razlikujejo predvsem po tem, kako

¹⁰⁰ Leta 2005 so bili definirani osnovni cilji projekta umetnih človeških možganov, poimenovanega Modri možgani, ki so našli svoj domicil na: École polytechnique fédérale de Lausanne (EPFL) v Švici.

interpretirajo izhodne signale na osnovi začetnih pogojev oz. vhodnih informacij:

- *biokemični računalniki*, ki uporabljajo za doseganje računalniške funkcionalnosti biološke kemične reakcije, tj. koncentracijo določene kemikalije;
- *biomehanični računalniki*, kjer je definiran izhodni signal kot *mehanska oblika* določenih molekul ali niz molekul pod vplivom niza začetnih pogojev;
- *bioelektronski računalniki*, ki na izhodu merijo električno prevodnost. Ti računalniki so zgrajeni iz posebnih biomolekul, ki prevajajo elektriko na poseben način.

Narejenih je bilo že kar nekaj enostavnih bioračunalnikov, vendar je njihova kapaciteta še vedno zanemarljiva v primerjavi z konvencionalnimi nebiološkimi računalniki. Zagotovo pa imajo ti računalniki pred seboj svetlo prihodnost. Dokaz njihovega potenciala se skriva predvsem v trenutno najbolj dovršenem in najbolj zmogljivem bioračunalniku – v človeških možganih. Ampak s to primerjavo smo ponovno na samem začetku, saj človeške možgane pravzaprav želimo na takšen ali drugačen način poustvariti in seveda bi to bilo najlažje narediti tako, da bi jih na nek način kar »prekopirali«.

- *DNK računalniki* oz. DNK računalništvo je oblika računanja, ki namesto tradicionalnih, na silikonu temelječih računalniških tehnologij, uporablja, podobno kot v prvem primeru, za svojo osnovo biokemijo in molekularno biologijo. DNK računalništvo, ali bolje splošno – biomolekularno računalništvo, je v osnovi podobno *vzporedni obdelavi podatkov*,¹⁰¹ le da je energijsko varčnejše, načelno pa je za večino primerov tudi hitrejše in prostorsko manj potratno. DNK računalništvo pa ne prinaša v računalniško teorijo nobenih novosti. Npr., če potreben računalniški »prostor« za rešitev določenega problema z zahtevnostjo problema pri konvencionalnih računalnikih narašča eksponentialno, potem velja enako tudi na DNK računalnikih. (Beneson et. al., 2004, Lewin, 2004)

¹⁰¹ Vzporedna obdelava (angl. parallel processing) je računalniška tehnologija, ki omogoča sočasno izvajanje več izračunov.

- *Kemični računalnik*, ki je poimenovan tudi *reakcijsko-difuzijski* računalnik, je računalnik, ki temelji na poltrdih kemičnih snoveh, pri katerih so podatki določeni s spreminjanjem koncentracije kemikalij. Računanje se izvaja z naravnimi kemičnimi reakcijami. Ta računalnik je še v zgodnji fazi razvoja, vendar pa ima, podobno kot tudi predhodno predstavljeni nekonvencionalni računalniki, velik potencial za računalniško industrijo, predvsem zaradi svoje funkcionalne enostavnosti in enostavnosti potrebne izdelovalne tehnologije.

Ker so valovi potencialnega gibanja pri kemičnih računalnikih v vseh možnih smereh (v nasprotju s konvencionalnimi mikroprocesorji, kjer se gibanje »informacij« lahko izvaja le po določenih poteh) in se morajo pri tem upoštevati predpisana pravila, kot npr. pri cestnem in železniškem, ladijskem ali letalskem prometu, bi lahko bili kemični računalniki milijon krat hitrejši kot so današnji konvencionalni računalniki. Kemične računalnike bi še najlaže (ponovno) primerjali z možgani. Čeprav mikroprocesorji sicer prenašajo informacije znatno hitreje kot nevroni v možganih, pa so možgani neprimerno uspešnejši pri izvajanju določenih (večine) nalog prav zato, ker lahko istočasno obdelujejo znatno več podatkov. (Adleman, 1994)

- *Kvantni računalniki*, ki uporabljajo principe teorije kvantne mehanike in se jim bomo nekoliko bolj posvetili v nadaljevanju, saj razvojno predstavljajo trenutno najbolj verjetno alternativo sedanjim, konvencionalnim digitalnim računalnikom. Najbolj verjeten alternativni model konvencionalnim računalniškim tehnologijam bi lahko temeljil na *holografski nevronske tehnologiji* (angl. Holographic Neural technology – HneT) in na linearnem faznem *koherentnem/dekoherentnem principu*,¹⁰² ki se skriva v kvantnem procesu ob uporabi *sinaptičnega algoritma*¹⁰³ (angl. Core synaptic algorithm). Bistvo tega modela temelji na popolnoma drugačnem preračunavanju in na popolnoma drugačni strojni opreми, na *kvantnem računalništvu*. (Abninder et. al., 2006)

¹⁰² Koherentna/dekoherentna superpozicija: Interferenca je sestavljanje (superpozicija) dveh ali več koherentnih valovanj, pri čemer nastane nov valovni vzorec. Pri sestavljanju valovanj z enako fazo se valovni hrbet prvega valovanja ujema z valovnimi hrbti drugega valovanja, zato je skupna amplituda večja. Pravimo, da pride do konstruktivne (koherentne) interference. Če pa sta valovanja v protifazi, tako da je njuna fazna razlika enaka 180°, pa pride valovni hrbet prvega valovanja na valovno dolino drugega valovanja in obratno, tako da se valovanja vzajemno popolnoma oslabita. Pojav imenujemo destruktivna (dekoherentna) interferenca.

¹⁰³ Sinaptični algoritem je zelo podoben kvantno mehanski valovni enačbi.

7.7.3 Kvantno računalništvo

Kakšno je stanje na področju procesorjev, uporabljenih v sodobnih računalnikih? Ali obstajajo principi, ki bi uspešno zamenjali digitalne principe delovanja? Ali bi lahko zamenjali današnje procesorje s še manjšimi miniaturnimi integriranimi vezji ali celo s posameznimi atomi, ioni, elektroni? To so vprašanja, ki označujejo današnji razvoj računalništva. Leta so se fiziki v laboratorijskih okoljih ukvarjali z delom na atomarnem in subatomarnem nivoju. V današnjem času poskušajo izvesti računske operacije z izkoriščanjem fizikalnih lastnosti, ki veljajo za atomarni nivo. Delo na tem področju se precej razlikuje od običajne uporabe binarne logike in principov *Boolove algebre*¹⁰⁴. Odkrite so nove zakonitosti, ki se jih da uporabiti za izvajanje računskih operacij. Običajna fizikalna pravila, ki veljajo tudi za najbolj miniaturizirana *CMOS logična vrata*,¹⁰⁵ tu ne veljajo več. Namesto tega veljajo za subatomarne delce pravila *kvantne mehanike*.

Kvantni računalnik prav v ničemer ni podoben konvencionalnim digitalnim računalnikom, narejenim na osnovi silicija iz tranzistorjev in diod, kot jih poznamo danes. Za kvantni računalnik potrebujemo popolnoma nove tehnologije, tehnologije, ki bodo omogočale *kvantnim bitom* (imenujmo jih kar po angleško '*qubiti*') obstajati kot *koherentna superpozicija* ničel in enic. Danes je izdelana že vrsta bolj ali manj uspešnih poskusov izdelave kvantnih računalnikov, komercialno zanimiva tehnologija za to proizvodnjo še vedno ni znana. Za razumevanje delovanja kvantnih računalnikov in problemov pri tem pa potrebujemo nekaj splošnega teoretičnega znanja.

TEORETIČNE OSNOVE KVANTNIH RAČUNALNIKOV

Današnje izboljšave na področju računalništva so prvenstveno usmerjene v zmanjševanje tranzistorjev, ki jih uporabljamo v sodobnih procesorjih. (Kurzweil, 2005) To pa se ne more nadaljevati v nedogled, saj, ko bodo tranzistorji postali dovolj majhni, zanje ne bo več veljala mehanika velikih objektov (npr. posplošena teorija relativnosti), temveč se bodo pričeli srečevati z vplivi kvantne mehanike. Ti

¹⁰⁴ Boolova algebra je računanje z logičnimi vrednostmi 1 (true/resnica) in 0 (false/neresnica). Za osnovno računanje uporabljamo logične operacije: AND (logični in), OR (logični ali), NOT, XOR, NOR.

¹⁰⁵ CMOS logična vrata so osnovni sestavni elementi logičnih vezij za izvajanje operacij, kot so AND, OR, negirani AND in OR itd.

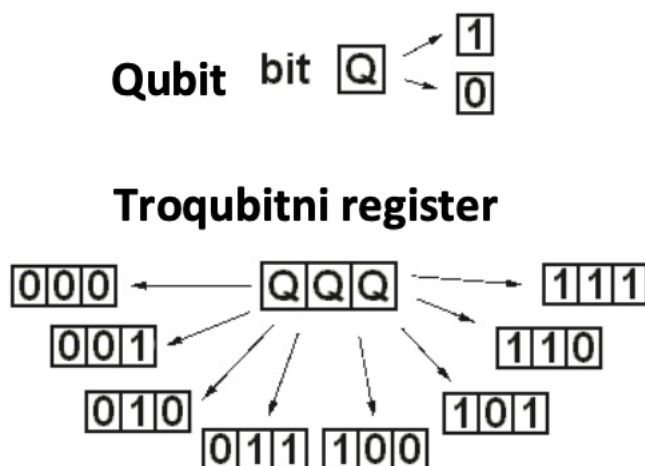
vplivi bodo verjetno predstavljali temeljne omejitve današnje računalniške tehnologije. In ob upoštevanju tega se nam lahko zastavita vprašanji, kje smo danes in koliko »manevrskega« prostora v razvoju »konvencionalnih« visokosposobnih digitalnih računalnikov sploh še imamo.

Leta 1982 je fizik Richard Freyman razvil idejo *kvantnega računalnika*, računalnika, ki uporablja vse pozitivne vplive kvantne mehanike. Ta ideja, ki je bila nekaj časa popolnoma potisnjena v ozadje, je danes ponovno aktualna, še posebej, ko je Peter Shor razvil poseben algoritem za obdelavo velikih števil na kvantnem računalniku. (Shor, 1997)

Pri klasičnem modelu računalnika so najbolj elementarni gradniki '*biti*'. Ti lahko obstajajo le v dveh diametralnih stanjih, to je ali v 0 ali v 1. Vsak delec pa lahko opišemo kot delec ali kot valovanje. Kadar subatomarni delec obravnavamo kot delec, potem ta lahko zaseda le diskretna energijska stanja, ki jim rečemo energijski kvanti (0 ali 1). Valovni opis pa govori o tem istem delcu v prostorsko-časovnem smislu. Ko se val, ki predstavlja elektron, razširi v prostoru in času, postane lokacija elektrona negotova; veljajo zakoni verjetnosti, kar izkoriščamo pri kvantnem računanju. Pri kvantnih računalnikih so torej pravila in možnosti popolnoma drugačna. (Kurzweil, 2005) Kvantni *bit*, poimenovali smo ga *qubit*, ne obstaja le v stanju 0 ali 1, temveč lahko obstaja tudi v *koherentni superpoziciji obeh*. Ko je *qubit* v takšnem stanju, si lahko zamislimo, kot da obstaja v dveh vesoljih, v enem kot 0 in v drugem kot 1. Torej se operacije na teh qubitih v bistvu izvajajo na obeh vrednostih istočasno. Pri tem pa je pomembno, da ko izvajamo posamično operacijo na qubitu, izvajamo to operacijo pravzaprav na dveh različnih vrednostih. Analogno temu torej pri dvoqubitnem sistemu izvajamo operacije na štirih vrednostih istočasno, pri troqubitnem sistemu pa že na osmih itd.

Povečevanje qubitov tako eksponentialno povečuje »*kvantni paralelizem*«. ¹⁰⁶ S pravilnim tipom algoritma je tako mogoče uporabiti ta paralelizem za reševanje določenih problemov (tudi problemov umetne inteligence), ki čas reševanja problema zmanjša le na droben delček časa klasičnega računalnika. S tem pa se hitrost reševanja nekega problema drastično skrajša.

¹⁰⁶ Kvantni paralelizem je metoda, s katero lahko kvantni računalnik izvaja dve računski operaciji istočasno.



Slika 7.19: Možne operacije pri troqubitnem kvantnem računalniku

Podatomska fizika temelji na kvantni mehaniki, kar je vzrok za to, da je kvantno računanje tako mogočno, pa tudi zato, da je računanje izjemno krhko in težko obvladljivo. Npr. predpostavimo, da je qubit v koherentnem stanju. Takoj, ko ga lahko izmerimo, bo ta zaradi vzajemnega vpliva okolja postal nekoherenten in se spremenil v eno od dveh klasičnih stanj, skratka ali v 0 ali v 1. To je problem dekoherence in je begajoč del kvantnih računalnikov, saj je potencialna moč kvantnih računalnikov odvisna od kvantnega paralelizma, ki je ustvarjen s *koherentnim stanjem*.¹⁰⁷ Ta problem je še toliko bolj pereč, da tudi, če na qubit gledamo, kot da lahko dekoherira, to celoten proces določanja rezultatov dela vsaj tako težavnega, kot je računanje samo.

7.7.4 Teorija univerzalnega računanja

Za vsak računalnik – od samega začetka računalništva (npr. Babbageovega analitičnega stroja iz leta 1837) pa vse do danes – je značilno, da se uklanja teoriji klasičnega računanja, ki jo je opisal Alain Turing. V bistvu Turing opisuje idejo univerzalnega Turingovega stroja, zelo enostavnega modela računalnika, ki je lahko

¹⁰⁷ Superpozicije so lahko koherentne – pri tem gre za t. i. koherentno stanje, kot je denimo idealizirana laserska svetloba, ki jo seva idealni laser – ali pa kaotične – termično stanje, ki opisuje svetlobo v toplotnem ravnovesju – sevanje črnega telesa.

programiran tako, da izvaja katerokoli izračunljivo operacijo. Vsi računalniki od začetka pa do danes so v bistvu implementacija takšnega univerzalnega Turingovega stroja. Čeprav so morda nekateri računalniki hitrejši, zmogljivejši, večji ali dražji od drugih, pa so vsi po osnovni funkciji enaki in izvajajo enak niz računalniških nalog.

UNIVERZALNI KVANTNI RAČUNALNIK

Church-Turingova teza: »Obstaja (ali ga lahko zgradimo) univerzalni računalnik, ki ga je mogoče programirati tako, da lahko izvaja katerokoli računsko nalogo, ki je izvedljiva s fizičnimi objekti« (v Kleene, 1952: 317)

V zadnjih desetletjih je bila izpostavljena vrsta prednosti kvantnega računanja pred konvencionalnim, diskretnim. Ena od prvih, verjetno pa tudi najpomembnejših je bilo odkritje Richarda Freynmana leta 1982, da lahko preprosti *univerzalni simulator* posnema obnašanje kateregakoli končnega fizikalnega objekta. David Albert je nato leta 1984 dodal drugo pomembno odkritje, ko je opisal *samomerilni kvantni avtomat*, ki lahko izvaja aktivnosti, ki jih ne more izvajati noben konvencionalni digitalni računalnik. S tem, ko samomerilni kvantni avtomat izvaja merjenje samega sebe, lahko dobi »osebne, njemu lastne« informacije, ki jih je popolnoma nemogoče dobiti z meritvami od zunaj. In tretji, morda najpomembnejši kamenček v mozaiku, je dodal David Deutsch leta 1989. Dokazal je, da se lahko vse računske zmožnosti kateregakoli končnega stroja, ki se uklanja zakonom kvantnega računanja, nahajajo v samo enem in edinem stroju, v *univerzalnem kvantnem računalniku*. Takšen računalnik bi lahko bil zgrajen iz vrat,¹⁰⁸ kvantnih ekvivalentov *Toffoli vrat*¹⁰⁹ z dodatkom nekaj posebnih operacij za linearno superponiranje stanja 0 in 1. To Deutschovo odkritje zahteva manjšo spremembo Church-Turingove teze. S tem dobimo Church-Turing-Deutschov princip, ki predstavlja močnejšo fizikalno obliko Church-Turingove teze in poudarja, da univerzalni računalnik lahko simulira vsak fizikalni proces (Deutsch, 2012) in bi se lahko glasil nekako tako: obstaja (ali ga lahko zgradimo), univerzalni *kvantni* računalnik, ki ga je mogoče programirati tako, da lahko izvaja katerokoli

¹⁰⁸ Vrata so osnovne enote, iz katerih so zgrajeni računalniki. Vrata, tudi logična vrata, imajo svojo osnovo v matematični logiki in so lahko IN, ALI ali NE, IZKLJUČOJOČI ALI ipd. (angl. AND, OR, NOT, XOR ipd.)

¹⁰⁹ Toffoli vrata: navadna, npr. IN vrata imajo dva vhoda in le en izhod, kar pomeni, da v procesu na poti od vhoda do izhoda izgubimo en bit informacij. Na ta način postane proces računanja nereverzibilen, saj iz izhodnih informacij ne moremo sklepati na vhodne informacije. Leta 1976 je Charles Bennett dokazal, da je mogoče zgraditi univerzalni računalnik, ki bo sestavljen izključno iz reverzibilnih vrat in pri tem ne bo bistveno upočasnjeno računanje. Takšna reverzibilna vrata so tudi Toffoli vrata.

računsko nalogo, ki je izvedljiva s fizičnimi objekti. Leta 1993 je Seth Lloyd pokazal, da veliko kvantnih sistemov, vključno s preprostim kristalom soli, lahko deluje kot kvantni računalnik. Istega leta je Peter W. Shor demonstriral, da kvantni računalnik lahko izvaja praktično delo hitreje kot katerikoli znani digitalni računalnik. Vsa ta odkritja so sprožila renesanso raziskav kvantnega računanja. Številne skupine skušajo zgraditi prototipne elemente kvantnih računalniških sistemov.

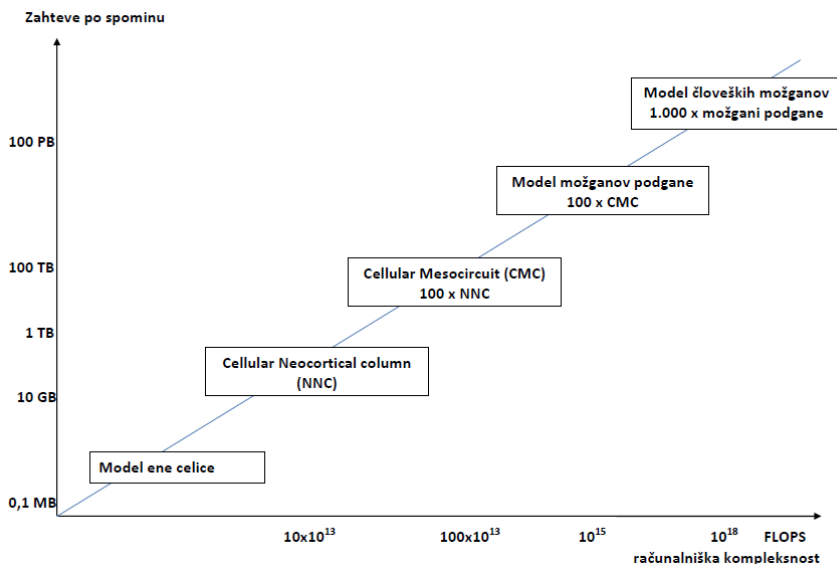
Teorije o kvantnem računanju imajo nekatere zanimive implikacije tudi za svet umetne inteligence. Razprave o tem, ali bodo stroji (računalniki) sploh kdaj sposobni biti inteligentni (tj. umetno inteligentni), danes temeljijo predvsem na filozofskih argumentih. Teorija kvantnega računanja nam dovoljuje (in omogoča), da na zavest gledamo z nekoliko drugačne perspektive. Pri tem je prvo omembe vredno dejstvo, da lahko vsak fizični objekt (od enostavnega kamna pa do celotnega vesolja) smatramo kot kvantni računalnik in da lahko vsak fizični proces, ki ga lahko zaznamo, smatramo kot računanje. (Kurzweil, 2005) Naslednja stopnja argumentov pa temelji na Church-Turingovi tezi oz. Church-Turing-Deutschovem principu in določa, da odkar je vsak računalnik vsakemu računalniku funkcijsko ekvivalenten in lahko vsak vsakega glede na njegovo osnovno funkcijo nadomesti. (Deutsch, 2012) Nekateri znanstveniki tako verjamejo, da bo kvantni računalnik ključ za rešitev problema UI, s čimer pa se, tako kot po navadi, mnogi ne strinjajo, npr. Roger Penrose, ki verjame, da zavest zahteva celo bolj sofisticirano (in danes še neznano) fizikalno teorijo od omenjene kvantne.

7.7.5 Projekt modri možgani

Verjetno je na področju kreiranja modela umetnih možganov prišel še najdlje projekt *Modri možgani*, zato si v nadaljevanju na kratko oglejmo njegove izsledke. Na osnovi zgoraj omenjenih izhodišč so bili v letu 2005 definirani osnovni cilji projekta človeških možganov, imenovanega *Modri možgani*, ki so našli svoj domicil na École polytechnique fédérale de Lausanne (EPFL) v Švica.

Projekt ustvarjanja umetnih človeških možganov bi lahko prišel do svojih rezultatov v naslednjih desetletjih. Osnovni cilj projekta je razrešiti vėdenje o skrivnosti možganov in poiskati alternativne možnosti zdravljenja mentalnih bolezni. Teoretično je to sicer možno, obstajata pa dva velika problema:

1. Še vedno ne poznamo vseh fenomenov delovanja človeških možganov. V predhodnem poglavju smo prikazali matematični model modeliranja umetnega nevrona in spoznali, da je to dokaj zapleten proces že na nivoju nevrona samega.
2. Še večji problem je potrebna računalniška moč. Na sliki 7.20 je nakazana potrebna moč za simuliranje kognitivnih procesov za različne zahteve po spominu in procesorski moči.



Slika 7.20: Potrebe po spominu in procesorski moči

Kot smo že pokazali, smo do sedaj sposobni modeliranja tako samostojne celice kot celičnih sistemov kortikalne kolumne, ki vsebujejo približno 10.000 celic, 10.000 nevronov. Strokovnjakom na EPFL je do sedaj uspelo najbolj modelirati možgane podgan, ki vsebujejo 10^8 celic. Ta bionevronska mreža je velikosti bučikine glave in ima 100.000 kortikularnih kolumnov, od katerih vsak vsebuje 10.000 nevronov v možganski opni in zahteva 20.000 krat več spomina kot za kortikularni kolumn. To pa je neznatno v primerjavi s človeškimi možgani, kjer ima človeška možganska opna nekako 2.000.000 kortikularnih kolumnov s po 100.000 nevroni. Tako bi simulacija človeških možganov zahtevala dodatno 1000-kratno povečanje tako spomina kot tudi računalniške moči. Ta razmerja so prikazana na sliki 7.20.

Vidimo, da so potrebe po računalniški moči ogromne. Samo za vsako simulacijo nevrona bi potrebovali ekvivalent enega notesnika. Za model celotnih možganov pa bi potrebovali na milijarde ekvivalentov notesnikov. Tehnologija superračunalnikov se sicer res hitro približuje nivoju, ko bo modeliranje možganov morda postalo realnost, predvideva se nekako do leta 2020 (ob upoštevanju Moorovega zakona), vendar pa so to le optimistične predpostavke, ki ne upoštevajo (ne morejo upoštevati) vseh možnih zapletov na poti razvoja. Pri tem lahko upoštevamo Bekensteinovo mejo¹¹⁰, ki predstavlja zgornjo mejo informacij, ki se lahko nahajajo na končnem prostoru, ki ima končno količino energije oz. z drugimi besedami predstavlja maksimalno količino informacij potrebnih za natančen opis nekega fizikalnega sistema vse do kvantnega nivoja. (Bekenstein, 1972)

Povprečni človeški možgani tehtajo 1,5 kg in imajo prostornino 1260 cm³. Energija (izračunana po enačbi $E = mc^2$) je 1.34813×10^{17} J. Če predpostavimo možgane kot kroglo ($V = 4 \pi r^3/3$), potem bi ta imela polmer 6.70030×10^{-2} m. Bekensteinova meja ($I \leq 2 \pi r E h c \ln 2$) tako znaša za povprečne človeške možgane 2.58991×10^{42} bitov in predstavlja zgornjo mejo potrebnih informacij za idealno pretvorbo povprečnih človeških možganov na kvantni nivo. Iz tega torej izhaja, da je število različnih stanj ($\Omega = 2^I$) človeških možganov (in tudi duha, ob fizikalistični predpostavki) največ $10^{7.79640 \times 10^{41}}$.

Modeli na celičnem nivoju, ki bi natančno predstavljali vso elektrofiziološko komunikacijo, bi potrebovali približno 20.000 diferencialnih enačb. Za dodatno modeliranje vaskularnega sistema, ustvarjanje virtualnih instrumentov (npr. virtualni EEG, virtualni fMRI) pa bi bilo seveda potrebno še dodatno povečanje teh zahtev. Nobena sodobna grafična postaja trenutno ni zmožna rešiti teh enačb v biološkem realnem času. Z drugimi besedami je edini način, kako reševati te in takšne probleme, da razvijemo in uporabimo visoko sposobne računalnike (angl. HPC - High Performance Computing), ob tem pa je seveda potrebno poudariti tudi to, da gre le za predvidevanja. Realne potrebe se bodo pokazale med samim izvajanjem (npr. projekta *Modri možgani*, če bo do izvedbe sploh prišlo).

¹¹⁰ Bekensteinova meja določa maksimalno mejo izvedenih informacij na nekem končnem območju.

7.8 Sklep

V sedmem poglavju s teorijo dinamičnih sistemov zaokrožujemo področje kognitivnega modeliranja in zagovarjamo tezo, da je vesolje dinamični sistem. Iz tega sledi, da je tudi človek in vse, kar se nanj nanaša oz. iz česar je človek sestavljen, dinamični sistem, zaradi česar lahko za razlago delovanja človeka uporabljamo isto teorijo, tj. teorijo dinamičnih sistemov. S pomočjo te teorije definiramo prostor stanja in atraktorje v njem, osnove teorije nihanja in se posvetimo odnosom med originalom in dinamičnim modelom, s posebnim poudarkom na matematičnem modeliranju človeškega telesa in duha.

Splošno teorijo dinamičnih sistemov prenesemo v kognitivno znanost in zagovarjamo tezo, da je možgane na teoretičnem nivoju možno modelirati kot kognitivne dinamične sisteme. Tezo o modeliranju duha potrdimo na poenostavljenem primeru rešitve modeliranja možganskih aktivnosti z uporabo novih oblik valovnih enačb. V enodimenzionalnem prostoru je rešitev odvisna od Lipschitzove zveznosti za začetne podatke in je tako edinstvena. To pomeni, da so odzivi na majhne motnje majhni kot se le da, kar je skladno tudi s fizikalnim glediščem. Model je za to razlago poenostavljen, z vpeljavo dodatnih spremenljivk pa bi lahko postal bolj realističen, kar pa za našo razlago ni nujno potrebno. Predstavljeni model predpostavlja:

- da je geometrija možganov enodimenzionalna (dvodimenzionalni model bi bil sicer bolj dosleden, vendar vpeljuje dodatne probleme);
- zvezna funkcija med nevronskimi mejami je homogena, saj ima več kot 50 % nevronskih mej homogene povezave;
- zunanji vnos p se sklada z $H^2([0, T], L^\infty(\Omega))$, v resnici pa je p lahko manj gladek ali bolj nelokalen.

Iz prikazanega je razvidno, da sicer diferencialne enačbe, s katerimi opisujemo elemente kognitivnih sistemov, same po sebi niti niso izrazito zahtevne, vendar pa je zahtevno njihovo reševanje, saj se lahko reševanje izvaja le z uporabo numeričnih metod, kar posledično zahteva ogromno računalniško moč in spominski prostor. S tem pa stopamo v prihodnost (po trenutnih predvidevanjih v leta 2025 do 2050) in na področje napovedovanja ter netradicionalnih znanosti, znanosti, ki so še danes v

razvoju ali jih še celo ne poznamo. Dotaknemo se vprašanja, ali lahko na diskretnih sistemih zares uspešno obravnavamo zvezne sisteme, kar možgani zagotovo so, ali pa smo z današnjo računalniško tehnologijo že dosegli njen »rob« in bo potrebno za naslednji korak narediti miselni in tehnološki preskok, t. i. »kvantni preskok«. Kot odgovor na ta vprašanja nakažemo možne potencialne (bolj ali manj verjetne) rešitve z različnimi nedigitalnimi pristopi k računanju in računalništvu od bioračunalnikov, delujočih na principu DNK in proteinov, pa do kvantnih računalnikov, katerih osnovni principi obdelav podatkov izhajajo iz kvantne teorije in kvantnega paralelizma.

Obstajajo trdni in utemeljeni razlogi, da so napovedi ustvarjanja umetnih možganov oz. kopiranja možganov v bližnji prihodnosti preveč optimistične. Še posebej zato, ker trenutno ne obstajajo potrebna znanja (in predvsem ne razumevanje znanj, ki jih imamo) niti o možganih niti o kogniciji in prav tako ne vemo, kakšno računanje in kakšni računalniki bi za to bili potrebni.

Poleg filozofskih dvomov so prisotni seveda tudi tehnološki problemi, problemi povezani z računalniško močjo in njihovim izkoristkom. Možgani porabijo približno 10–20 W moči, v nasprotju s tem pa superračunalniki porabijo celo več deset tisoč MW, če pri tem upoštevamo *Landauerjevo limito*¹¹¹, ki je 3.5×10^{20} operacij/sekundo/W pri sobni temperaturi in ob upoštevanju *Bekensteinove meje*, ki je sicer po predvidevanjih le zgornja meja informacij na nekem podanem končnem območju (npr. v možganih ta limita znaša 2.58991×10^{42} operacij – bitov). Razen tega se pojavi še en problem, to je sproščena energija, največ v obliki toplote, saj vemo, da je že pri današnjih procesorjih, ki so po svojih zmogljivostih in porabljeni energiji še »svetlobna leta« od te meje, osnovni problem hlajenje oz. odvajanje te sproščene energije.

Ollar Stone Fuller (1990) je v svoji študiji o zapletenih dinamičnih sistemih izpostavil, da imajo naravni sistemi skupno lastnost, da delujejo zelo blizu mejne vrednosti parametra, kar se kaže v neurejenem obnašanju celotnega sistema (sistemi so nestabilni, lahko so nekaj, ali pa nekaj popolnoma drugega). Pravi, da to dovoljuje najučinkovitejšo obdelavo informacij v vseh delih sistema skozi čas in prostor, kar

¹¹¹ Landauerjeva limita definira, koliko operacij v časovni enoti (sekundi) se lahko izvede na W moči pri sobni temperaturi.

je potrebno za primerno vzdrževanje in delovanje sistema. Ta ideja se zdi intuitivno smiselna, ker se to delovanje na mejni točki kaosa dogaja tam, kjer bo sistem lahko našel največ svobode in možnosti za drastične spremembe smeri, rezultata ali stanja. S tem bodo spremembe začetnih, trenutnih ali naknadnih pogojev minimalne in zaradi kaotičnih lastnosti občutljive na začetne pogoje.

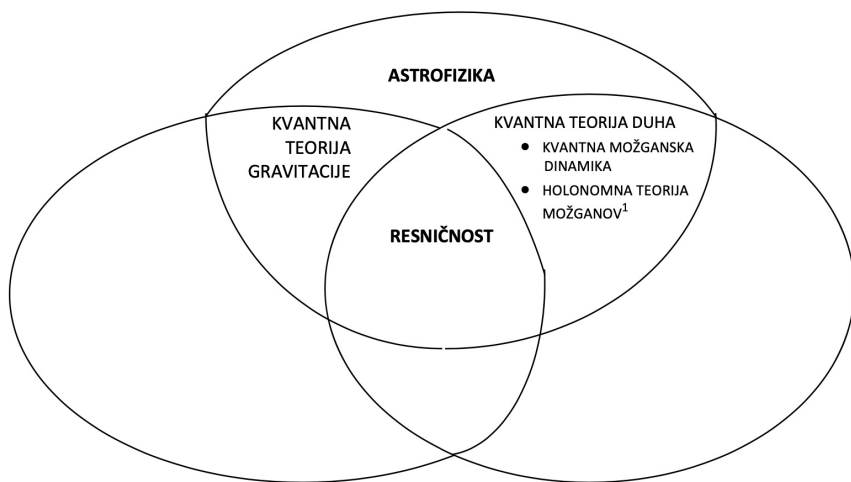
Nasproti vsemu temu in kot odgovor na izhodiščno vprašanje pa so nekateri filozofi, ki so se ukvarjali z epistemologijo, med njimi tudi David Hume, Immanuel Kant in Charles Saunders Pierce glede »mej človeškega razumevanja« menili, da obstajajo stvari, ki jih nikakor ne moremo vedeti, vsaj ne na trenutni stopnji razvoja človeške rase.

8 Resničnost, prostor in čas

Prostor in čas sta eni od najbolj vznemirljivih področij človekovega raziskovanja, saj tvorita areno resničnosti, snov iz katere ja zgrajeno vesolje. Vse, kar počnemo, razmišljamo in doživljamo, poteka v nekem delu prostora in v določenem času. Pa znanost še vedno ne razume povsem, kaj prostor in čas sploh predstavljata in kako bi si ju lahko razlagali s pomočjo znanosti. Ali sta prostor in čas realni fizikalni veličini ali le uporabni ideji, sta del materialnega ali idealističnega sveta? In če sta realni, materialni ali sta osnovni, neodvisni veličini, ali izhajata iz še bolj temeljnih sestavin? Kaj pomeni, da je prostor pretežno prazen? Ali ima čas začetek? In ne nazadnje, ali lahko resnično nadziramo prostor in čas? V nadaljevanju bomo poskušali zelo na kratko orisati približno tristoletna prizadevanja iskanja odgovorov na ta osnovna vprašanja. Dotaknili se bomo definicije prostora stanja, kjer igra eno od ključnih vlog tudi čas (Green, 2006). Na spodnji sliki 8.1 so prikazane povezave med vedami, ki v današnjem svetu ustvarjajo videz realnosti v sodobni znanosti.

Pri kakršnihkoli znanstvenih raziskavah se vedno srečamo s širšim, a zelo izmuzljivim vprašanjem: *Kaj pravzaprav je resničnost?* Kako vemo, da je kakršnakoli naša predpostavka, ideja, teorija ali ne nazadnje matematični model raziskovanega sistema odraz resničnega stanja tega sistema v realnem prostoru in času? Kako smo lahko ljudje, ki imamo dostop le do notranjih zaznav in misli, prepričani, da le-te v resnici odražajo zunanji svet? Filozofi so že zdavnaj prepoznali ta problem, ostali znanstveniki pa se zavedajo, da svet, ki ga opazujemo, ki se razvija na prizorišču

prostora in časa, morda v resnici nima veliko skupnega z želeno resničnostjo. Predvsem v zadnjih sto letih so znanstvena odkritja privedla do ugotovitev, ki so tako dramatične, osupljive in presenetljive, kot je lahko najbolj domiselna znanstvena fantastika. A gremo po vrsti.



Slika 8.1: Povezave med vedami, ki ustvarjajo resničnost.

Mnoga zgoraj zastavljena vprašanja, morda v različnih preoblikah, so si zastavljali že Aristotel, Galileo, Newton, Einstein in mnogi drugi, in tako lahko zasledujemo odgovore, ki jih je ena generacija postavila, naslednja ovrgla ter v naslednjih stoletjih izboljšala in nadgradila sodobna znanost. Tako lahko npr. sledimo odgovorom na vprašanje o popolnoma praznem prostoru kot realni entiteti ali zgolj abstraktni zamisli od Isaaca Newtona, ki je v 17. stoletju trdil, da je prostor resničen, preko Ernsta Macha, ki je v 19. stoletju zaključil, da ni, pa do Alberta Einsteina, ki je v 20. stoletju dramatično preoblikoval osnovno vprašanje in spojil prostor in čas ter v veliki meri ponovno ovrgel trditve predhodnika Macha. Nato, v nadaljevanju, se z redefinicijo pojma »prazno«, vprašanje postavlja na novo, kajti prostor naj bi bil prežet s t. i. kvantnimi polji in morda tudi z razpršeno enakomerno energijo, imenovano kozmološka konstanta – sodobnim odmevom starega in ovrženega Descartesovega pojma o etru, ki preplavlja prostor. Naša doba pa je ponudila dognanja, da je prostor, ki ga naseljujemo, morda le drobec v veličastnejšem Kozmosu. Osredotočili se bomo le na delčke, izseke, na najvznemirljivejše in najbolj

bistvene značilnosti, pomembne za celovito podobo resničnosti, ki jo slika sodobna znanost s posebnim poudarkom na mikrokozmosu, –na povezavo med kvantno mehaniko in duhom, ki jo lahko na kratko poimenujemo kvantna teorija duha (zavesti).

Kot je dejal v Alžiriji rojeni filozof Albert Camus, napredek v fiziki, npr. razumevanje števila prostorskih dimenzij, napredek v nevropsihologiji, kot je poznavanje organizacijskih struktur v možganih ali ne navsezadnje napredek na kateremkoli znanstvenem področju prispeva pomembne podrobnosti, a vpliv teh izsledkov na našo oceno življenja in resničnosti naših spoznanj ostaja neznaten. Gotovo pa ni resničnost tisto, kar mislimo, da je; resničnost se nam razkrije šele skozi naše izkušnje. (Camus, 1999) Ob znanstvenih prizadevanjih zadnjih desetletij smo se naučili, da so človeške izkušnje pogosto zavajajoč vodnik k pravi naravi resničnosti, saj tik pod površino vsakdana obstaja nek drug svet, ki ga povečini komaj prepoznamo.

V zadnjih treh stoletjih so fizikalna odkritja razkrila, da sta prostor in čas najbolj begajoča in skrivnostna koncepta, in hkrati najbolj bistvena za kakršnokoli znanstveno analizo vesolja in življenja v njem. Razvoj je tudi pokazal, da sta prostor in čas na samem vrhu seznama prastarih znanstvenih konstruktov, ki so jih najnaprednejše raziskave korenito spremenile. Za Isaaca Newtona sta prostor in čas preprosto obstajala – tvorila sta neaktivno, univerzalno kozmično prizorišče, na katerem so potekali vsi dogodki. Za njegovega sodobnika Gottfrieda Wilhelma Leibniza pa sta bila prostor in čas le veličini za podajanje informacij o tem, kje so telesa in kdaj se dogodki zgodijo in nič več. Za Alberta Einsteina sta bila prostor in čas surovina v ozadju resničnosti. S svojima teorijama relativnosti je Einstein v celoti spremenil naše razmišljanje o prostoru in času in razkril njuno vodilno vlogo, ki jo imata za razvoj vesolja. Od njegovih časov sta tako postala bistvena elementa, postala sta znana, a ob enem tuja; popolno razumevanje prostora in časa je postalo največji izziv sodobne fizike. (Green, 2006)

Temeljnih fizikalnih zakonitosti se bomo dotaknili le bežno, v našem znanstvenem prizadevanju se bomo nekoliko bolj osredotočili predvsem na filozofske osnove. Ker zgodba prostora in časa še zdaleč ni do konca zapisana, tudi mi ne bomo prišli do dokončnih odgovorov. Spoznali pa bomo niz dognanj, nekatere tudi zelo nenavadne, druge globoko pomirjajoče, nekatere eksperimentalno potrjene, druge

povsem hipotetične, ki bodo pokazala, v kolikšni meri smo z duhom zajeli vse možne znanstvene konsekvence in se dotaknili prave teksture resničnosti.

KLASIČNA RESNIČNOST

Zgodovinarji niso enotni, kdaj se je sodobna znanost začela, a ko so na sceno stopili Galileo Galilei, René Descartes in Isaac Newton je zagotovo že bila v polnem razmahu. V tistih časih se je vztrajno oblikovalo novo znanstveno razmišljanje. Ti začetki sodobne znanstvene misli so trdili, da je dogajanje v znanosti, če ga pogledamo s prave perspektive, ne le razložljivo, pač pa tudi predvidljivo. Od tega trenutka naprej je bila razkrita moč znanosti, da dosledno napoveduje določene vidike prihodnosti.

Prve znanstvene raziskave so se osredotočale na vsakodnevne življenjske izkušnje in njihove entitete. Cilj teh raziskav je bil s porajajočim znanstvenim instrumentarijem prisluhniti naravi. Pri tem je bil Isaac Newton osrednji lik, ki je združil bogastvo fizikalnih pojavov v en sam teoretični okvir. V svojem delu *Principia mathematica* je izrazil tudi svoje pojmovanje prostora in časa in ju razglasil kot absolutna in nespremenjena elementa, kar je veljalo naslednjih dvesto let. V desetletjih po Newtonu so njegovi nasledniki njegove enačbe umestili v zapletene matematične strukture, kar je bistveno razširilo njihov domet in praktično uporabnost.

RELATIVISTIČNA RESNIČNOST

Klasična fizika je ponudila trdno osnovo človeški intuiciji in leta 1894 je ugledni eksperimentalni fizik Albert Michelson dejal, da je »večina pomembnih temeljnih načel povsem dognana«. Leta 1900 je Kelvin zapisal, da sta ostala le še »dva oblaka«, eden povezan z lastnostmi gibanja svetlobe in drugi v zvezi s sevanjem segretega telesa, ki bosta nedvomno kmalu razrešena. Naslednje desetletje pa je njegovo trditev v celoti postavilo na glavo. *Relativnostna revolucija*, ki se nanaša na prvega od njegovih »oblakov«, sega v leta od 1905 do 1915. Albert Einstein je v tem času dokončal svojo splošno in posebno teorijo relativnosti. V njej ugotavlja, da je Newtonovo dojemanje prostora in časa, na katerem temelji klasična fizika, pomanjkljivo. V začetku leta 1905 je dognal, da prostor in čas nista neodvisna in absolutna, pač pa prepletena in relativna na način, ki nasprotuje vsakdanjim izkustvom. Ni pokazal, da sta prostor

in čas del celote, pač pa je pokazal tudi, da preko ukrivljenja sodelujeta v kozmični evoluciji, sta skratka prožna in dinamična.

KVANTNA RESNIČNOST

Drugi Kelvinov »oblak« pa je privedel do *kvantne revolucije*, ene največjih prelomnic, ki jim je bilo sodobno človeško razumevanje kadarkoli priča. Ko so se začetna nasprotovanja polegla, je pretekli sijaj klasične fizike za stalno obledel v luči kvantne resničnosti.

Ključna značilnost klasične fizike je, da izhodiščne enačbe na osnovi položajev in hitrosti vseh teles v določenem trenutku napovedo njihov položaj in hitrost v vsakem drugem trenutku, preteklem ali prihodnjem, s čimer brez dvomov klasična fizika napove, da morata biti preteklost in prihodnost vtisnjeni prav v sedanost, kar pravzaprav menita tudi obe relativnostni teoriji.

V tridesetih letih prejšnjega stoletja so bili fiziki prisiljeni uvesti popolnoma nov pojem, namreč *kvant*, na katerem temelji t.i. *kvantna mehanika*. Povsem nepričakovano so odkrili, da le kvantni zakoni lahko razrešijo množice ugank in razložijo podatke, ki so prihajali iz atomskega in subatomskega sveta. A po kvantnih zakonih tudi v primeru, da opravite najboljšo možno meritev današnje situacije, lahko napoveste kvečjemu verjetnost, da bodo stvari takšne ali drugačne v nekem določenem trenutku v prihodnosti ali da so bile stvari drugačne v nekem izbranem trenutku v preteklosti. Skratka, po tej teoriji vesolje (kot najširši termin) ni vtisnjeno v sedanost, ampak le sodeluje v igri naključij.

Kvantna mehanika, poimenovana včasih tudi kvantna fizika ali kvantna teorija, je del fizike, ki se ukvarja z matematičnim opisovanjem lastnosti dvojnosti delca, ki je lahko tako delec kot val, torej s povezavami med materijo in energijo. Kvantna mehanika opisuje časovni razvoj fizikalnih sistemov s pomočjo matematičnih struktur, poimenovanih *valovne funkcije*. Valovna funkcija napoveduje verjetnost, da se lahko sistem nahaja v podanem stanju v določenem času.

Zgodovina kvantne mehanike sega v leto 1838, ko je Michael Faraday odkril katodni žarek. Temu je sledil leta 1859 problem sevanja črnega telesa, katerega avtor je bil Gustav Kirchhoff, in nato leta 1877 predlog Ludwiga Boltzmann o diskretni obliki

energije fizikalnih sistemov, čemur je sledila leta 1900 kvantna hipoteza Maxa Plancka. (Mehra, Rechenberg, 1982) Planckova hipoteza, da je energija lahko sevana ali absorbirana v diskretnih »kvantih«, »elementih energije«, natančno ustreza ugotovitvam opazovanj vzorcev sevanja črnega telesa. Po Plancku je vsak element energije E sorazmeren njegovi frekvenci ν :

$$E = h \nu, \quad (8.1)$$

pri čemer je h Planckova konstanta. Planck je pri tem trdil, da je to preprosto proces absorpcije in sevanja in nima prav nič skupnega s fizično realnostjo sevanja samega (Kuhn, 1978). Nasprotno pa je Albert Einstein Planckovo kvantno hipotezo interpretiral bolj realistično in za to uporabil razlago fotoelektričnega učinka, po kateri sevanje svetlobe na površino materiala lahko izbije elektron iz tega materiala.

Osnove kvantne mehanike so postavili v prvi polovici 20. stoletja fiziki, kot so Niels Henrik David Bohr, Werner Karl Heisenberg, Erwin Schrödinger, Vladimir Aleksandrovič Fok, Paul Adrien Maurice Dirac in drugi. Nekatere osnovne vidike teorije še vedno dejavno raziskujejo, po drugi strani pa izsledke kvantne mehanike že dolgo uporabljajo številne veje fizike in kemije, med njimi fizika kondenzirane snovi, kvantna kemija in fizika osnovnih delcev.

Kvantna mehanika je bila v izvorni obliki namenjena boljšim razlagam atoma, še posebej za razlago razlik v spektru svetlobe, ki so ga sevali različni izotopi istega elementa. Kvantna teorija atoma je bila odkrita kot razlaga za elektrone, ki so ostajali v svojih orbitah in jih ni bilo možno razložiti niti z Newtonovimi zakoni gibanja niti z Maxwellovimi zakoni klasičnega elektromagnetizma. Širše rečeno pa se kvantna mehanika loteva reševanja štirih fenomenov, na katere klasična fizika ne more najti odgovorov, in to:

- kvantizacije nekaterih fizikalnih lastnosti,
- val-delec dualizma,
- princip nedoločenosti,
- kvantna razmerja.

Opis teorije

Kvantna mehanika opisuje trenutno stanje sistema z valovno funkcijo, s katero je povezana verjetnostna gostota vseh merljivih lastnosti ali *opazljivke*. Opazljivke sistema so lahko energija, lega, gibalna količina, vrtilna količina ipd. V kvantni mehaniki opazljivkam ne moremo pripisati določenih vrednosti, ampak lahko sklepamo le o njihovih verjetnostnih porazdelitvah. Valovno obnašanje snovi lahko razložimo z interferenco valovnih funkcij.

Valovne funkcije so lahko odvisne od časa. V nekem trenutku lahko denimo delec v praznem prostoru opišemo z valovno funkcijo, ki je valovni paket s središčem v neki povprečni legi. V nekem poznejšem času se valovni paket spremeni, s tem pa je tudi večja verjetnost, da delec najdemo na nekem drugem mestu. Časovni razvoj valovnih funkcij opisuje Schrödingerjeva enačba.

Nekatere valovne funkcije opisujejo verjetnostne gostote, ki se s časom ne spreminjajo. Mednje sodijo tudi mnogi sistemi, ki bi jih v klasični mehaniki obravnavali dinamično. Zgled je elektron v nevzbujenem atomu, ki ga klasično opisujemo kot delec, ki kroži okoli atomskega jedra, v kvantni mehaniki pa ga opišemo s statičnim krogelno simetričnim oblakom verjetnostne gostote, v katerega središču je atomsko jedro.

Z merjenjem določene opazljivke sistema vedno zmotimo valovno funkcijo, tako da ta zavzame eno od t. i. lastnih stanj te opazljivke. Verjetnost za posamezno lastno stanje določa stanje valovne funkcije, tik preden smo jo zmotili. Za zgled si oglejmo delec, ki se giblje v praznem prostoru. Če izmerimo lego delca, bomo dobili neko naključno vrednost x . Na splošno njene natančne vrednosti ne moremo napovedati vnaprej, je pa verjetneje, da bomo izmerili vrednost blizu središča valovnega paketa, kjer je amplituda verjetnostne gostote večja. V trenutku, ko meritev izvedemo, se valovna funkcija »sesede« v lastno stanje, ki je ostro nakopičeno okoli izmerjene vrednosti x . Med samim procesom sesedanja valovne funkcije za slednjo ne velja Schrödingerjeva enačba. Ta je deterministična glede na to, da za valovno funkcijo v nekem trenutku povsem natančno napoveduje njeno vrednost v nekem poznejšem času. Med meritvijo pa je lastno stanje funkcije, v katero se sesede valovna funkcija, določeno verjetnostno in ne deterministično. Verjetnostna narava kvantne mehanike tako izhaja iz samega dejanja merjenja. Ena od posledic sesedanja valovnih funkcij

je ta, da določenih parov opazljivk, kot sta denimo lega in gibalna količina, ne moremo obenem določiti s poljubno natančnostjo. To je znano kot Heisenbergovo načelo nedoločenosti.

Kvantna mehanika je Heisenbergova zasluga. Leta 1927 je postavil svoje temeljno načelo nedoločenosti. Načelo pravi, da nobenemu nebesnemu, atomskemu ali podatomskemu telesu ne moremo istočasno z enako stopnjo natančnosti določiti lege in hitrosti v prostoru. Einstein je kvantno mehaniko zavračal, saj je menil, »*da bog ne kocka*«. S tem je poudaril svoje prepričanje, da naj bog pač ne bi prepuščal, da bi se stvari odvijale zgolj naključno. Angleški teorijski fizik Stephen Hawking, je ugotovil, kakor se je izrazil v svoji zbirki esejev *Kratka zgodovina časa*, »/.../ bog ne samo, da rad kocka, ampak vrže kocko tudi tja, kjer je mi ne moremo več zaznati /.../«. (Hawking, 1988: 7) Prav naključja so glavna domena te zanimive teorije, ki je delovala kot protitež splošni in posebni teoriji relativnosti. Fiziki in drugi znanstveniki so imeli veliko težav, ko so poskušali iznajti teorijo vsega in jim to do sedaj še ni uspelo. Ta teorija naj bi združila vse v eni preprosti fizikalni enačbi, ki bi bila prilagodljiva, in bi se jo dalo uporabiti povsod in bi dala odgovore na vsa znana vprašanja. Združila naj bi tudi vse štiri glavne fizikalne sile.

Kvantna mehanika s tem kaže na možnost preseganja prostora, vsaj pod določenimi pogoji; daleč segajoče vezi lahko tudi obidejo prostorsko oddaljenost. Npr. dve telesi sta lahko prostorsko zelo oddaljeni, a sta, kar se tiče kvantne mehanike, ena sama entiteta. In celo še več, zaradi tesne povezave med prostorom in časom imajo kvantne povezave tudi časovni karakter. Kljub vsem tem velikim odkritjem pa osnovne značilnost časa, namreč da neposredno kaže iz preteklosti v prihodnost, ne pojasnjujeta ne relativnost ne kvantna mehanika. Prepričljivo so glede tega vprašanja napredovale teorije na področju fizike, imenovane *kozmozologije*. (Green, 2006)

Med značilnosti vsakdanjih izkustev, ki se upirajo popolni razlagi, sodi tudi verjetno največja še nerazrešena uganka sodobne fizike, uganka, ki ji je veliki britanski fizik Arthur Eddington rekel *puščica časa*. Samoumevno se zdi, da stvari v času potekajo v določeni smeri – spomini so stvar preteklosti, nikoli prihodnosti; ljudje se staramo, nikoli pomlajujemo ipd.

APLIKACIJE

Kvantna mehanika je uspešna pri razlagi mnogih fenomenov današnjega sveta, ki jih brez nje ne bi mogli znanstveno dojeti, razložiti in razumeti. Vedenje posamičnega podatomskega delca, ki lahko ustvarja najrazličnejše oblike materije – elektrone, protone, nevtrone, fotone in druge –, lahko primerno razložimo le z uporabo kvantne mehanike. Kvantna mehanika je zelo vplivala na razvoj teorije superstrun, ta pa je obetala poenoteno teorijo znanosti. Prav tako je kvantna mehanika pomembna za razumevanje tega, kako se posamični atomi združujejo v molekule. Relativistična kvantna mehanika lahko v principu matematično opiše tudi večino kemijskih problemov, omogoča vpogled v ustvarjanje procesov ionskih in kovalentnih vezi in s tem eksplicitno pokaže, katera molekula je energetsko najbolj primerna za neko drugo in tudi koliko (Bregant, et al., 2010). To je izredno pomembno tudi za naš poskus razlage zavesti z uporabo teorije kvantne mehanike.

8.1.1 Poenotena resničnost

Albert Einstein je s svojima teorijama relativnosti združil prostor, čas in gravitacijo. Uspeh ga je nato navdihnil, da je pričel razmišljati še širše. Želel je poiskati en sam, vseobsegajoči okvir, ki bi zajel vse naravne zakonitosti, a mu ni uspelo. Ta okvir je imenoval *poenotena teorija*. Glavna ovira za uresničitev poenotene teorije je temeljno neskladje med dvema ključnima področjema fizike 20. stoletja: splošne relativnosti in kvantne mehanike. Čeprav ta dva teoretična okvirja največkrat uporabljamo v povsem različnih svetovih – splošno relativnost za velike stvari (npr. vesolje in vse, kar je v njem) in kvantno mehaniko za majhne stvari, kot so molekule in atomi – pa za vsako velja, da naj bi bila univerzalna, da torej lahko delujeta v vseh svetovih. A če ti dve teoriji uporabimo hkrati, njune združene enačbe dajejo nesmiselne rešitve. Čeprav so se Einsteinova prizadevanja v njegovem času zdela iluzorna, pa se sodobni znanstveniki, že približujejo harmonični spojitvi zakonov velikega in majhnega. Pristop, ki naj bi največ obetal, se imenuje *teorija superstrun*.

Teorija superstrun je poskus razložiti tako vse delce kot tudi temeljne sile narave z eno samo teorijo tako, da jih vse obravnava kot nihanje drobne, supersimetrične strune. Teorija superstrun je okrajšava za *supersimetrično teorijo strun* in je verzija teorije

strun, ki upošteva *fermione*¹¹² in *supersimetrijo*¹¹³. Osnovna ideja te teorije je, da je temeljni delček realnosti struna Planckove dolžine ($1,6 \cdot 10^{-35}$ m), ki niha z resonančno frekvenco. Po tej teoriji ima vsaka struna svojo lastno resonanco (lastno frekvenco) oz. harmoničnost. Različne harmonije pa definirajo različne osnovne sile. Napetost strune je reda Planckove sile (10^{44} N).

Teorija superstrun je na začetku ponudila nov odgovor na staro vprašanje: Kaj so najmanjše, nedeljive sestavine snovi? Mnoga desetletja se je konvencionalni odgovor glasil, da snov sestavljajo delci – elektroni in kvarki – ki jih lahko predstavimo kot točke, ki so nedeljive in nimajo ne velikosti ne notranje strukture. Po konvencionalni teoriji, ki je tudi eksperimentalno podprta, se ti delci različno kombinirajo in dajo protone, nevtrone ter širok razpon atomov in molekul, ki na koncu sestavljajo vse, kar poznamo. Teorija superstrun ubere popolnoma drugačen pristop. Ne zanika ključne vloge, ki jo imajo elektroni, kvarki in druge vrste delci, ki so jih pokazali eksperimenti, vendar trdi, da ti delci niso točke. Po teoriji superstrun vsak delec sestavlja majcena nitka energije, okoli 10^{-18} krat manjša od atomskega jedra (skratka veliko manjša, kot lahko trenutno sploh zaznavamo) in ima obliko strune. Prav kakor lahko struna violine niha v različnih vzorcih, da ustvari različne glasbene tone, tudi nitke teorije superstrun nihajo v različnih vzorcih. A ta nihanja ne dajejo različnih glasbenih tonov, pač pa različne lastnosti delcev. Nihanje majcene strune po enem vzorcu ima maso in električni naboj elektrona; po teoriji superstrun je takšna nihajoča struna to, kar smo tradicionalno imenovali elektron. Nihanje majcene strune v drugačnem vzorcu ima lastnosti, na osnovi katerih jo lahko označimo kot kvark, nevtrino ali kakšen drug delec. V teoriji superstrun so združene vse vrste delcev, kajti vsak izhaja iz drugačnega nihanja vzorca, ki ga izvaja ista podstat. Skratka, teorija superstrun združi splošno relativnost in kvantno mehaniko v eno samo skladno teorijo in odpravi usodne neskončne verjetnosti, ki so zaznamovale prejšnje poskuse združitve. Teorija superstrun je ob tem pokazala tudi širino, potrebno za to, da sestavi vse sile narave in vso snov v isto teoretično sestavljanjo. Teorija superstrun je, skratka, glavni kandidat za Einsteinovo poenoteno teorijo, seveda pa, ob upoštevanju revidiranja našega dojemanja prostora in časa. Namesto

¹¹² *Fermioni* so imenovani po italijanskem fiziku Enricu Fermiju in so delci, ki sestavljajo povsem antisimetrična sestavljena kvantna stanja. Vsi osnovni delci so bodisi fermioni bodisi bozoni. V standardnem modelu elektrošibke in močne interakcije so fermioni gradniki snovi, ki sestavljajo npr. protone in nevtrone.

¹¹³ *Supersimetrija* je v fiziki delcev domnevna simetrija, ki povezuje delce z nekim določenim spinom z delci, ki se od njih razlikujejo za polovico enote za spin. Te delce imenujemo superpartnerje. Skratka, supersimetrija povezuje delce in njihove superpartnerje.

treh prostorskih dimenzij in ene dimenzije časa, ki so znane in razumljive vsem, razumevanje teorije superstrun terja devet prostorskih dimenzij in eno časovno dimenzijo, kar pa je vsaj delno razumljivo le ozkemu krogu znanstvenikov, ki se s tem neposredno ukvarjajo. Zato in ker je to pravzaprav onstran naših zastavljenih ciljev, v pričajučem delu teoriji superstrun ne bomo posvečali posebne pozornosti.

8.2 Problem duha in telesa: osnovne rešitve

Po Kimu obstajajo trije pomembni načini razmišljanja o problemu duha in telesa, ki smo jih v takšni ali drugačni obliki že omenili (Kim, 1996). Naj jih ponovimo:

1. Dualizem – katerega glavni predstavnik in zagovornik je bil René Descartes in zagovarja tezo, da je telo materialno, duh pa je nematerialen.
2. Teorija identitete – trdi, da so procesi duha identični s procesi v možganih, torej so duhovni procesi možganski procesi. Vsakemu duhovnemu procesu odgovarja samo ena fizikalna realizacija.
3. Funkcionalizem – mentalno stanje je lahko realizirano na različne načine (večkratna realizacija). To, kar naredi nekaj duhovno, ni odvisno od strukture, temveč od njene funkcije. Vsa duhovna stanja so torej vzročno-funkcionalne oblike. Ta teorija ni omejena le na ljudi, saj trdi, da ima vse primerno funkcijo mentalnosti.

8.3 Redukcija

V nadaljevanju se bomo osredotočili predvsem na teorijo identitete oz. redukcijo, ki se nam pri rešitvi problema duha in telesa zdi najbolj obetavna. Temeljno izhodišče redukcionizma je teza, da je možno neko teorijo ali fenomen reducirati na neko drugo teorijo ali fenomen, npr. neko biološko teorijo na neko fizikalno teorijo. Poznamo različne vrste teoretične redukcije, ki je danes priljubljena predvsem v nevroznanosti.

8.3.1 Modeli teoretične redukcije

Teoretično redukcijo lahko predstavimo s tremi temeljnimi modeli, pristopi, kako nek izhodiščni problem, neko izhodiščno entiteto pretvoriti v temeljno teorijo. Teorijo, ki jo bomo reducirali, bomo dosledno označevali kot *ciljno teorijo* (*C*), teorijo,

na katero bomo želeli reducirati, pa *temeljno teorijo* (*T*). Po letu 1920 so se izoblikovali trije glavni načini razumevanja redukcije, in sicer:

1. Teorijo *C* reduciramo na teorijo *T*, če lahko vse resnice teorije *C* (vključno z njenimi zakoni) prevedemo v jezik teorije *T*. To imenujemo *prevajalni model redukcije* (angl. translation model). Njegova glavna zagovornika sta bila zgodnja logična pozitivista Rudolf Carnap in Otto Neurath. (Carnap, 1934, Neurath, 1931/1959)
2. Teorijo *C* reduciramo na teorijo *T*, če lahko vse zakone teorije *C* izpeljemo iz zakonov teorije *T*. To imenujemo *izpeljevalni model redukcije* (angl. derivation model). Glavna zagovornika tega modela sta bila kasnejša logična empirista Carl Hempel in Ernest Nagel. (Hempel, 1965, Nagel, 1961).
3. Teorijo *C* reduciramo na teorijo *T*, če lahko vsa opažanja, ki jih razložimo s teorijo *C*, razložimo tudi s teorijo *T*. To je *razlagalni model redukcije* (angl. explanation model), katerega glavna zagovornika sta bila John Kemeny in Paul Oppenheim. (Kemeny, Oppenheim, 1956).

Splošen cilj teoretične redukcije je promocija *poenotenja znanosti*, torej da se lahko različne vede reducirajo na temeljno (fiziko), kar z drugimi besedami tudi pomeni, da so vsi fenomeni (tudi kot v našem primeru pojav duha, kot je npr. zavest) identični fizikalnim fenomenom. Tako redukcijo prvenstveno razumemo kot način, kako poenotiti znanost, idejo, o kateri je začel sanjati že Albert Einstein s svojo *poenoteno teorijo*¹¹⁴.

Vsi ti trije modeli redukcije omogočajo, vsak po svoje, poenotenje znanosti, kar pomeni, da lahko zanje uporabljamo enak jezik in resnice vseh znanosti razlagamo z njimi. Prav tako pa znanosti postanejo poenotene tudi, če so lahko zakoni ene teorije izpeljani iz zakonov druge teorije. Do *popolne* (idealizirane) poenotene teorije po Einsteinovem modelu pa je pot zagotovo še dolga, saj smo trenutno sposobni le zmanjševati število potrebnih teorij za razlago različnih pojavov, kar bomo poskušali dokazati.

¹¹⁴ Poenotena teorija je vseobsegajoč okvir, ki naj bi po Albertu Einsteinu zajel vse naravne zakonitosti.

V nadaljevanju si nekoliko podrobneje ogledimo zgoraj omenjene modele. Pri tem je potrebno poudariti, da si ti med seboj niso konkurenčni, temveč se celo dopolnjujejo, z njimi lahko ustvarjamo nekakšne hibridne modele.

REDUKCIJA KOT PREVAJANJE

Carnap opisuje prevajalni model z naslednjo tezo:

»Objekt (ali koncept) rečemo da je reducibilen na enega ali več objektov, če vse izjave o tem prvem objektu lahko pretvorimo v izjave o teh drugih objektih«. (Carnap, 1928/1967: 6). Njegova glavna teza se glasi: »/.../ znanost je enotnost, kjer se vse empirične izjave lahko izrazijo le v samo enem jeziku, vsa stanja dogodkov so samo enovrstna in ugotovljiva z isto metodo«. (Carnap, 1934: 32).

Skratka, Carnap (1963) je zagovarjal možnost redukcije vseh izjav na jezik fizike, tj. na en univerzalni, fizikalni jezik. Takoj se zastavljata dve temeljni vprašanji. Prvo, zakaj bi nekdo želel prevesti vse izjave v samo en skupen, univerzalen jezik? In drugo, kakšno povezavo bo imel ta prevod z idejo poenotene znanosti? Carnapov in Neurathov interes po poenotenju izhaja iz frustracij glede humanističnih in družboslovnih znanstvenih metod, saj dokazovanje te discipline temelji predvsem na subjektivnih metodah, kot sta intuicija in empatija. Za pozitiviste so bile te metode izjemno problematične in sporne. Želeli so jih zamenjati z metodami, ki se uporabljajo v naravoslovju, predvsem v fiziki. Zato sta zagovarjala teze, da potrebujemo teorijo (ali skupek teorij), ki bodo sposobne upoštevati vse pojave. Na ta način bi lahko za vse dogodke, ki so se zgodili, bili sposobni podati kompleksno razlago in za vsak dogodek, ki naj bi se zgodil, bili sposobni napovedati, da se bo zgodil (tako v času kot v prostoru). Iz tega izhajata dve Neurathovi ideji (1931/1959):

- minimalizacija števil izrazov, ki se v teoriji uporabljajo in s tem uvajanje teoretičnih poenostavitev, skratka vpeljevanje »znanstvenega esperanta« v znanstveno komunikacijo;
- opisovanje pojavov različnih znanosti na podoben (skupen) način.

Poleg tega je pomembno poudariti tudi to, da ta skupen jezik omogoča sporazumevanje med vsemi disciplinami. To bi spodbujalo naslednje:

- občutek za neskladja, pojasnjevalne vrzeli (angl. explanatory gap) in s tem odpravljanje neskladij med teorijami;

- odpravljanje odvečnosti in s tem zagotavljanje poenostavitve;
- možnost močnejšega potrjevanja teorij.

Ta pozitivistični koncept redukcije je pomemben, saj omogoča prepoznavanje neodvisnosti posameznih gibanj redukcionizma s stališča videnja fizike kot vede, na katero lahko skrčimo vse ostale vede, kar omogoča sodelovanje med disciplinami, s tem pa povečevanje raziskav za boljše potrjevanje in razumevanje različnih teorij.

REDUKCIJA KOT IZPELJAVA

Carl Hempel (1966) je razumel redukcijo kot pristop, ki mora izpolniti dve nalogi:

- *Pogoje* ciljne teorije prevesti v osnovni, razumljivejši jezik. »Definicije v vprašanih bi težko bile analitične /.../ vendar pa /.../ naj bodo razumljene v manj strogem smislu, takšnem, ki ne zahteva, da imajo enak pomen ali namen, kot definiendum, imajo naj le enako rešitev ali uporabo.« (Hempel, 1966: 103);
- *Zakone* ciljne teorije moramo izpeljati iz zakonov temeljne teorije.

Najboljši znan model redukcije kot izpeljave lahko najdemo v delu Ernesta Nagela *The Structure of Science* (1961). Njegova redukcija je učinkovita, če so zakoni ciljne znanosti *C* logična posledica teoretičnih predpostavk osnovne znanosti *T*. *Ko in če* to zahtevo izpeljemo, lahko ugotovimo, da obstaja samo en osnovni niz principov, ki so potrebni za dosego resnice v obeh teorijah. Za Nagela je eden od ciljev redukcije približati znanost k izhodiščni ideji, saj trdi, da morajo sodobne znanstvene teorije integrirati vsa področja naravoslovnih znanosti v obliko skupnih nizov principov.

V nasprotju s Hempelom pa Nagel ni trdil, da vse redukcije v izhodišču zahtevajo prevod izrazov. Razločeval je med *homogenimi* in *heterogenimi* redukcijami in zagovarjal tezo, da samo v heterogeni redukciji ciljna znanost vsebuje tudi izraze, ki niso zajeti že v temeljni znanosti. To pomeni, da ni nujno potrebno prevesti vseh izrazov iz ciljne znanosti *C* v jezik temeljne znanosti *T*. Na primer, če želimo reducirati psihologijo na fiziko, bomo hitro opazili, da ciljna znanost psihologija vsebuje izraze, kot so 'prepričanje', 'usoda' ali 'bolečina', ki jih ne poznamo v temeljni znanosti, to je v fiziki. V tem primeru moramo prevesti te predpostavke (izraze) v jezik fizike, tj. vzpostaviti povezavo med psihološkimi in fizikalnimi izrazi. Te predpostavke

imenujemo *premostitveni zakoni* in omogočajo izpeljavo zakonov ali teoremov ciljne znanosti iz zakonov ali teoremov temeljne znanosti.

Nagelovsko redukcijo bi lahko zapisali v naslednji obliki:

T_1 in T_2 sta izraza v jeziku temeljne znanosti in C_1 in C_2 pa izraza v jeziku ciljne znanosti. Dogodek T_1 povzroči dogodek T_2 (zakon v temeljni znanosti).

Če je nekaj T_1 , potem je tudi C_1 (premostitveni zakon).

Če je nekaj T_2 , potem je tudi C_2 (premostitveni zakon).

Zato:

Dogodek C_1 povzroči dogodek C_2 (zakon v ciljni znanosti). (Hempel, 1966: 346)

Rezultat tega postopka je redukcija zakona ciljne znanosti na zakon temeljne znanosti. Osnovni problem pri tem je, kakšne zahteve mora izpolnjevati *premostitveni zakon* za legitimno izpeljavo redukcije. Skratka, ključnega pomena je premostitveni zakon s pogojno obliko. Vendar bo izpeljava uspešna, če povezovalne izjave v premostitvenem zakonu niso le pogojne oblike, ampak so identitete.

Za ilustracijo te ideje si oglejmo domnevno redukcijo teorije termične prevodnosti na teorijo električne prevodnosti. Wiedemann-Franzev zakon je enostaven fizikalni zakon, ki predpostavlja, da je razmerje med termično in električno prevodnostjo proporcionalno temperaturi. Če ta zakon uporabimo kot premostitveni zakon, potem lahko na enostaven način sistematično izpeljemo dejstva o električni prevodnosti kovin iz njihove termične prevodnosti. Iz zakona izhaja, da ima kovina določeno električno prevodnost, ki jo predstavlja določena temperatura, *če in samo če* ima kovina takšno in takšno termično prevodnost. Če redukcijo izpeljemo na ta način, z uporabo bikondicionalnih premostitvenih zakonov, lahko skržimo teorijo električne prevodnosti na teorijo termične prevodnosti. Vendar pa bi mnogi temu oporekali. Kot je zapisal Lawrence Sklar:

»Ali ta zakon vzpostavlja redukcijo teorije toplotne prevodnosti na teorijo električne prevodnosti? Nihče tega ni nikoli niti trdil. To kar razlagata oba, električni in termični zakon, kakor tudi Wiedemann-Franzev zakon, je, da je to redukcija makroskopske teorije snovi na

teorijo njene atomske mikroskopske ravni. Korelacijske točke redukcije pa ne zagotavljajo redukcije same». (Sklar, 1967: 119)

S takšno redukcijo pa se ne strinja niti Jaegwon Kim in trdi, da premostitveni zakoni, ki imajo status identitete, niso uporabni kot del pristnih redukcij saj:

»Analitično sporno je, da mora redukcija *poenostavljati*, navsezadnje mora redukcija skržiti /.../ Premostitveni zakon oblike [nekaj je C_1 , če in samo če je T_1] se lahko zahteva na različne načine. Odkar velja izraz [nekaj je C_1 , če in samo če je T_1] se to smatra kot pogojni zakon (angl. contingent law), koncepta [C_1] in [T_1] pa ostaneta različna; zato premostitveni zakon ne vodi do poenostavitve. Nadalje, ker imamo samo pogojno, *že – če* povezane lastnosti [C_1 in T_1 , C_1 in T_1], ostajata različni lastnosti. Iz tega pa zato ne izhaja nobena ontološka poenostavitev ... Če želimo ontološko poenostavitev predstavljene redukcije, moramo nekako poiskati pot za krepitev poenostavitvenih zakonov /.../ v identitete«. (Kim, 1998: 96–97)

Skratka, samo v primerih, kjer imajo premostitveni zakoni status identitet, izpeljava zakonov pomeni redukcijo. To je splošna teza v literaturi s področja filozofije duha in je po navadi tudi cilj teh redukcij; želi poiskati rešitev problema duha in telesa tako, da njen cilj ni poenotenje znanstvenih teorij, kar sta zagovarjala Carnap in Neurath, temveč le ontološka poenostavitev.

REDUKCIJA KOT RAZLAGA

To je zadnji, v preteklosti zelo vpliven model redukcije. Ta model je zgodovinsko povezan s filozofoma Johnom Kemenyem in Paulom Oppenheimom in je podrobneje predstavljen v njunem članku *On reduction* (1956). Splošna definicija te redukcije se glasi takole:

Teorija C se lahko reducira na drugo teorijo T če:

1. slovar teorije C vsebuje tudi izraze, ki niso v slovarju teorije T ;
2. je katerikoli del opazovanja O , ki je razločljiv s teorijo C , razločljiv tudi s teorijo T ;
3. je teorija T prav tako dobro sistematizirana kot teorija C .

Niz opazovanih podatkov O pomeni relativizacijo podatkov, ki jih zahteva razlaga v določenem trenutku poskusa redukcije.

Obstaja kar nekaj elementov te definicije, ki potrebujejo dodatno razlago, še posebej zapis o sistemizaciji, ki sta ga vpeljala Kemeny in Oppenheim. Sistematizacija teorije je merilo, kako dobro je neka kompleksnost v teoriji kompenzirana z dodatno trdnostjo te teorije, da bi lahko razložila ali predvidevala dodatna opazovanja. Za to pa je nujno potrebna definicija točke 3. Če temu ne bi bilo tako, bi nekdo lahko vpeljal opazovanje v temeljno teorijo T in na ta način ustvaril novo teorijo $T+O$, ki bi vplivala na *ad hoc* redukcijo, to pa ni princip, po katerem redukcija deluje. Namesto tega se pričakuje, da ima temeljna teorija vrline, tipične za znanstveno teorijo, in je vsaj tako sistematizirana, kot je ciljna teorija. Ta aspekt redukcije je ključnega pomena. Pove nam, da »vloga teorije ni podati še več dejstev, temveč organizirati ta dejstva v praktičen in obvladljiv sistem«. (Kemeny, Oppenheim, 1956: 11). Cilj redukcije je racionalizirati našo splošno znanstveno sliko sveta in izgnati teorije, katerih področje opazovanja bi lahko bolj sistematično pokrivalo bolj obsežne teorije. Tako je enostavno videti, kako lahko rezultati te redukcije dobro služijo ciljem predhodnikov poenotene znanosti Carnapa in Neuratha.

Tisti, ki dobro poznajo metafiziko in filozofijo duha, vedo, da ta model redukcije ni pogosto omenjen v strokovni literaturi o redukcionizmu. V enem svojih zadnjih del je John Bickle zagovarjal tezo, ki jo je imenoval »*neusmiljen redukcionizem*«, kjer kaže uspešen model redukcionizma pri »molekularni in celični« razlagi kognicije – področju raziskav, ki se nahaja znotraj celične in molekularne nevroznanosti, trenutno najbolj prevladujoče smeri nevroznanosti. (Bickle, 2006) Zaradi predvsem uspešnih in popolnih eksperimentalnih metod in vse bolj natančnih izidov sta vsaj dve trditvi sodobnih filozofov, nasprotnikov redukcionizma, in kognitivnih znanstvenikov napačni:

- da o delovanju možganov ne vemo veliko;
- da z nižje nivojsko nevroznanostjo ne moremo neposredno razložiti kognicije in kompleksnih obnašanj.

Eksperimentalne prakse raziskujejo neposredno molekularne komponente sestavnih delov celic in genske zapise v nevronih in pri tem neposredno merijo določena obnašanja z uporabo splošno priznanih testov eksperimentalne psihologije za študij psiholoških fenomenov, kot so spomin, pozornost ali zaznava. Ta teza neusmiljenega redukcionizma je v mnogočem podobna Kemeyn-Oppenheimovemu modelu redukcije brez zanašanja na dosledno teorijo in dihtonomijo opazovanj.

(Bickle, 2006: 429) Upošteva primer redukcije psihologije na nevroznanost, Bickle opisuje redukcijo kot vpeljavo naslednje enostavne prakse: »Poseganje v nevralnost in sledenje vedenjskim učinkom«. (Bickle, 2006: 425). Njegov pogled je, da je v praksi redukcija dokončna, ko eksperimentalno najdemo uspešno pot interveniranja na kemičnem ali celičnem nivoju, da povzročimo spremembe obnašanja. Tako tudi trdi:

»Če je takšna strategija uspešna, celični ali molekularni dogodki v določenem nevronu na katerem izvajamo eksperiment /.../ *direktno razlagajo* vedenjske podatke. Ti nizi razlag poleg intervencijskih razlag vsebujejo tudi psihološke, kognitivne informacijske in celo kognitivno-nevroznanstvene procese /.../ Ti uspehi pa kažejo redukcijo duha na molekularne poti v nevronih.« (Bickle, 2006: 426)

Tako za Bickla kot za Kemenyja in Oppenheima redukcija deluje, če lahko najdemo teorijo (v prikazanem primeru nevroteorijo, ki opisuje molekularne ali celične mehanizme), ki lahko razloži podatke druge teorije, v našem primeru nekatere aspekte psihologije.

FUNKCIONALISTIČNA REDUKCIJA

Ko danes trdimo, da se dajo posamezne vede reducirati na fiziko, to po navadi pomeni, da je v nekem smislu možno izpeljati model redukcije, torej da lahko neposredno ali s pomočjo premostitvenih zakonov vse zakone posamičnih ved izpeljemo iz fizike. To pa seveda zahteva, da poiščemo fizikalno korelacijo vsem izrazom, ki se pojavljajo v zakonih posebnih (ciljnih) znanosti. 'Celice', 'bolečina', 'zavest', 'denar': vsi ti izrazi morajo imeti svoje fizikalne korelate, tako da lahko formuliramo poenostavitvene zakone, ki bodo usmerjali izpeljavo (na biologijo, psihologijo ali ekonomijo). Redukcionisti v metafiziki in filozofiji duha, ki sledijo smernicam Sklarja in Kima, verjamejo, da morajo ti premostitveni zakoni imeti status identitet. Taka redukcija vseh posebnih (ciljnih) znanstvenih teorij na fiziko je mišljena kot prenos vseh entitet (preko identifikacija) v fiziko in njihova pojasnitev z njenim jezikom, npr. Nagelov primer 'toplota' pomeni 'gibanje molekul'.

Tem izhodiščem in takšni redukciji pa mnogi, npr. Place (1956) ali Smart (1959), nasprotujejo. Smart trdi naslednje:

»Najprej si podrobneje oglejmo tezo, da je občutek možganski proces. Ne obstaja teza, ki npr. 'paslika' (angl. after-image) ali 'bolečino' enači z 'možganskim procesom vrste x'. /.../ To je tako, kot bi 'odsev-podoba' ali 'bolečino' označevali kot proces, proces, ki *naj bi se dogodil* kot

možganski proces. Iz tega pa sledi, da teza ne trdi, da izjavo 'občutka' lahko *prevedemo* na izjavo o možganskih procesih.« (Smart, 1959: 144)

Za Smarta in Placea je torej resnica o redukcionizmu duha nekaj, česar se lahko naučimo le z opazovanjem in ne zgolj z refleksijo pomena psiholoških izrazov. Vsi pa se tudi ne strinjajo, da jzgotlj zagotovitev identitete zadostuje za uspešno redukcijo. Pri tem je problem predvsem v tem, da naj bi bile redukcije tudi *razlagalne*, samo kompletiranje vseh izpeljav pa še ne prikazuje, kako smo premostitveno identiteto dobili. To lahko ponazorimo s Kimovim komentarjem Nagelovske redukcije termodinamike na statiko. Fiziki pri tem uporabljajo naslednji premostitveni zakon:

toplota = gibanje molekul

Ta premostitveni zakon nato dovoljuje izpeljavo termodinamičnih zakonov iz zakonov statične mehanike, ki ureja gibanje molekul. Kimov dvom pri tem je, da tudi če ta redukcija uspe, nekdo še vedno ne bo razumel, kako lahko termodinamika temelji na statični mehaniki, saj v redukciji ni razložena identitetna izjava.

»Zakaj ne bi enostavno rekli, da je ena in ena dve? Ali, dajte mi dober razlog, da tega ne bi izrazili na ta način! Menim, da se moramo potruditi in najprej zagotoviti pozitiven razlog, da nato lahko rečemo, da je nekaj, kar obravnavamo kot različno, v bistvu eno in isto.« (Kim, 1998: 98)

To kar je po mnenju Kima in mnogih drugih, npr. Francka Jacksona (1998) ali Davida Chalmersa (1996) pri tej redukciji potrebno, je, da identitete za svoje temelje potrebujejo to, kar poznamo kot *funkcionalistična redukcija*.

Funkcionalizem je novejši pristop k problemu duha in telesa. Poudarja pomen funkcionalne vloge duševnih stanj, to pomeni, da se osredotoča na vhodne in izhodne podatke ter pri tem izpostavlja odnose med notranjimi stanji. Posamično duševno stanje opredeljuje v razmerju do drugih duševnih stanj ter z njihovimi učinki na obnašanje posameznikov. Najpreprosteje funkcionalizem ponazarjamo s primerjavo z računalniki in njihovim odnosom med strojno in programsko opremo. Pri tem pa je potrebno poudariti, da se funkcionalizem kot teorija pri tem ukvarja le s programsko opremo (funkcijo sistema) in v celoti zanemara strukturo tega sistema. Po tem je še najbolj podoben behaviorizmu.

Funkcionalistična redukcija deluje v dveh fazah:

- V *prvi fazi* upoštevamo fenomene posebne (ciljne) znanosti, ki jo želimo reducirati, in jih 'priređimo' za redukcijo. Npr., če poskušamo reducirati kemične fenomene, kot je vretje, lahko te razlagamo kot lastnost, ki jo ima substanca, ko se pojavljajo mehurčki na njeni površini in s tem povzročajo uparjanje.
- V *drugi fazi* poiščemo lastnosti v temeljni znanosti, ki lahko realizira takšen opis.

Ko je to narejeno, lahko določimo fenomene posebne (ciljne) znanosti s fenomeni temeljne (fizikalne) znanosti. Če nadaljujemo zgornji primer, bi lahko našli v fiziki (in očitno je to poenostavitev) tezo, da so, ko atomi v substanci dosežejo določen povprečni vzgon in je tlak v substanci nižji kot je atmosferski tlak v okolici substance, ti mehurčki sposobni pobegniti s površine. Tako lahko vidimo, kako se delajo mehurčki na površini substance in posledica tega je uparjanje. Ko je ta razlaga podana, lahko sedaj prepoznamo naslednje:

X-vretje je identično z x-tok, ko x-atomi dosežejo določen zagon in je x-notranji tlak nižji kot x-tlak okolice.

S takšno razlago postane jasno, zakaj in kako smo to identiteto dobili. To je zato, ker je drugi del redukcije samo fizikalni fenomen, ki zahteva, da x ve, kako se vrenje izvaja, pa je razloženo v prvem delu redukcije, tj. s tem, da je x-tok.

Funkcionalistična redukcija torej dopušča vpogled v to, kako smo v nekem primeru dobili identiteto. Identitetno teorijo Smarta in Placea lahko nadomestimo s funkcionalistično redukcijo, tako da razložimo premostitvene zakone s statusom identitete.

ARGUMENT IZ VEČVRSTNE REALIZACIJE

Argument iz večvrstne realizacije je zgodovinsko povezan s Putnamom in Fodorjem (Putnam, 1975, Fodor, 1975). Trdila sta, da ni mogoče najti resnične izjave identitete v obliki, ki jo zahtevajo redukcije specialnih znanosti. Nazorneje si to oglejmo na primeru redukcije psihologije na fiziko. Če želimo, da bo ta redukcija uspešna, potem moramo najti fizikalno korelacijo vseh psiholoških lastnosti, tj. povezavo vsakega

psihološkega pogoja z ustreznim fizikalnim pogojem, npr. z neko fizikalno lastnostjo F . Zato mora obstajati resnična identitetna izjava naslednje oblike:

Za vse X (X je bolečina = X je fizična lastnost F)

ali bolj splošno

za vse X (X je lastnost specialne znanosti S = X je lastnost fizikalne znanosti F)

Kot je izpostavil Putnam, je velik izziv za vsakega redukcionista najti ustrezno fizikalno lastnost, ki bo služila zastavljenemu namenu. Putnam in Fodor sta namesto redukcionizma zagovarjala funkcionalizem: bolečina je funkcionalna lastnost, ki ima mnogo različnih zavedanj, vendar nobeno od teh ni fizikalna lastnost, s katero bi lahko lastnost bolečine identificirali.

8.3.2 Novi trendi v redukciji: algoritemska redukcija

Poskus vzpostaviti dinamično kognitivno znanost kot teoretični okvir celotne kognitivne znanosti ima obilico skupnega z Nagelovskim modelom redukcije. Predvsem poskus unifikacije različnih nivojev, v našem primeru kognitivnega in nevronskega, tako da za oba lahko uporabljamo unificiran slovar (jezik), to potrjuje. V nasprotju z Nagelovim pogledom na redukcijo pa Kelso (1995) ni zagovarjal takšne, imenujmo jo poenotene, kognitivne znanosti. Zagovarjal je tezo, da kognitivnih dogodkov ni mogoče razložiti s terminologijo (jezikom) živčnih dogodkov in pri tem uporabljati 'jezikov' različnih sistemov diferencialnih enačb v funkciji premostitvenih zakonov. Menil pa je, da je mogoče enaka teoretična orodja, enake sisteme diferencialnih enačb, uporabljati za razlago tako kognitivne kot tudi nevrnalne dinamike. To je, kot smo videli, prikazal na primeru gibanja prstov, kjer z enakimi enačbami definira tako motorično kot tudi nevropsihološko ozadje tega gibanja. Če tezo zastavimo ostreje, lahko trdimo, da z istimi diferencialnimi enačbami lahko ne zgolj opišemo različne dogodke, ampak da te diferencialne enačbe lahko tvorijo določene algoritme, ki opisujejo celotne procese, in da z istimi procesi (algoritmi) lahko opisujemo različne dogodke, v našem primeru tako kognitivne kot tudi nevrnalne procese. Na ta način postane redukcija *algoritemska* in je bolj natančna ter bližja Nagelovski teoretični redukciji.

V nekoliko spremenjeni obliki lahko Kelsovo idejo tako uporabimo za izhodišče našega dokazovanja. Seveda obstajajo razlike med realnim (konkretnim, kompleksnim) dinamičnim sistemom, kot je npr. sončni sistem ali Wattov centrifugalni regulator, ki je pogosto uporabljen za dokazovanje dinamičnih sistemov, in matematičnim, abstraktnim modelom, kot sistemom diferencialnih gibalnih enačb, s katerimi opisujemo te realne sisteme. Zanimivo pri tem pa je, da lahko več različnih dinamičnih sistemov opišemo z enakimi (celo istimi) matematičnimi modeli, t. j. *enake* diferencialne gibalne enačbe lahko opišejo serijo *različnih* realnih sistemov, kar smo videli v predhodnih poglavjih. Kelsova ideja je torej nekaj, kar lahko v našem primeru uporabimo za povezavo duha in možganov (kognitivne in nevrnalne procese), saj, kot smo pokazali, jih lahko ne-le opišemo z diferencialnimi enačbami enake oblike, temveč uporabimo za to tudi enak sistem, postopek (algoritem) opisovanja.

Izhajajoč iz Kelsove ideje lahko tako zagovarjamo tezo, da lahko reduciramo kognicijo na nevrofiziološki nivo s pomočjo enakih matematičnih modelov, s pomočjo enakega opisa, enakega algoritma opisa osnovne dinamike, iz česar lahko izpeljemo novo redukcijo, imenovano *algoritemska redukcija* (AR).

V čem pa bi bila algoritemska redukcija drugačna od vseh v tem poglavju opisanih redukcij? Zakaj ni niti Nagelovska niti funkcionalistična? Nagelovski model redukcije upošteva redukcijo kot povezavo dveh znanstvenih teorij, temeljne teorije in ciljne teorije, kjer lahko ciljno teorijo reduciramo na temeljno samo, če lahko zakone ciljne teorije reduciramo na zakone temeljne teorije s pomočjo premostitvenih zakonov. Funkcionalistična redukcija pa temelji na opisovanju mentalnih lastnosti kot funkcionalnih lastnosti, ki jih lokalno realizirajo različne fizične baze. Kakšna je predlagana algoritemska redukcija (AR)?

AR je hibridna redukcija, kombinacija Nagelovske in funkcionalistične redukcije. Opišemo jo lahko kot dvostopenjski proces, pri katerem v prvem koraku za osnovo uporabimo funkcionalistično redukcijo in nato, v drugem koraku, Nagelovsko redukcijo. Oglejmo si princip AR podrobneje. V prvem koraku morajo biti kognitivna stanja opisana v pojmi njihovih vzročno-funkcionalnih zvez (kar je prvi korak v funkcionalistični redukciji). V naslednjem, drugem, koraku pa zapišemo diferencialne enačbe, ki služijo kot »premostitveni zakonik« med psihologijo in nevrofiziologijo (kar je prvi korak Nagelovske redukcije). To potem pomeni, da

lahko duh reduciramo na možgane in iz njih, preko matematičnih algoritmov, izpeljemo resnice o duhu. AR zaključí, da so npr. bolečina, prepričanja ali ljubezen možganska stanja, ki so opisana z ustreznimi diferencialnimi enačbami ali matematičnimi algoritmi.

Ta dvostopenjski proces redukcije kognitivnih sposobnosti bi lahko bil to, kar je predlagal Kelso. Na kratko: Če torej lahko najdemo (in to lahko, kar smo dokazali v predhodnih poglavjih) enake diferencialne enačbe, ki krmilijo višje nivojske (vedenjske) in nižje nivojske (nevrofiziološke) procese, lahko ugotovimo, da so ti izomorfni, in zato lahko kognitivno obnašanje reduciramo na možganske procese, oziroma duha na telo.

8.4 Sklep

Že v četrtem poglavju smo nakazali, da sta za inteligentno delovanje osnovna kriterija učljivost (in na podlagi tega adaptivnost) ter zavest. Zvest ostaja ena od največjih še nerešenih skrivnosti in mnogi menijo, da je prav zavest tista, zaradi katere je problem duh—telo tako težko rešljiv. Skrivnost je predvsem v tem, da ne vemo, kako bi jo uskladili z vsem ostalim znanjem, tj. kako bi odgovorili na vprašanja kot so: Zakaj zavest obstaja, kaj je njena funkcija, kako lahko vznikne iz živčnih procesov v možganih itd. Zvest so v zgodovini razlagali na različne načine. Po eni strani se zdi, da je zavest (in duševnost) povsem drugačna od vsega ostalega v svetu in mora zato imeti poseben ontološki status, po drugi strani pa sprejemamo, da je človek biološko bitje in da je del narave ter da so procesi, ki se dogajajo v njem, vključno z duševnimi procesi, tako tudi naravni procesi. Ker zagovarjamo naturalistični pristop k reševanju problema duh—telo, se nam je še najbolj obetavna razlaga zavesti zdela kvantna razlaga, ki daje možnost redukcije duha na kvantno raven. Tako smo izhajajoč iz Kelsove ideje izpeljali novo redukcijo, imenovano *algoritemska redukcija (AR)*, kjer smo dokazali, da lahko reduciramo kognicijo na nevrofiziološki nivo preko enakih matematičnih modelov, tj. s pomočjo enakega opisa in enakega matematičnega algoritma osnovne dinamike, ki rabi kot premostitveni zakon med ciljno in temeljno znanostjo, kar je glavni sklep te naloge.

9 Sklep

Že starogrški filozofi so si postavljali vprašanja o naravi duše in telesa ter o povezavi med njima. Kakšna je vloga jezika? Ali duša preživi smrt telesa? Kakšen je odnos med duševnimi in telesnimi procesi? Ali smo svobodni v svojih odločitvah? Kako vemo, da nekaj vemo? Pomembni antični filozofi Platon, Aristotel Demokrit in Lukrecij so v svojih delih nanja odgovarjali na različne načine, njihovo delo pa so nadaljevali Descartes, Spinoza, Hume, Kant in mnogi drugi. Tudi danes, v tehnološki dobi, si sodobni raziskovalci s področja filozofije, kognitivne znanosti, nevrobiologije, nevroračunalništva in umetne inteligence zastavljajo podobna, morda le nekoliko bolj s tehnološkim razvojem obarvana vprašanja. Med njimi so tudi vprašanja o korelaciji med človekom in strojem ter implikacijami, ki jih to ima za reševanje tradicionalnih problemov znotraj filozofije duha, tj. problem duha in telesa, problem mentalne vzročnosti in problem zavesti. S terminom 'stroj' smo v delu poimenovali vsak sistem, ki v sodobnem svetu poskuša človeku olajšati delo in življenje ali ga celo nadomeščati, tako človeka kot fizično bitje (npr. stroj kot humanoidni robot, ki lahko izvaja različne fizične aktivnosti, ...) kot tudi človeka kot mentalno (predvsem razmišljujoče) bitje, npr. računalnik, ki simulira človeške kognitivne procese.

V nalogi smo se posvetili iskanju in utemeljevanju vzporednic med naravno in umetno inteligenco na epistemološkem in ne toliko na hevrističnem nivoju. Pri tem smo se v izhodišču nekoliko posvetili paradigmam in strategijam učenja kot enem

od pomembnih delov inteligentnega delovanja, ki smo jih predstavili v prenovljeni kibernetični didaktiki. Kibernetična didaktika temelji na kibernetičnih načelih in spoznanjih o krmiljenju in reguliranju tehničnih, organskih in družbenih sistemov. Med slednje uvrščamo tudi učni sistem. Če naj računalniški programi (inteligentni agenti) delujejo inteligentno, morajo vsebovati predstavo o svetu in o tem, kako si te predstave razlagamo, ter kako delujejo osnovne zakonitosti v njem. Vsak ustvarjalec umetne inteligence mora vedeti, kaj je znanje in kako do njega pridemo. Prav zaradi tega se nekateri tradicionalni filozofski problemi pojavljajo tudi v umetni inteligenci in obratno, njihove rešitve znotraj umetne inteligence lahko uporabimo pri reševanju tradicionalnih filozofskih problemov. Tako smo v nalogi izhajajoč iz matematičnih algoritmov za učenje predstavili t. i. algoritemsko redukcijo mentalnega, pri kateri bi človekovo duševno življenje po vzoru dinamičnih modelov učenja reducirali na matematične algoritme.

Izhajajoč iz teze, da je korelacija človek – stroj možna, v nalogi zagovarjamo ter dokazujemo tezo, da je tudi možno ustvariti inteligenten avtonomen adaptiven sistem (to je lahko tako program, računalnik ali humanoidni stroj – robot), ki bi se bil sposoben učiti, se prilagajati novo nastalim okoliščinam in ob tem izvajati tudi (kritično) samoevalvacijo. Tu smo prišli do točke, ko smo si zastavili naslednje vprašanje, ki je bilo osnovno izhodišče zastavljenega problema: *Ali lahko človeški duh formaliziramo in reduciramo na jezik znanosti in, če da, kar smo tudi dokazali, kakšna orodja potrebujemo za to?* Skratka, v delu smo dokazali, da se duh da formalizirati, zapisati v fizikalni obliki kvantne zavesti ter nato z uporabo teorije dinamičnih sistemov koncepta duha reducirati nanje.

Najprej smo se na osnovi dileme o redukciji človeškega duha osredotočili na tradicionalna filozofska izhodišča o duhu in se pri tem na temelju zgodovinskega razvoja posvetili novjšim filozofskim pogledom na mentalno ter se dotaknili vprašanj redukcije in emergence. Temu je sledil oris modeliranja s poudarkom na kognitivnih procesih, pri čemer je bilo izpostavljeno, da na področju modeliranja duha obstajajo trije prevladujoči pristopi – modeli: diskretni, zvezni in dinamični oz., povedano z drugimi besedami, simbolni, konekcionistični in dinamični. Predstavili smo nekaj idej s področja filozofije duha, s posebnim poudarkom na naturalističnem/redukcionističnem pristopu, ki ga v nalogi zagovarjamo, in premisi, da je duh lahko reduciran, in to tako, da ga zapišemo z jezikom znanosti, fizike, nevroznanosti ali z jezikom drugih naravoslovnih znanosti. Pri tem izpostavljamo

predvsem dva možna pristopa, hibridnega, kombinacijo simbolnih in konekcionističnih pristopov, katerih najpomembnejši predstavnik je ACT-R, in dinamičnega, ki se intenzivno razvija v zadnjih letih in predstavlja rdečo nit te naloge.

V delu sta izpostavljeni dve diametralno nasprotni stališči do narave duševnosti, dva sklopa domnev. Po eni strani se nam zdi, da je duševnost povsem drugačna od vsega ostalega v svetu in mora zato imeti poseben ontološki status, po drugi strani pa sprejemamo, da je človek biološko bitje in je del narave, torej so tudi procesi, ki se dogajajo v njem, vključno z duševnimi procesi, naravni procesi. Pri domnevah prvega pristopa filozofi večinoma izhajajo iz svojih vsakdanjih izkušenj, domneve drugega pristopa pa so vezane predvsem na znanstvena spoznanja. Zagovorniki naturalistične teorije (domneve drugega pristopa), med katere sodimo tudi sami, nekakšni znanstveni optimisti, moramo poiskati razlago, kako so ti pojavi mogoči v naravi. Zagovorniki dualistične teorije (domneve prvega pristopa) pa menijo, da ne obstajajo prepričljivi dokazi za naturalizem mentalnega in zagovarjajo razlage, ki vključujejo nematerialnost in drugačnost duševnega stanja. Če se osredotočimo na drugi pristop, ki ga v delu zagovarjamo, pa so lahko razlage emergentne ali redukcionistične.

Poudariti moramo, da postaja v sodobni kognitivni znanosti nevroznanost vse pomembnejša in potrebnejša za razumevanje, razlaganje in tudi za modeliranje kognitivnih sistemov. Tako smo se v delu dotaknili tega področja, ki se je zaradi izjemno močnega tehnološkega razvoja na področju empiričnih raziskav in metod zbiranja podatkov (fMRI, EMR, BOLD, ...) v zadnjih desetletjih razvijalo izrazito intenzivno. Danes nam nevroznanost omogoča dokaj natančen vpogled v delovanje možganov. To je temelj razvoja raznovrstnih metod modeliranja procesov in elementov tako celotnega živčnega sistema kot njegovega centralnega dela, to je možganov. Vpogled sam pa zagotovo ni dovolj, saj nam ne omogoča razumevanja procesov, ki tam tečejo. Zaključimo lahko, da so možgani v svojem bistvu bionevronska mreža, katere osnovni gradniki so nevroni. Posvetili smo jim pomemben del dela, raziskovali smo, kako se informacije v možganih (nevronih) prenašajo, kako se signali pretvarjajo od električnih do kemičnih reakcij in kako se pri tem ustvarja ustrezen akcijski potencial, ki definira dostopnost do določenih informacij in verjetnost njihovega priklica v spomin. To pa je osnova kakršnegakoli učenja.

Na tej osnovi smo prešli k bistvu raziskovalnega problema in se poglobili v kognitivno znanost in povezanost le-te s kognitivno teorijo učenja ter s procesi in modeli procesiranja informacij. Učenje in pridobivanje novih znanj in veščin je poleg zavesti in svobodne volje eden od ključnih dejavnikov, ki ločujejo inteligentne od neinteligentnih sistemov. Zato smo se posvetili predvsem metodam pridobivanja novih znanj s stališča, kako kognitivna nevroznanost obravnava učenje in spomin. Pri tem smo izpostavili teorijo sistemov in kibernetično didaktiko, tako da pojmuemo učni proces kot svojevrsten dinamičen in odprt sistem, ki ga sestavljajo povezani deli, odnosi med njimi in njihova struktura.

Vseskozi smo poudarjali, da je tako kot nevroznanost tudi kognitivna znanost izrazito interdisciplinarna, in pogled iz samo enega zornega kota, zornega kota ene vede, ne more dati pravih rezultatov, najpogosteje pa daje celo popolnoma napačne in zavajajoče rezultate. Ne zadostuje, da imamo parcialne rešitve, na primer izsledke empiričnih raziskav, temveč je ključnega pomena razumevanje vseh pojavov in njihovih vzajemnih povezav. Temeljna opredelitev kognitivne znanosti – raziskovanja o tem, kako človek procesira informacije – je tako široka, da dopušča različna tolmačenja in pristope h kognitivni arhitekturi. Z izrazom arhitektura smo poudarili, da pri modelu ni pomembno le njegovo obnašanje, temveč tudi strukturne lastnosti in postopek modeliranja sistema. Pri tem ni nujno potrebno, da so vse lastnosti fizikalne, lahko so tudi virtualne, take, ki jih vključimo v fizikalne stroje (npr. v možgane ali računalnik). Kognitivne arhitekture so lahko simbolno računalniške, konekcionistične v obliki nevronske mreže in dinamične.

Vsak od pristopov h kognitivni arhitekturi poudarja posamezne značilnosti kognicije in se kaže kot delno primerna arhitektura zanje. Tako se simbolni klasični modeli kažejo kot precej uspešen pristop za modeliranje višjih kognitivnih funkcij, ki zahtevajo uporabo jezika, konekcionistični modeli pa pri prepoznavanju vzorcev. Pristop teorije dinamičnih sistemov in »utelešene kognicije« nas opozarjata, da moramo najprej razumeti biološke rešitve za preproste naloge, kot so npr. občutljivost na svetlobo, premikanje in izbira dejanja.

Dalje nakazujemo, da sta za inteligentno delovanje osnovna kriterija učljivost (in na podlagi tega adaptivnost) ter zavest. Zavest ostaja ena največjih še nerešenih skrivnosti in mnogi menijo, da je prav zavest tista, zaradi katere je problem duh–telo tako težko rešljiv. Zavest so v zgodovini tolmačili na različne načine. Po eni

strani se nam zdi, da je zavest (in duševnost) povsem drugačna od vsega ostalega v svetu in mora zato imeti poseben ontološki status. Po drugi strani pa sprejemamo, da je človek biološko bitje in da je del narave ter da so procesi, ki se dogajajo v njem, vključno z duševnimi procesi, tako tudi naravni procesi. Ker v tem delu zagovarjamo naturalističen pristop, se nam je še najbolj obetavna razlaga zavesti zdela kvantna razlaga in možnost redukcije le-te preko matematičnih algoritmov na kvantni nivo (algoritemska redukcija). Takšen način obravnave zavesti pa se elegantno vključuje tudi v dinamični pristop razlage duha, ki temelji na algoritemski redukciji.

Že na začetku smo nakazali, da obstajajo teorije, ki podpirajo diskretno, in teorije, ki podpirajo zvezno modeliranje kognitivnih procesov. Tako obstajajo trije ločeni pristopi modeliranja duha, to so *kognitivno simbolni* sistemi, ki temeljijo na simbolnem računanju in informacijsko procesnem modelu, *kognitivno konekcionistični*, ki za izhodišče modela uporabljajo nevronske mreže, in sistemi *modeliranja na osnovi teorije dinamičnih sistemov*, ki zastopajo prepričanje, da se da kognitivni razvoj bolje predstaviti s fizikalnimi teorijami, in da so diferencialne enačbe najprimernejše orodje za modeliranje človeškega vedenja. Poleg tega pa se predvsem v zadnjih desetletjih vse pogostejše pojavljajo mešani modeli, ki smo jih poimenovali hibridni sistemi. Ker menimo, da so se simbolni sistemi v samostojni obliki že nekoliko izpeli in da z njimi pač ne moremo opisovati vse potrebne kompleksnosti kognitivnih procesov, smo se v delu osredotočili predvsem na konekcionistične in dinamične sisteme, simbolne sisteme pa obravnavamo le v povezavi s konekcionističnimi, kot delom hibridnih sistemov. V nadaljevanju dela smo se tako najprej posvetili kognitivnim konekcionističnim modelom. Definirali smo, kaj so biološke in kaj formalne nevronske mreže, pregledali zgodovinski oris uporabe teh modelov v kognitivni znanosti, kako nevroznanost obravnava nevronske mreže, predvsem formalne, in na osnovi njih modelira umetne možgane. Nakazali smo smernice razvoja in prikazali možnosti izgradnje silicijevih nevronskih mrež kot osnove umetnih možganov. In na koncu smo izpostavili paradoks, da danes še vedno poskušamo modelirati delovanje duha (delovanje možganov), ki izkazuje izrazito zvezni proces delovanja, na konvencionalnih, digitalnih računalnikih, ki delujejo izključno na diskretnem principu.

Potem smo predstavili kombinacijo konekcionističnih in simbolnih sistemov, ki jih imenujemo tudi hibridni sistemi. Fizikalni simbolni sistemi, ki jim v delu nismo posvetili posebne pozornosti, so bili v zgodovini prvi, ki se jih je prijel vzdevek

inteligentni sistemi. Ogledali smo si le nekaj osnovnih primerov in na temelju le-teh poskušali razložiti principe njihovega delovanja. Na kratko smo opisali trenutno najbolj aktualne simbolne arhitekture, njihovo delovanje pa smo prikazali s simbolno arhitekturo SOAR, za kar smo uporabili nalogo, ki zahteva za rešitev problema uporabo verjetnostnega znanja. Na začetku je SOAR uporabil preudarno sklepanje, s katerim je izračunal verjetnost pričakovane vrednosti, ki jo je nato lahko pri odločanju kombiniral s hevrističnim znanjem in modeli nasprotovanja. Pokazali smo, da so za takšne naloge *gradniki* za posredovanje informacij izredno primerni, saj lahko z njihovo pomočjo sistem kombinira različne tipe znanj in ustvarja nova pravila, nato pa s pomočjo utrjevalnega učenja ta pravila prilagaja tako dolgo, da z njimi doseže znatne izboljšave želenih lastnosti. Za prikaz delovanja je bila igra s kockami izjemno primerna in je dokazala, da so simbolni kognitivni sistemi zelo uspešni za reševanje takšnih in podobnih problemov. Danes pa se večina strokovnjakov strinja, da le-to ni dovolj za inteligentno delovanje.

Izpostavili smo nekatere osnovne kontraste med simbolnim in podsimbolnim, konekcionističnim nivojem in ugotovili, da je *gradnik* sam po sebi simbolni način posredovanja informacij. Zgrajen je iz osnovnih simbolov in tudi njegov način delovanja znotraj arhitekture je določen s simbolno strukturo. Vendar pa sam od sebe ne more narediti ničesar, saj je shranjen znotraj deklarativnega spomina oz. deklarativnega modula. Da bi lahko deloval znotraj celotnega sistema (kot je to v možganih), mora biti priklican iz spomina in nameščen v medpomnilnik. Ta sistem, ta priklic, pa usmerja podsimbolna enačba, ki določa aktivacijski nivo vsakega gradnika kot funkcijo njegove uporabnosti v preteklosti in koristnosti za trenutno situacijo, kar ponazarja nevronske povezave. In to je tudi osnovni problem, zakaj simbolni sistemi ne morejo izkazovati inteligentnega delovanja samostojno. Ta enačba priklica je podsimbolna zato, ker je popolnoma neodvisna od gradnikove notranje simbolne strukture. Iz tega izhaja, da vsaj bližnja prihodnost zagotovo ni ne črna in ne bela oz. ni ne mrežna, konekcionistična, ne fizikalno simbolna, temveč je nujna njuna kombinacija. To hipotezo dokazujemo na osnovi trenutno verjetno najbolj kompleksne in tudi najbolj delujoče kognitivne hibridne arhitekture, na primeru ACT-R, ki nam omogoča simuliranje vrste kognitivnih procesov in je po svojem principu in po svojem delovanju danes še najbližji delovanju človeških možganov, kar potrjujejo tudi empirične raziskave v nevroznanosti.

ACT-R omogoča možnost razmišljanja o modelu celovitega organiziranja duha, ki kombinira elemente dveh popolnoma različnih konceptov procesiranja informacij, simbolnega, ki je neposredno povezan s hipotezo fizikalnega simbolnega sistema na eni strani in podsimbolnega, ki je povezan z modelom umetnih nevronske mreže, na drugi strani.

Res je, da avtorji ACT-R na ta tip podsimbolnega procesiranja v resnici niso, vsaj ne v začetku, gledali na opisani način, tj. kot na implementiranje koncepta umetnih nevronske mreže, oziroma kot na prevlado umetnih nevronske mreže nad *simbolnim* procesiranjem informacij, zato te arhitekture ne moremo neposredno preslikati na razliko med fizikalnim simbolnim sistemom in umetnimi nevronske mrežami. Vendar pa menimo, da se kljub temu iz tega lahko naučimo vsaj dvoje, in sicer da pravilno razmišljanje o modularni organiziranosti duha zahteva razmišljanje tudi o tem, kako bi lahko različni moduli sodelovali med seboj pri procesiranju informacij, in kako bi lahko različni deli arhitekture duha izkoriščali različne modele procesiranja informacij. Nekateri naloge že po naravi težijo bolj k simbolnemu načinu reševanja, medtem ko druge bolj k podsimbolnemu, če na problem seveda ne gledamo iz drugega zornega kota, iz zornega kota teorije dinamičnih sistemov. Znanstvene razprave o teh modelih procesiranja informacij zagotovo niso in ne morejo biti v okviru načela vse ali nič, samo eno ali samo drugo.

S teorijo dinamičnih sistemov smo zaokrožili področje kognitivnega modeliranja in zastavljenega problema. S pomočjo te teorije smo definirali prostor stanja in atraktorje v njem ter osnove teorije nihanja in se v nadaljevanju posvetili odnosom med originalom in dinamičnim modelom, s posebnim poudarkom na matematičnem modeliranju človeškega telesa in duha. Izhajajoč iz teze, ki jo zagovarjamo, da je vesolje dinamičen sistem, je tako tudi vsak podsistem ali element dinamičnega sistema dinamičen sistem. To redukcijo lahko izvajamo vse do najmanjšega delca. Posledično sledi, da je tudi človek in vse, kar se nanj nanaša oz. iz česar je človek sestavljen, dinamičen sistem, in tako lahko za razlago delovanja človeka uporabljamo isto teorijo, tj. teorijo dinamičnih sistemov.

Nato splošno teorijo dinamičnih sistemov prevedemo na hipotezo dinamičnih sistemov v kognitivni znanosti in poglavje nadaljujemo z zagovorom in dokazovanjem možnosti redukcije duha na dinamične sisteme ter z dokazovanjem uporabnosti predlagane teorije tako na enostavnejših primerih kot tudi na

kompleksnejšem primeru dinamičnega modela možganskih aktivnosti. V delu zagovarjamo tezo, da je možgane na teoretičnem nivoju mogoče modelirati kot kognitivne dinamične sisteme. Tezo o modeliranju duha potrdimo na kompleksnem, a še vedno poenostavljenem primeru rešitve modeliranja možganskih aktivnosti z uporabo novih oblik valovnih enačb in s prikazom možne izvedbe redukcije na osnovno znanost, fiziko (kvantno mehaniko), in z njo povezane algoritemske redukcije. V enodimenzionalnem prostoru je rešitev odvisna od Lipschitzove zveznosti za začetne podatke in je tako edinstvena. To pomeni, da so odzivi na majhne motnje majhni kot se le da, kar je skladno tudi s fizikalnim glediščem. Model je za to razlago poenostavljen, z vpeljavo dodatnih spremenljivk pa bi lahko postal bolj realističen, kar pa za našo razlago in dokazovanje ni nujno potrebno. Predstavljeni model predpostavlja:

- da je geometrija možganov enodimenzionalna (dvodimenzionalni model bi bil sicer bolj dosleden, vendar vpeljuje dodatne probleme);
- zvezna funkcija $f_{ke}(x, X) = \frac{1}{2\sigma_e} e^{-\frac{|x-X|}{\sigma_e}}$ med nevronskimi mejami je homogena, saj ima pravzaprav več kot 50 % nevronskih mej homogene povezave, tako da bi lahko enostavno dodali dodaten izraz, ki bi predstavljal heterogenost;
- zunanji vnos p se sklada z $H^2([0, T], L^\infty(\Omega))$, v resnici pa je p lahko manj gladek ali bolj nelokalen.

Iz prikazanega je razvidno, da sicer diferencialne enačbe, s katerimi opisujemo elemente kognitivnih sistemov same po sebi niti niso izrazito zahtevne, vendar pa je zahtevno njihovo reševanje, saj se lahko reševanje izvaja le z uporabo numeričnih metod, kar posledično zahteva ogromno računalniško moč in spominski prostor. S tem pa stopamo v »daljno« prihodnost (po trenutnih predvidevanjih v leta 2025 do 2050) in na področje napovedovanja neznanega.

Na koncu smo se dotaknili še vprašanja, ali lahko na diskretnih sistemih zares uspešno obravnavamo zvezne sisteme, kar možgani zagotovo so, in vprašanja, ali je izdelava umetnih možganov v prihodnosti realna. Kot odgovor na prvo vprašanje smo nakazali možne potencialne (bolj ali manj verjetne) rešitve z različnimi nedigitalnimi pristopi k računanju in računalništvu od bioračunalnikov, delujočih na

principu DNK in proteinov, pa do kvantnih računalnikov, katerih osnovni principi obdelav podatkov izhajajo iz kvantne teorije in kvantnega paralelizma.

Glede drugega vprašanja pa menimo, da obstajajo trdni in utemeljeni razlogi, da so napovedi ustvarjanja umetnih možganov oz. kopiranja možganov v bližnji prihodnosti preveč optimistične. Še posebej zato, ker trenutno ne obstajajo potrebna znanja niti o možganih niti o kogniciji in prav tako ne vemo, kakšno računanje in kakšni računalniki bi za to bili potrebni. Danes določena znanja sicer imamo, problem pa je lahko tudi v tem, da morda teh znanj niti ne razumemo v popolnosti niti ne vemo, ali jih razlagamo in uporabljamo pravilno.

Poleg teh filozofskih dvomov pa so tu še tehnološki problemi, problemi povezani z računalniško močjo in njihovim izkoristkom. Možgani porabijo približno 10–20 W moči, v nasprotju s tem pa bi superračunalniki lahko porabili celo več deset tisoč MW, če pri tem upoštevamo *Landauerjevo limito*, ki je 3.5×10^{20} operacij/sekundo/W pri sobni temperaturi in ob upoštevanju *Bekensteinove meje*, ki je sicer po predvidevanjih le zgornja meja informacij na nekem podanem končnem območju (npr. v možganih ta limita znaša 2.58991×10^{42} operacij – bitov). Izhajajoč iz tega pa se pojavi problem sproščene energije, največ v obliki toplote, saj vemo, da je že pri današnjih procesorjih, ki so po svojih zmogljivostih in porabljeni energiji še »svetlobna leta« od opisane meje, osnovni problem hlajenje oz. odvajanje sproščene energije.

Nasproti vsemu temu in kot odgovor na izhodiščno vprašanje pa so nekateri filozofi, med njimi tudi David Hume, Immanuel Kant in Charles Saunders Pierce, glede mej človeškega razumevanja menili, da obstajajo stvari, ki jih nikakor ne moremo vedeti, vsaj ne na trenutni stopnji razvoja človeške rase.

Na koncu lahko poudarimo, da so poskusi, da bi izdelali inteligentne stroje, povzročili revolucijo v mišljenju o mišljenju. Kot pravijo kognitivni znanstveniki, predstavlja računalnik naše novo zrcalo. Računalniška metafora je odprla nove, revolucionarne poti v raziskovanju človekovih spoznavnih procesov. Možnost modeliranja človekovega spoznavanja je namreč zares omogočil šele pojav računalnika. Pred tem so obstajali zgolj opisi miselnih procesov, ki so temeljili le na notranjem uvidu, na introspekciji. V nalogi smo razvili in dokazali možnost nove vrste redukcije, algoritemsko redukcijo, ki zagotavlja pozitivne odgovore na

zastavljena vprašanja. Algoritemska redukcija je izviren in v tej nalogi prvič uporabljen način dokazovanja. Opišemo jo lahko kot dvostopenjski proces, pri katerem v prvem koraku za osnovo uporabimo funkcionalistično redukcijo in nato, v drugem koraku, Nagelovsko redukcijo. V prvem koraku z dinamičnimi enačbami opišemo kognitivna stanja in to v pojmi vzročnih zvez pomembnih spremenljivk kognitivnega sistema (kar je tudi prvi korak v funkcionalistični redukciji). V drugem koraku pa zapišemo diferencialne enačbe, ki opisujejo oboje, kognitivno in možgansko dinamiko, in so enake oblike, torej izomorfne (kar je prvi korak Nagelovske redukcije), kar pomeni, da lahko duha reduciramo na kvantni nivo (osnovne fizikalne delce) in iz njega sklepamo na možganske procese preko matematičnih algoritmov. Tako lahko pri algoritemski redukciji zaključimo, da so npr. bolečina, prepričanje ali ljubezen možganska stanja, ki so opisana z ustreznimi diferencialnimi enačbami ali matematičnimi algoritmi, kar je nov in svež pristop v sodobnem redukcionizmu.

Literatura

- Aberšek, B. (1993). *Analiza modela širjenja kratkih utrujenostnih razpok na zobeh zobnikov* [doktorska disertacija]. Maribor: FS.
- Aberšek, B., Flašker, J. (2004). *How Gears Break*. Southampton, Boston: WIT Press.
- Aberšek, B., Popov, V. (2004). Intelligent tutoring system for training in design and manufacturing. *Advances in engineering software*, 35, 461–471.
- Aberšek, B. (2012). *Didaktika tehnike med teorijo in prakso*. Ljubljana: Založba ZRSŠ.
- Aberšek, B. (2013). Cogito ergo sum homomachine? *Journal of Baltic science education*, vol. 12, no. 3, str. 268–270.
- Aberšek, B., Borstner, B., Bregant, J. (2014). *The virtual science teacher as a hybrid system*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Aberšek, B. (2014). *Filozofska analiza mehanizmov naravnega mišljenja in umetne inteligence* [doktorska disertacija]. Maribor: FF.
- Aberšek, B. (2018). *Problem-based learning and proprioception*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Aberšek, B., Flogie, A. & Pesek, I. (2023). *AI and cognitive modelling for education*. Springer.
- Abninder, L., Eliasmith, C., Kroon, F. W., Wiendstein, S., Thagard, P. (2006). »Is the Brain Quantum Computer?«, *Cognitive Science*, 30, 593–603.
- Abramson, D. (2008). »Turing's Responses to Two Objections«, *Mind & Machines*, 18: 147–167.
- Adleman, L. M. (1994). »Molecular computation of solutions to combinatorial problems«, *Science*, 266 (5187), 1021–1024.
- Akvinski, T. (1989). *Summa Theologiae* (ST). McDermott, A. T. (ur.), *A concise translation*, Texas: Christian Classics.
- Alexander, S. (1920). *Space, Time and Deity*. London: Macmillan and Co.
- Bedau, M. (1997). »Weak emergence«. V Tomberlin, J. (ur.), *Philosophical Perspectives: Mind, causation and world*, Vol. 11., Malden, MA: Blackwell.
- Agar N. (2012). "Why is it possible to enhance moral status and why doing so is wrong?," *Journal of Medical Ethics*, 2012, 39, pp. 67–74.
- Alligood, K. T., Sauer, T., Yorke, J. A. (1997). *Chaos: An Introduction to Dynamical Systems*. New York: Springer-Verlag.
- Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe*. Oxford: Oxford University Press.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., Qin, Y. (2004). »An Integrated Theory of the Mind«, *Psychological Review*, Vol. 111, No. 4, 1036–1060.
- Anderson, J. R. (1982). »Acquisition of cognitive skill«, *Psychological Review*, 98 (4), 369–406.
- Anderson, J. R., Bower, G. H. (1980). *Human Associative Memory: revised edition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1985). *Cognitive Psychology and its Implications*, 2nd Ed. New York: Freeman.
- Anderson, J. R., Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah: Erlbaum.
- Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.
- Anderson, J. R. (1996). »ACT: A Simple Theory of Complex Cognition«, *American Psychologist*, 51 (4), 355–365.
- Aristotel, (1993). *O duši*. Ljubljana: Slovenska matica.

- Aristotle, *Metaphysics*, Book H 1045a.
- Armstrong, D. M. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Ashby, R. (1952). *Design for a Brain*. London: Chapman-Hall.
- Atkinson, A. P. (1998). *Wholes and their Parts in Cognitive Psychology: Systems, Subsystems, and Persons*. Pridobljeno: 15. 9. 2012 s <http://cogprints.org/337/>
- Ausubel, D., Novak, J., Hanesian, H. (1978). *Educational Psychology: A Cognitive View*, 2nd Ed. New York: Holt, Rinehart & Winston.
- Baars, B. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford: Oxford University Press.
- Beer, R. (2000). »Dynamical Approaches to Cognitive Science«, *Trends in Cognitive Sciences*, 4(3), 91–99.
- Bechtel, W., Abrahamsen, A. (2002). *Connectionism and the Mind*. Oxford, UK: Blackwell Publisher.
- Bechtel, W., Mandik, J., Mundale, J., Stufflebeam, R. S. (2001). *Philosophy and the Neurosciences*. Oxford, UK: Blackwell Publisher.
- Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. London: Routledge
- Bekenstein, J. D. (1972). »Nonexistence of baryon number of static black holes«, *Physiological Review*, D 5, 2403–2412.
- Beeman, D. (2005). »Introduction to Realistic Neural Modeling«, *Brains, Minds, and Media*, 1, 1–22.
- Bell, C. G., Newell, A. (1971). *Computer Structures: Readings and Explanation*. New York: McGraw-Hill.
- Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., Shapiro, E. (2004). »An Autonomous molecular Computer for Logical Control of Gene Expression«, *Nature*, 429 (6990): 423–429.
- Bermudez, J. J. (2010). *Cognitive Science*. Cambridge: Cambridge University Press.
- Berners-Lee, T., Handler, J., & Lassila, O. (2001). The semantic web. *Scientific American*. May 2001. <https://www.scientificamerican.com/article/the-semantic-web/>
- Bickle, J. (1998). *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.
- Bickle, J. (2001). »Understanding Neural Complexity: A Role For Reduction«, *Minds and Machines*, 11, 467–481.
- Bickle, J. (2003). *Philosophy and Neuroscience: A Rutlessly Reductive Account*. Dorchester: Kluwer Academic Publisher.
- Bickle, J. (2006). »Reducing Mind to Molecular Pathways: Explicating the Reductionism Implicit in Current Cellular and Molecular Neuroscience«, *Synthese*, 151, 411–434.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals by a committee of college and university examiners (Handbook I: Cognitive domain)*. Longmans Publishing.
- Bohm, D. (1980). *Wholeness and the Implicate Order*. Amsterdam: John Benjamins.
- Boolos, G., Burgess, J., Jeffrey, R. (2002). *Computability and Logic*, 4th ed. Cambridge: Cambridge University Press.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.
- Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1). <https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/718>
- Bryson, J. J. (2018). »Patency is not a virtue: the design of intelligent systems and systems of ethics,« *Ethics and Information Technology*, 20, pp. 15–26.
- Bower, J. M., Beeman, D. (2007). »Constructing Realistic Neural Simulations with GENESIS«, *Methods in Molecular Biology*, Vol. 401, 103–125.
- Bower, J. M., Beeman, D. (1998). *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural Simulation System*, 2nd edition. New York: Springer-Verlag.
- Bregant, J. (2004). *Misel kot vzrok*. Maribor: Pedagoška fakulteta.
- Bregant, J. (2009). »Physicalism, or Something Near Enough: Good Enough to be a Global Theory?«, *Croatian journal of philosophy*, Vol. 9, No. 26, 219–232.
- Bregant, J. (2010). »Ali lahko stroj misli?«, *Analiza*, 14, št. 4, 55–72.
- Bregant, J., Stožer, A., Cerkenvenik, M. (2010). »Molecular Reduction: Reality or Fiction?«, *Synthese*, 172, 437–450.

- Broadbent, D. (1958). *Perception and Communication*. London: Pergamon Press.
- Broadman, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in Ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Barth.
- Broad, C. D. (1925). *Mind and its Place in Nature*. London: Kegan Paul.
- Brooks, F. P., Jr. (1962). »Architectural philosophy«. V Buchholz, W. (ur.), *Planning a Computer System*, New York: McGraw-Hill, 5–16.
- Brooks, R. (1991). »Intelligence without representation«, *Artificial Intelligence*, 47, 135–159.
- Brooks, R. (1999). *Cambrian Intelligence, The Early History of the New AI*. Cambridge, MA: MIT Press.
- Bruner, J. R., Goodnow, J. J., Austin, G. A. (1956). *A Study of Thinking*. New York: Wiley.
- Bullock, T. H. (1980). »Reassessment of Neural Connectivity and its Specification«. V Pinsker, H. M., Willis, W. D. (ur.), *Information Processing in the Nervous System*, New York: Raven Press, 119–220.
- Busemeyer, J. R., Townsend, J. T. (1993). »Decision Field Theory: A Dynamic Cognition Approach to Decision Making«, *Psychological Review*, 100, 432–459.
- Caine, R. N., & Caine, G. (1991). *Making connections: Teaching and the human brain*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Camus, A. (1999). *The Myth of Sisyphus and other Essays*. New York: Vintage Books, Random House.
- Carnap, R. (1928/1967). *The Logical Structure of the World and Pseudoproblems in Philosophy*. Berkeley, California: University of California Press.
- Carnap, R. (1934). *The Unity of Science*. London: Kegan Paul, Trench, Trubner, and Co.
- Carnap, R. (1963). »Autobiography«. Schilpp, P. A., (ur.), *The Philosophy of Rudolf Carnap*, Chicago, IL: Open Court, 1–84.
- Causey, R. L. (2001). *Logic, Sets, and Recursion*. Burlington, MA: Jones & Bartlett.
- Chader, G. J., Weiland, J., Humayun, M. S. (2009). »Artificial vision: needs, functioning, and testing of a retinal electronic prosthesis«, *Progress in Brain Research*, Vol. 175, 317–332.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York, NY: Oxford University Press.
- Chklovskii D.B. (2004). Synaptic connectivity and neuronal morphology: two sides of the same coin. *Neuron*, 43(5):609–17. doi: 10.1016/j.neuron.2004.08.012. PMID: 15339643.
- Chomsky, N. (1956). »Three Models for the Description of Language«, *IRE Transactions on Information Theory*, Vol. 2, 2, 113–123.
- Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.
- Churchland, P. (1981). »Eliminative Materialism and the Propositional Attitudes«, *The Journal of Philosophy*, 78, 67–90.
- Churchland, P. (2007). *Neurophilosophy at Work*. Cambridge: Cambridge University Press.
- Clark, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.
- Collins, A. M., Quillion, M. R. (1972). »How to make a language user«. V Tulvniq, E., Donaldson, W. (ur.), *Organization of Memory*, New York: Academic Press, 309–351.
- Copeland, J. (1993/2007). *Artificial Intelligence: A philosophical Introduction*. Oxford: Blackwell Publishing.
- Cosmides, L., Tooby, J., Barkow, J. (1992). »Evolutionary psychology and conceptual integration«. V J. Barkow, L. Cosmides, L., Tooby, J. (ur.), *The adapted mind: Evolutionary psychology and the generation of culture*, New York: Oxford University Press, 1–19.
- Cube, F. (1982). *Kybernetische Grundlagen des Lernen und Lehrens*. Stuttgart: Klett-Cotta.
- Damasio, A. R. (1996). »The somatic marker hypothesis and the possible functions of the prefrontal cortex«, *Philosophical Transactions: Biological Sciences*, 351(1346), 1413–1420.
- Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Press.
- Davidson, J. (2004). *The Secret of the Creative Vacuum: Man and the Energy Dance*, New York: C. W. Daniel.
- Dennett, D. C. (2005). *Sweet Dreams: Philosophical Obstacles to Science of Consciousness*. Cambridge: MIT Press.

- Descartes, R. (1985). *Strasti duše*. V *The Philosophical Writings of Descartes Vol. I* (CSM I); prevedli J. Cottingham, R. Stoothoff, D. Murdoch. Cambridge: Cambridge University Press.
- Descartes, R. (1984). Šesta zbirka odgovorov. V *The Philosophical Writings of Descartes Vol. II* (CSM II); prevedli J. Cottingham, R. Stoothoff, D. Murdoch. Cambridge: Cambridge University Press.
- Descartes, R. (1957). *Razprava o metodi*. Ljubljana: Slovenska matica.
- De Schutter, E., Bower, J. M. (1994). »An Active Membrane Model of the Cerebellar Purkinje Cell: I. Simulation of current clamps in slice«, *Journal of Neurophysiology*, Vol. 71, 1, 375–400.
- Dethier, J., Gilja, V., Nuyujukian, P., Elassaad, S.A., Shenoy, K. V., Boahen, K. (2011). »Spiking Neural Network Decoder for Brain-Machine Interfaces«, *IEEE EMBS Conference on Neural Engineering*, IEEE Press, 396–399.
- Deutsch, D. (2012). *The Beginning of Infinity: Explanations That Transform the World*. Harlow: Penguin Books.
- Dignum, V. (2018). Ethics in artificial intelligence: introduction to the special issue, *Ethics and Information Technology* 20:1–3.
- Downes, S. (2005). An introduction to connective knowledge. https://www.researchgate.net/publication/248290359_An_Introduction_to_Connective_Knowledge
- Dreyfus, H. (1979). *What Computers Can't Do*. New York: MIT Press.
- Duch, W., Oentaryo R. J., Pasquier, M. (2008). »Cognitive architectures: where do we go from here?«. V Wang, P., Goertzel, B., Franklin, S. (ur.), *Frontiers in Artificial Intelligence and Applications*, Vol. 171, Amsterdam: IOS Press, 122–136.
- Eccles, J. (1985). *Mind and Brain: The Many-Faceted Problems*. New York: Paragon House.
- Eliasmith, C. (1997). »Computation and dynamical models of mind«, *Minds and Machines*, 7, 531–541.
- Elman, J. L. (1998). »Connectionism, artificial life, and dynamical systems«. V Bechtel, W., Graham, G. (ur.), *A companion to cognitivescience*, Oxford: Basil Blackwood, 1–32.
- Erlhagen, W., Schoener, G. (2002). »Dynamic field theory of movement preparation«, *Psychological Review*, 109, 545–572.
- Ernst, G. W., Newell, A. (1969). *GPS: A case study in generality and problem-solving*. New York: Academic Press.
- Ethics Commission (2017). *Automated and Connected Driving*. Retrieved from: https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile.
- Ferreirós, J. (2001). »The Road to Modern Logic-An Interpretation«, *Bulletin of Symbolic Logic*, 7 (4), 441–484.
- FitzHugh R. (1961). »Impulses and physiological states in theoretical models of nerve membrane«, *Biophysical Journal*, 1, 445–466.
- Flogie, A., Barle Lakota, A., & Aberšek, B. (2018). The Psychosocial and Cognitive Influence of ICT on Competences of STEM Students. *Journal of Baltic Science Education*, 17, 267–276.
- Flogie, A., Aberšek, B. (2015). Transdisciplinary approach of science, technology, engineering and mathematics education. *Journal of Baltic science education*, 14, (6), 779–790.
- Fodor, J. A. (1975). *The Language of Thought*. Cambridge: Harvard University Press.
- Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- Frank, H. (1993). *Kybernetische Grundlagen der Pädagogik*. Prag: Kava-Pech.
- Frank, H., Mader, B. S. (1971). *Einführung in die kybernetische Pädagogik*. München: Wissenschaftliche Reihe.
- Freyman, R.P. (1982). »Simulating physics with computers«, *International Journal of Theoretical Physics*, 21, 467–488.
- Freud, S. (1925). *Collected papers*. London: Institute for Psychoanalysis and Hogarth Press.
- Fuller, O. S. (1990). *Factors affecting the balance of cooperation and conflict between the yucca moth, Tegeticula yuccasella, and its mutualist, Yucca glauca*. New Mexico: University of New Mexico.

- Gardner, H. (2011). *Truth, beauty, and goodness reframed: Educating for the virtues in the 21st century*. New York: Basic Books.
- Giunti, M. (1997). *Computation, dynamics and cognition*. Oxford: Oxford University Press.
- Gottfredson, L. S. (1997). »Foreword to 'Intelligence and social policy'«, *Intelligence*, 24 (1), 1–12.
- Gogus, A. (2012). Bloom's taxonomy of learning objectives. In N. M. Seel (Eds.), *Encyclopedia of the sciences of learning*. Springer. https://doi.org/10.1007/978-1-4419-1428-6_141
- Green, B. (2006). *Tkanina vesolja, Prostor, čas in tekstura resničnosti*. Tržič: Zbirka žepna knjiga.
- Gregory, R. L. (1961). »The brain as an engineering problem«. V Thorpe, W. H., Zangwill, O. L. (ur.), *Current Problems in Animal Behaviour*, Methuen, MA: C & MofP, 547–565.
- Grossberg, S. (1976). »Adaptive pattern classification and universal resampling: I. Parallel development and coding of neural feature detector«, *Biological Cybernetics*, 23, 121–134.
- Haken, H., Kelso, J. A. S., Bunz, H. (1985). »A theoretical model of phase transitions in human hand movements«, *Biological Cybernetics*, 51, 347–356.
- Haugeland, J. (1981). »Semantic Engines: An Introduction to Mind Design«. V Haugeland, J. (ur.), *Mind Design*, Bradford Books, MIT Press, 1–34.
- Hawking, S. (1988). *A Brief History of Time*. New York: Bantam Press.
- Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.
- Hempel, C. G., Oppenheim, P. (1948). »Studies in the logic of explanation«, *Philosophy of Science*, Vol. 15, 135–175.
- Henshall, W. (2024). When might AI outsmart us? It depends on who you ask. *Time*. <https://time.com/6556168/when-AI-outsmart-humans/>
- Hobbes, T., (1651/1994). *Leviathan*. Indianapolis: Hackett.
- Hofstadter, D. R. (1996). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books.
- Hodgkin, A., Huxley, A. (1952). »A quantitative description of membrane current and its application to conduction and excitation in nerve«, *Journal of Physiology*, 117, 500–544.
- Holland, J. H. (1975/1992). *Adaptation in Natural and Artificial Systems*. Cambridge, MA: MIT Press.
- Horgan, T. (1997). »Connectionism and the Philosophical Foundations of Cognitive Science«, *Metaphilosophy*, 28, 1–30.
- Horgan, T., Tienson, J. (1996). *Connectionism and the Philosophy of Psychology*. New York: Bradford Books, MIT Press.
- Horst, S. (2007). *Beyond Reduction: Philosophy of Mind and Post-Reductionist Philosophy of Science*. Oxford: Oxford University Press.
- Hume, D. (2000). *A Treatise of Human Nature*. Oxford: Oxford University Press.
- ITEA (2007). *Standard for technological literacy, Content for the study of technology*. Reston, Virginia: International Technology Education Association.
- ITEA (2003). *Advancing Excellence in technological literacy: Students assessment, professional development and program standard*. Reston, Virginia: International Technology Education Association.
- Jackson, F. (1982). »Epiphenomenal Qualia«, *Philosophical Quarterly*, 32, 127–136.
- Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Jaušovec, N. (1994). *Flexible thinking: An explanation for individual differences in ability*. Cresskill, NJ: Hampton Press Inc.
- Jeffress, L., A. (ur.) (1951). *Cerebral mechanisms in behavior, the Hixon symposium*. New York: Wiley.
- Jirsa, V. K., Haken, H. (1996). »Field Theory of Electromagnetic Brain Activity«, *Physical review letters*, Vol. 77, No. 5, 960–963.
- Jirsa, V. K. (2004). »Connectivity and dynamics of neural information processing«, *Neuroinformatics*, 2, 183–204.
- Jirsa, V. K., Kelso, J. A. S. (2005). »The excitator as a minimal model for the coordination dynamics of discrete and rhythmic movements«, *Journal of Motor Behavior*, 37, 35–51.
- Jost, J. (2005). *Dynamical systems*. Berlin: Springer-Verlag.

- Jrاده M., Bergounioux, M. (2008). *On Solutions of a Non linear Wave Equation Derived From Brain Activity Modeling*. Pridobljeno 20.07.2012 iz: <http://hal.archives-ouvertes.fr/docs/00/27/45/70/PDF/JMAA-BJrev.pdf>
- Jung, C. G. (1933). *Modern man in search of a soul*. London: Routledge.
- Kelso, J. A. S. (1995). *Dynamic patterns: The self-organisation of brain and behavior*. Cambridge, MA: MIT Press.
- Kelso, J. A. S. (2003). »Cognitive coordination dynamics«. V Tschacher, W., Dauwalder, J. P. (ur.), *The Dynamical Systems Approach to Cognition: Concepts and Empirical Paradigms Based on Self-Organization, Embodiment and Coordination Dynamics*, Singapore: World Scientific, 45–71.
- Kemeny, J., Oppenheim, P. (1956). »On Reduction«, *Philosophical Studies*, 7, 6–19.
- Kiguradze, T. (2001). »On bounded and time-periodic solutions of nonlinear wave Equations«, *Journal of Mathematical Analysis and Application*, 259 (1), 253–276.
- Kim, J. (1997). »Explanation, prediction and reduction in emergentism«, *Intellectica*, 25 (2), 45–57.
- Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: MIT Press.
- Kim, J. (1999). »Making sense of emergence«, *Philosophical Studies*, 95, 3–36.
- Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
- Kim, J. (2006). »Emergence: Core ideas and Issues«, *Synthese*, 151, 547–559.
- Kleene, S. C. (1952). *Introduction to Metamathematics*. Amstrdam: North-Holland Publishing Co.
- Kljajić, M. (2002). *Teorija sistemov*. Kranj: Moderna orgaizacija.
- Koestler, A. (1967). *The ghost in the machine*. New York: Macmillan.
- Kordež, U. (2004). *Od resnice do zaupanja*. Ljubljana: Studia Humanitatis.
- Kordež, U., Markič, O. (2007). *Kognitivna znanost v Ljubljani, Možnosti za študij in raziskovalno delo*. Ljubljana: Pedagoška fakulteta.
- Kordigel Aberšek, M. (2012). Neuroscience, world wide web and reading curriculum. *Problems of education in the 21st century*, vol. 46. 66–73.
- Kordigel Aberšek, M., Aberšek, B. and Flogie, A. (2017). Writing versus typing during science teaching: case study in Slovenia. *Journal of Baltic science education*. 17(1), 84-96.
- Kordigel Aberšek, M., Aberšek, B. (2021). *Society 5.0 and Literacy 4.0 for 21st Century*. NOVA Science Press.
- Kordigel Aberšek, M. (1991). *Recepcija slovenske znanstvenofantastične literature : doktorska disertacija*. Ljubljana
- Kordigel Aberšek, M., Dolinar, D. (urednik). (1994). *Znanstvena fantastika*. DZS.
- Kos, J. (1973). *Temelji filozofije*. Ljubljana: Cankarjeva založba.
- Kosslyn, S. H. (1996). *Image and brain: The resolution of the imagery debate*. Cambridge: MIT Press.
- Kutscher, C.F., Milford, J.B., Kreith, F. (2018). *Principles of Sustainable Energy Systems*. CRC Press
- Kuhn, T.S (1978). *Black-body theory and the quantum discontinuity 1894–1912*. Oxford: Clarendon Press.
- Kurzweil, R. (1998). *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York: Viking.
- Kurzweil, R. (2006). *The Singularity is Near*. New York: Viking Press.
- Laird, J. E., Derbinsky, N., Tinkerhess, M. (2011). *A Case Study in Integrating Probabilistic Decision Making and Learning in a Symbolic Cognitive Architecture*. Palo Alto, CA: Association for The Advancement Of Artificial Intelligence.
- Laird, J., Newell, A., Rosenbloom, P. (1987). »SOAR: An Architecture for General Intelligence«, *Artificial Intelligence*, 33, 1–64.
- Landa, L. N. (1974). *Algorithmization in Learning and Instruction*. New Jersey: Educational Technology.
- Lewes, G. H. (1875). *Problems of Life and Mind*, Vol. 2. London: Kegan Paul, Trench, Turbner, and Co.
- Lewin, D. I. (2002). »DNA computing«, *Computing in Science & Engineering*, 4 (3), 5–8.
- Lewis, M. D. (1980). »Mad Pain and Martian Pain«. V Block, N. (ur.), *Readings in the Philosophy of Psychology*, Vol. 1. Cambridge, Massachusetts: Harvard University Press, 216–222.
- Lurija, A. R. (1987). *The mind of a mnemonist: a little book about vast memory*. Cambridge, MA: Harvard University Press.
- Markič, O., Bergant, J. (2007). *Narava mentalnih pojavov*. Maribor: Aristej.

- Markič, O. (2008). Filozofija v kognitivni znanosti, *Družboslovne razprave*, 24, 59, 85–98.
- Markič, O. (2010). Kognitivna znanost: *Filozofska vprašanja*. Maribor: Aristej.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- McKraw, T. K. (2007). *Prophet of innovation: Joseph Schumpeter and creative destruction*. Harvard University Press.
- McCulloch, W. S., Pitts, W. H. (1943/1990). »A logical Calculus of the Ideas Immanent in Nervous Activities«, *Bulletin of Mathematical Biophysics*, 5, 115–133.
- McGinn, C. (1999). *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books, Perseus Books Group.
- Mead, C. A. (2002). *Collective Electrodynamics: Quantum Foundations of Electromagnetism*. Cambridge: MIT Press.
- Mehra, J., Rechenberg, H. (1982). *The historical development of quantum theory*. Stuttgart: Springer-Verlag.
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO 2023. ISBN 978-92-3-100612-8.
- Miao, F., & Tawil, S. (2024). Steering the digital transformation of education: UNESCO's human-centered approach. *Frontiers of Digital Education*, 1(1), 51–58. <https://doi.org/10.3868/s110-009-024-0005-6>
- Miller, G. (1956). »The magical number seven, plus or minus two: Some limits on our capacity for information processing«, *Psychological Review*, 63 (2), 81–97.
- Minsky, M. L., Papert, S. A. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Miščević, N., Markič, O. (1998). *Fizično in psihično: uvod v filozofijo psihologije*. Šentilj: Aristej.
- Morris, R., Filenz, M. (2003). *Science of the brain an introduction for young students*. Liverpool, UK: The British Neuroscience Association.
- Morris, R., Filenz, M. (2007). *Prvi koraki v nevroznanost, znanost o možganib*. Ljubljana: Sinapsa, Izobraževalni in raziskovalni inštitut Ozara.
- Mountcastle, V. B., (1975). »The view from within: Pathways to the study of perception«, *Johns Hopkins Medical Journal*, 136, 109–131.
- Myers, R., (1995). *Problem-Based Learning and Technology*. Wheeling: NASA Classroom of the Future, Wheeling Jesuit University.
- Nagel, E. (1961). *The structure of science. Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace World, Inc.
- Nagel, E. (1982). *The structure of science*. London: Routledge and Kegan Paul.
- NAEP (2018). *Technology and Engineering Literacy Assessment* Retrieved from: <https://www.nationsreportcard.gov/tel/>.
- NAGB (2018). *Technology and Engineering Literacy Framework for the 2018 National Assessment of Educational Progress*. Washington, DC: National Assessment Governing Board. Retrieved form: www.nagb.gov.
- Nagumo, J., Arimoto, S., Yoshizawa, S. (1962). »An active pulse transmission line simulating nerve axon«, *Proc IRE*, 50, 2061–2070.
- Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Lochlin, J. C., Perloff, R., Sternberg, R. J., Urbina, S. (1996). »Intelligence: Knowns and unknowns«, *American Psychologist*, Vol. 51 (2), 77–101.
- Neurath, O. (1931/1959). Sociology and Physicalism. V Ayer, A. J. (ur.), *Logical Positivism*, New York: The Free Press.
- Newell, A. (1980). »Physical symbol systems«, *Cognitive science*, 4, 135–183.
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., Simon, H. (1956). »The logic theory machine«, *IRE Transactions on information Theory*, 3, 61–79.
- Newell, A., Simon, H. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.
- Nunez, P. L. (1995). »Mind, brain, in electroencephalography«. V Nunez, P. L. (ur.), *Neocortical dynamics and human EEG rhythms*, New York: Oxford University Press, 133–194.
- Paivio, A (1986). *Mental representations: a dual coding approach*. Oxford: Oxford University Press.

- Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics*. Oxford: Oxford University Press.
- Peruš, M. (2000). *Biomreže, mišljenje in zavest*. Ljubljana: Založba DZS.
- Philips, W. A. (1997). »Theoris of Cortical Computation«. V Rugg, M. D. (ur.), *Cognitive Neuroscience*, Cambridge, MA: MIT Press, 11–46.
- Piaget, J. (1970). *Genetic epistemology*. New York: Columbia University Press.
- Piaget, J. (1985). *The equilibration of cognitive structures*. Chicago, IL: University of Chicago Press.
- Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: MIT Press
- PISA (2015). *Pisa 2015: draft reading literacy framework*. Paris: OECD Publications Service
- PISA (2018). *OECD Programme for International Student Assessment (PISA)*
<https://www.pisa.tum.de/en/domains/>
- Place, U.T. (1956). »Is Consciousness a Brain Process?«, *British Journal of Psychology*, 47, 44–50.
- Port, R. F. (2002). »The Dynamical Systems Hypothesis in Cognitive Science«, *Encyclopedia of Cognitive Science*, Vol. 1, 1027–1032.
- Port, R. F., Cummins, F., McAuley, J. D. (1995). »Naive time, temporal patterns and human audition«. V Port, R. F., van Gelder, T., (ur.), *Mind as Motion*. Cambridge, MA: MIT Press, 339–437
- Poper, K., Eccles, J. (1978). *The Self and Its Brain*. New York: Springer.
- Pratschke, B.M. (2024). *Generative UI and Education. Digital Pedagogies, Teaching Innovation and Learning Design*. Springer
- Putnam, H. (1975). *Mind Language and Reality: Philosophical Papers, Vol. 2*. Cambridge: Cambridge University Press.
- Pylyshyn, Z. W. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.
- Radford, A., & Narasimhan, K. (2018). *Improving language understanding by generative pre-training*. OpenUI.
- Ramsey, W., Stich, S. P., Garon, J. (1991). *Connectionism, Eliminativism and the Future of the Folk Psychology*. London: Erlbaum.
- Repovš, G. (2007). »Kognitivna nevroznanost: multidisciplinarni program raziskovanja odnosa med umom in možgani«. V Kordeš, U., Markič, O. (ur.), *Kognitivna znanost v Ljubljani, Možnosti za študij in raziskovalno delo*, Ljubljana: Pedagoška fakulteta, 116–124.
- Rolls, E., Treves, A. (1998). *Neural Networks and Brain Function*. Oxford: Oxford University Press.
- Rosenblatt, F. (1958). »The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory«, *Psychological Review*, Vol. 65, 6, 386–408.
- Rosenblatt, F. (1962). *Principles of Neurodynamics*. Washington, DC: Spartan Books.
- Rothschild, R. M (2010). »Neuroengineering tools/applications for bidirectional interfaces, brain–computer interfaces, and neuroprosthetic implants – a review of recent progress, *Frontiers in Neuroengineering*, Vol. 3, Article 112.
- Rumelhart, D. E., McClelland, J. L. (1986). *Parallel Distributed Processing – Part 1 and 2*. Chester, NJ: A Bradford Book, MIT Press.
- Schwartz, S. H. (1990). Individualism-collectivism: Critique and proposed refinements. *Journal of Cross-Cultural Psychology*, 21(2), 139–157. <https://doi.org/10.1177/0022022190212001>
- Searle, J. R. (1980). »Minds, brains, and programs«, *Behavioral and Brain Sciences*, 3 (3), 417–457.
- Shadmehr, R., Krakauer, J. W. (2008). »A computational neuroanatomy for motor control«, *Experimental Brain Research*, 185, 359–381.
- Shor, P. W. (1997). »Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer«, *SIAM, Journal on Computing*, 26 (5), 1484–1509.
- Sklar, L. (1967). »Types of Inter-theoretic Reduction«, *The British Journal for the Philosophy of Science*, 18, 109–124.
- Silberstein, M. (2002). »Reduction, emergence and explanation«. V Machamer, P., Silberstein, M. (ur.), *Blackwell guide to the philosophy of science*. Oxford: Blackwell.
- Simon, H. A. (1979). *Models of thought*. New Haven: Yale University Press.
- Smart, J. J. C. (1959). »Sensations and Brain Processes«, *The Journal of Philosophy*, 68, 141–156.

- Smith, L. B. (1999). »Knowing in the cortex of acting: the task dynamics of the A – nor – B error«, *Psychological Review*, 106, 235–260.
- Smith, L. B., Thelen, E. (2003). »Development as a dynamic system«, *TRENDS in Cognitive Sciences*, Vol.7, 8, 343–348.
- Smith, E. E., Shoben, E. J., Rips, L. J. (1974). »Structure and process in semantic memory: A featural model for semantic decisions«, *Psychological Review*, 1, 214–241.
- Smolensky, P. (1995). »Computational Models of Mind«. V Guttenplan, S. (ur.), *A Companion to the Philosophy of Mind*, Cambridge: Blackwell, 176–185.
- Sotomayor, J., Fernando Mello, L., De Carvalho Braga, D. (2007). »Bifurcation analysis of the Watt governor system«, *Computational & Applied Mathematics*, Vol. 26, 1, 19– 44.
- Spencer, J. P., Smith, L. B., Thelen, E. (2001). »Test of dynamic systems account of the A – nor – B error: the influence of prior experience on the spatial memory abilities of two-year-olds«, *Child Development*, 72, 1327–1346.
- Strmčnik, F. (1978). *Sodobna šola v luči programiranega pouka*. Ljubljana: DDU Univerzum.
- Strmčnik, F. (2001). *Didaktika: Osrednje teoretične teme*. Ljubljana: Znanstveni inštitut Filozofske fakultete.
- Suppes, P. (1966). The uses of computers in education. *Scientific American*, 215(3), 206–223.
- Thelen, E., Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Thelen, E., Schöner, G., Scheier, C., Smith, L. B. (2001). »The dynamics of embodiment: a field theory of infant perseverative reachings«, *Behavioural and Brain Sciences*, 24, 1–86.
- Tiku, N. (2022). The Google engineer who thinks the company's AI has come to life. *The Washington Post*. <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>
- Tooby, J., Cosmides, L. (1992). »The psychological foundations of culture«. V Barkow, J., Cosmides, L., Tooby, J. (ur.), *The adapted mind: Evolutionary psychology and the generation of culture*, New York: Oxford University Press.
- Tooby, J., Cosmides, L. (1990). »On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation«, *Journal of Personality*, 58, 17–67.
- Turing, A. M. (1936). »On Computable Numbers, with an Application to the Entscheidungsproblem«, *Proceedings of the London Mathematical Society*, 2, 42, 230–265.
- Turing, A. (1950). »Computing Machinery and Intelligence«, *Mind*, 59, 434–460.
- Turkle, S. (1995). *Life on the Screen: Identity in the Age of the Internet*. New York: Simon & Schuster.
- van Gelder, T. (1991). »Connectionism and dynamical explanation«, *Proceedings of the thirteenth annual conference of the cognitive science society*, Hillsdale, NJ: Erlbaum Associates, 499–503.
- van Gelder, T. (1999). »Revisiting the Dynamical Hypothesis«, Preprint No. 2/99, Melbourne: University of Melbourne. <http://www.arts.unimelb.edu.au/~tgelder/papers/Brazil.pdf>
- van Gelder, T., Port, R. (1995). »It's about time: an overview of the dynamical approach to cognition«. V Port, R., Van Gelder, T. (ur.), *Mind as motion*, Cambridge, MA: MIT Press, 1–43.
- van Gulick, R. (2004). »Consciousness«. *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Zalta, E. N. (ur.), URL = <http://plato.stanford.edu/archives/spr2014/entries/consciousness/>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need*. <https://arxiv.org/abs/1706.03762v1>
- von Bertalanffy, L. (1950). »An Outline of General System Theory«, *The British Journal for the Philosophy of Science*, Vol. 1, 2, 134–165.
- von Foerster, H. (1974). *Cybernetics of Cybernetics*. Urbana Illinois: University of Illinois.
- Voss, P. (2006). »Essentials of general intelligence: The Direct Path to Artificial General Intelligence«. V Goertzel, B., Pennachio, C. (ur.), *Artificial General Intelligence*, Berlin: Springer, 131–159.
- Walmsley, J. (2008). »Explanation in dynamical cognitive sciences«, *Minds and Machines*, 18, 331–348.
- Walmsley, J. (2010). »Emergence and reduction in dynamical cognitive sciences«, *New ideas in Psychology*, 28, 247–282.

- Weidman, N. M. (2006). *Constructing Scientific Psychology: Karl Lashley's Mind-Brain Debates*. Cambridge: Cambridge University Press.
- Wellman, H. M., Cross, D., Bartsch, K. (1986). »Infant search and object permanences: a meta-analysis of the A – nor – B error«, *Monographs of the Society for Research in Child Development*, 51 (3), 1–51, 62–67.
- Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., Danny Axsom, D. (2000). »Focalism A Source of Durability Bias in Affective Forecasting«, *Journal of Personality and Social Psychology*, Vol. 78, 5, 821–836.
- Zaghloul, K. A., Boahen, K., Demb, J. B. (2003). »Different circuits for ON and OFF retinal ganglion cells cause different contrast sensitivities«, *Journal of Neuroscience*, 23, 2645–2654.
- Zaghloul, K. A., Boahen, K., Demb, J. B. (2005). »Contrast adaptation in subthreshold and spiking responses of mammalian Y-type retinal ganglion cells«, *Journal of Neuroscience*, 25, 860– 868.
- Zaghloul, K. A., Manookin, M., Borghuis, B. G., Boahen, K., Demb, J. B. (2007). »Functional circuitry for peripheral suppression in mammalian Y-type retinal ganglion cells«, *Journal of Neurophysiology*, 97, 4327–4340.
- Zhijian, Y. (2003). »Initial boundary value problem for a class of non-linear strongly damped wave equations«, *Mathematical Methods in the Applied Sciences*, 26 (12), 1047–1066.
- Zuse, K. (1993). *Der Computer. Mein Lebenswerk*, 3rd ed. Berlin: Springer-Verlag.

Stvarno kazalo

A

ACT-R, 6, 12, 54, 111, 115, 147, 200, 231,
252, 257, 259, 260, 261, 262, 263, 264,
265, 266, 267, 269, 270, 271, 272, 273,
274, 275, 277, 278, 279, 281, 282, 283,
284, 285, 332, 373, 376, 377
agenti, 48, 96, 112, 151, 252, 253, 372
akcijski potencial, 6, 164, 166, 167, 168, 169,
180, 188, 310, 373
akcijski potencial, 164
Aksoni, 166
algoritemske redukcije, 5, 378
algoritmi, 45, 48, 49, 99, 108, 121, 129, 136,
185, 203, 271, 299, 367, 369, 380
analitične gramatike, 110
Analogija, 30, 248, 278
animalični blapovi, 59, 60
atraktor, 95, 319

B

Bioračunalniki, 334
Boolove algebre, 205, 337

Č

človeška inteligenca, 18

D

dedukcijo, 72, 74, 93
Deep Blue, 47
delta pravilo, 46, 191, 214, 215, 216
Dendriti, 165
Digitalna pedagogika, 37
dinamični sistemi, 92, 96, 101, 116, 141, 196,
197, 287, 300, 304, 305
dolgoročni spomin, 66, 125, 126, 178, 200
Dualizem, 57, 77, 357

dualizem lastnosti, 57

duševne reprezentacije, 28, 29, 30

E

ekspertni sistem, 25, 46
Eliminativizem, 31
Eliza, 228, 229, 230, 241, 242, 243
emergenca, 69, 71, 72, 74, 75, 76, 95
empirična raziskovanja, 27, 31
Epistemološki del, 24
Etika in tehnologija, 150
Evklidov algoritem, 134

F

filozofije duha, 3, 18, 24, 25, 57, 68, 115, 362,
371, 372
filozofije jezika, 24
filozofije znanosti, 24, 62
filozofska dilema, 54
fonološka sbramba, 178
Formalna gramatika, 110
funkcija, 87, 139, 145, 158, 186, 191, 205,
214, 215, 249, 269, 280, 291, 317, 322,
325, 326, 327, 328, 344, 351, 353, 369, 378
funkcionalistična redukcija, 72, 79, 365
Funkcionalizem, 28, 357, 365

G

Generativizem, 38
generativna umetna inteligenca (GUI), 34
generativne gramatike, 110
Globoko učenje, 49
Gradnik, 271

H

hevristični del, 24, 33, 252
 hibridni sistemi, 6, 221, 283, 375
 HKB model, 90, 319, 321

I

Inteligenca, 102
Intelligentni agent, 112, 141
 Intencionalni realizem, 245
introspekcija, 124

J

jezik misli, 23, 244, 248, 249, 250

K

Kibernetična didaktika, 52, 132, 137, 140, 372
 kibernetike, 51, 84, 120, 132, 182, 220
 Kibernetški pristop, 17
 kognicija, 23, 82, 91, 98, 99, 123, 142, 219,
 259, 266, 279, 281, 306, 331
 Kognitivna arhitektura, 141, 144, 145, 255
 kognitivna filozofija, 23
 kognitivnega nevromodeliranja, 4, 19
 kognitivno teorijo učenja, 5, 160, 374
 kognitivno znanost, 5, 76, 82, 121, 122, 143,
 160, 170, 199, 210, 217, 219, 220, 300,
 306, 307, 331, 344, 367, 374
koharentna superpozicija, 337
 konekcionistične mreže, 189, 198, 210, 213,
 216
Konekcionistični model, 99
konekcionističnih modelov, 83, 195, 199, 200
Konekcionizem, 52, 132, 190, 209, 211
kratkoročni spomin, 125, 126, 127, 177
 krovnega zakona, 74, 86, 93, 94, 95
Kvantna mreža, 201
 kvantni nivo, 7, 343, 375, 380
 Kvantni računalnik, 337

L

lingvistika, 23, 51, 120, 124, 143

M

mentalni fenomeni, 26
 mentalnih procesov, 5, 315
 metafizično vprašanje, 27
Misterianistični, 3
Mnogoslojna fraktalna mreža, 201
močna UI, 50, 333
 model procesiranja informacij, 124, 125
Modularna nevronska mreža, 201
monizem, 57
Moorov zakon, 47
morfodinamika, 97

N

Nagelovsko redukcijo, 7, 72, 76, 95, 361, 368,
 380
naturalističen pristop, 4, 5, 7, 375
Naturalistični, 4
nereduktivni fizikalizem, 27, 30, 64
 nevrološki dinamični model, 89
nevromorfno inženirstvo, 173, 208
 Nevron, 163
 nevroni, 6, 133, 146, 164, 165, 167, 173, 175,
 176, 180, 183, 184, 188, 189, 190, 201,
 202, 203, 207, 208, 213, 312, 336, 342, 373
 nevronske mreže, 33, 46, 49, 82, 99, 111, 132,
 188, 189, 190, 191, 200, 201, 202, 203,
 206, 208, 209, 210, 214, 221, 259, 262,
 263, 287, 332, 334, 375
 nevroračunalništva, 4, 19, 371
 nevrottransmitor, 168, 169, 176, 177
Nevrottransmitorji, 167, 168
 nevroznanost, 5, 15, 23, 26, 27, 51, 67, 85,
 119, 120, 122, 143, 155, 156, 157, 160,
 169, 171, 172, 180, 192, 210, 216, 217,
 218, 221, 263, 333, 364, 373, 374, 375,
 387, 388
 Newel, 387
 Nomološko pojavnost, 74
Normativno, 26

P

paralelno procesiranje (PDP), 123
perceptron, 46, 183, 184, 193, 194, 200
plastičnost, 175, 177, 213
Podsimbolni nivo, 270
Problem intencionalnosti, 64
Problem zavesti, 65

procesni modul, 267, 273, 280
programiranim poukom, 138
Programski jezik, 109
propozicijska logika, 248
 Prostor in čas, 347
prostor stanja, 97, 289, 290, 291, 292, 344, 377
prvotna duša, 61
Psihofizični antiredukcionizem, 27

Q

qualia, 31, 69, 200

R

Računalniška arhitektura, 144
 računalniška metafora, 142, 217
 receptorji, 168, 176
 redukcionizem, 5, 27, 67, 69, 72, 73, 78, 79, 80, 81, 363
reduktivni fizikalizem, 27
redundantna teorija, 134
Rekurzivne funkcije, 110
 roboti, 95, 108, 113, 114, 206

S

semantična (vsebinska) raven, 143
 semantično vrednost, 29
senzorični spomin, 125, 126
Simbolni nivo, 270
 simbolno modeliranje, 182, 184
 Simon, 387
sinergija, 97
 sintaktična raven, 29, 143
Situacijska kognicija, 99
 Splošna umetna inteligenca, 50
 stališča semantike, 248
 stališča sintakse, 248
 STATFAG, 25
 STEAM, 148, 149, 150
 strojno inteligenco, 18
 strojno učenje, 45, 108, 111, 114, 129, 130, 160

struktura, 83, 132, 140, 144, 145, 146, 165, 226, 231, 244, 249, 256, 262, 266, 270, 278, 374
 strukturalne arhitekture, 29
 supervenienca, 71, 72, 73
 Supervenienca, 71
 supervenira, 63, 71

Š

šibko UI, 50

T

Taksonomija učenja, 131
 takšnosti, 31, 69, 200
Tehnološka in inženirska pismenost, 149, 152
Teorija algoritmov, 134
teorija informacij, 133
teorija komunikacij, 133, 134
 teorijo dinamičnih sistemov
 teorija dinamičnih sistemov, 6, 52, 82, 91, 97, 288, 315, 344, 377
 Trajektorija, 289, 291, 292

U

Učenje, 102, 128, 129, 131, 133, 157, 158, 160, 161, 184, 251, 374
 ustvarjalnost, 105, 106, 107, 148
 Utelešena kognicija, 97

V

Velike podatkovne baze, 50
vidno-prostorska skicirka, 178
Virtualna mreža, 201
vzvratno inženirstvo, 171

Z

Zavest, 7, 26, 369, 374

FILOZOFSKA ANALIZA MEHANIZME KOGNITIVNEGA MODELIRANJA IN UMETNE INTELIGENCE V IZOBRAŽEVANJU

BORIS ABERŠEK,¹ METKA KORDIGEL ABERŠEK²

¹ Univerza v Mariboru, Fakulteta za naravoslovje in matematiko, Maribor, Slovenija
boris.abersek@um.si

² Univerza v Mariboru, Pedagoška fakulteta, Maribor, Slovenija
metka.kordigel@um.guest.si

Tehnologija je dvoje, je vzrok in posledica za vse hitreje spreminjajočo se družbo. Ko danes govorimo o tehnologiji, govorimo predvsem o inteligentnih sistemih, o tako imenovanih kibernetičnih sistemih, pri katerih sistem vsebuje, če govorimo v jeziku tehnologije, *strojno opremo*, stroje in naprave (človeško telo), in *programsko opremo* – inteligenco – duha, kar bi v jeziku humanistike lahko poimenovali tudi *občutke*, *intencionalna stanja*, *čustva* ipd. Mnogi problemi, ki so povezani s sodobnimi tehnologijami, so zato povezani tudi z (umetno) inteligenco in učenjem. Za reševanje problemov v zvezi z umetno inteligenco potrebujemo predvsem splošna znanja o svetu, družbi in ljudeh, kar je *filozofski problem*. Zato je za vse, ki morajo sprejemati odločitve s področja uporabe umetne inteligence, izjemno pomembno, da *razumejo* osnovne splošne mehanizme inteligence – filozofijo duha, tj. kako deluje naravna, kako deluje človeška inteligenca, kako se človek uči in prilagaja okolju, kar je še posebej pereče na področju vzgoje in izobraževanja.

DOI

[https://doi.org/
10.18690/um.fnm.x.2025](https://doi.org/10.18690/um.fnm.x.2025)

ISBN

978-961-299-076-3

Ključne besede:

filozofija uma,
kognitivna znanost,
nevroznanost,
kognitivno modeliranje,
umetna inteligenca,
dinamični sistemi,
izobraževanje



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.fnm.x.2025](https://doi.org/10.18690/um.fnm.x.2025)

ISBN
978-961-299-076-3

Keywords:

philosophy of mind,
cognitive science,
neuroscience,
cognitive modelling,
artificial intelligence,
dynamical system,
education

PHILOSOPHICAL ANALYSIS OF THE MECHANISMS OF COGNITIVE MODELING AND ARTIFICIAL INTELLIGENCE IN EDUCATION

BORIS ABERŠEK,¹ METKA KORDIGEL ABERŠEK²

¹ University of Maribor, Faculty of Natural Sciences and Mathematics, Maribor, Slovenia
boris.abersek@um.si

² University of Maribor, Faculty of Education, Maribor, Slovenia
metka.kordigel@um.guest.si

Technology is both a cause and a consequence for an ever faster changing society. Speaking about technology today, we speak above all about intelligent systems, so called cybernetic systems, the system which contains, if we speak in technological language, *engine equipment*, machines and devices (the human body), and *program equipment* – intelligence – mind (in the language of humans), also *feelings*, *intention conditions*, *emotions*, etc. Many problems connected with modern technologies are therefore also connected with (artificial) intelligence (AI). To solve problems connected with artificial intelligence, above all we need common knowledge about the world, society, and human beings, which is a *philosophical problem*. Therefore, it is very important for everyone who has to take decisions in the field of artificial intelligence *to understand* the basics of general mechanisms of intelligence – philosophy of the mind, to understand the way natural and artificial intelligence functions especially in the area of education.



University of Maribor Press



Univerza v Mariboru

Fakulteta za naravoslovje
in matematiko

