Olga Markič

Kognitivizem in konekcionizem – dva pristopa v kognitivni znanosti

UVOD

Kognitivna znanost ali "Nova znanost duha", kot jo imenuje Howard Gardner, je interdisciplinarno področje, ki ga sestavljajo filozofija, psihologija, računalništvo, lingvistika, nevroznanost in antropologija. Z združenimi močmi skušajo s sodobnimi sredstvi in metodami odgovoriti na stara epistemološka vprašanja, ki so si jih zastavljali že starogrški filozofi, na primer: Kaj pomeni to, da nekaj vem? Kje so izvori védenja? Kakšni so mehanizmi spominjanja, učenja, zaznavanja? Kakšna je vloga jezika?... Pri tem pomembno vlogo igra računalnik. Na eni strani govorimo o računalniški metafori, o računalniku kot modelu za psihološko teorijo, na drugi strani pa predstavlja močno orodje, s katerim lahko teorijo v praksi preizkušamo. Prav to, kako gledamo na računalnik – ali igra obe vlogi ali pa je le primeren pripomoček za simulacije, loči dva pristopa v kognitivni znanosti, o katerih bo govor v nadaljevanju.

NASTANEK KOGNITIVNE ZNANOSTI

Začetki kognitivne znanosti in njenega jedra, kognitivizma, segajo v petdeseta leta in pomenijo reakcijo na takrat predvsem v anglosaksonskem svetu prevladujočo behavioristično psiholo-

gijo. Pomemben mejnik je bil Hixon Symposium na California Institute of Technology leta 1948, 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 podal izziv doktrini behaviorizma.

Behavioristični psihologi so se v želji, da bi bila njihova psihološka raziskovanja res znanstvena, omejili le na proučevanje obnašanja. Zavračali so introspekcijo, ki ne more nuditi javne dostopnosti predmeta raziskovanja in zato ni primerna kot znanstvena metoda eksperimentalne znanosti. Ameriška psihologa J. B. Watson in B. F. Skinner sta zastopala stališče, da je edini ustrezni predmet preučevanja behavioristične psihologije odnos med dražljaji iz okolice organizma in reakcijami organizma na te dražljaje. Odločilen faktor pri razlagi je predstavljalo okolje. Posamezniki se ne obnašajo tako, kot se, zaradi lastnih idej in namer, ampak le pasivno odslikujejo sile iz okolja. Behavioristi se se veliko ukvarjali s principi pogojevanja in ojačanja, saj so menili, da bodo s pomočjo teh principov lahko razložili, kako se posameznik nauči določenega obnašanja. Za psihološka stanja, katerih rezultat ni opazljivo obnašanje organizma, so v razlago vključili še dispozicijska stanja. Ta naj bi predstavljala urojena ali pogojena nagnjenja k refleksnemu reagiranju organizma na določene dražljaje. Določiti naj bi jih bilo mogoče s testi in s tem zvesti na potencialno opazljivo obnašanje. Behavioristi bi se na ta način izognili po njihovem mnenju sumljivim mentalističnim entitetam.

Behavioristično metodologijo si najlaže 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 obnašanjem, ne spušča pa se v raziskovanje same "črne škatle". Vse, kar se dogaja znotraj nje, se ne dogaja na ravni, ki zanima znanstveno psihologijo. Na ta način se behavioristični psiholog pri razlagi izogne tako sklicevanju na nevrofiziološka kot tudi na mentalna stanja. Pokazalo pa se je, da je projekt, po katerem bi lahko vse pojme z mentalistično vsebino definirali s pomočjo operacijskih definicij (kot dejanske ali potencialne reakcije organizma na določene dražljaje), neuresničljiv. Psihologi so se obrnili k vsakdanji zdravorazumski psihologiji, ki uporablja prepričanja, želje, strahove in namere za razlago obnašanja in je nudila okvir, ki ga je bilo treba napolniti z znanstveno teorijo. S sprejetjem mentalističnih pojmov, ki igrajo določeno vzročno vlogo v razlagi obnašanja, se je odprl nov prostor raziskovanja, kognitivizem.

Kognitivizem je od behaviorizma sprejel pogled, po katerem je cilj psihologije razlaga in napoved obnašanja, kot tudi to, da je za to ustrezna raven funkcionalna raven, ne pa nevrofiziologija ali kakšna druga naravoslovna znanost. Loči pa se od njega v tem, da je odprl "črno škatlo" in vanjo postavil mentalna stanja kot člene v vzročni verigi razlage in napovedi obnašanja. Pri tem se je zgledoval predvsem pri delovanju računalnika. Analogija med miselnimi procesi in računalnikom je odprla možnosti za nove hipoteze in za njihovo preverjanje. Vprašanja o naravi jezika, načrtovanju, reševanju problemov, domišljiji ..., ki so se v času behaviorizma le redko pojavljala, so postala osrednje teme raziskovanja. Za "rojstno leto" kognitivne znanosti večina priznava leto 1956, ko sta na področju psihologije izšli deli "Study of Thinking" (J. Bruner, J. Goodnow in G. Austin) ter "The magical number seven" (G. Miller), na področju lingvistike "Three models of language" Noama Chomskega in na področju računalništva članek "Logical Theory Machine", v katerem sta Allen Newell in Herbert Simon opisala prvi popolni dokaz izreka, ki je bil izveden na računalniku.

Kognitivna revolucija, nov pristop v proučevanju duha, pa ne bi bila možna brez nekaterih teoretskih prispevkov. Ključnega pomena so bili dosežki matematike in logike, razvoj informacijske teorije, kibernetika, modeliranje nevronov in raziskovanje spoznavnih nezmožnosti, ki so posledice poškodb možganov.

Prav gotovo je imelo izredno velik vpliv delo britanskega matematika Alana Turinga. V članku o izračunljivih številih, ki ga je napisal leta 1936, je definiral komputacijo kot formalno manipulacijo z neinterpretiranimi simboli, ki se izvaja z uporabo formalnih pravil. Za ponazoritev je razvil pojem preprostega stroja, ki ga danes imenujemo Turingov stroj. Ta "teoretičen" stroj, v bistvu predmet abstraktne matematike, je sestavljen iz dveh delov, neskončno dolgega traku, ki se pomika skozi stroj, in glave, ki bere, kar je na traku zapisano. Trak je razdeljen na polja, ki so ali prazna ali pa je v njih zapisan znak. Stroj lahko izvede štiri operacije: trak lahko premakne v levo, desno, zbriše zapis na traku ali pa doda novega. Turing je pokazal, da lahko s takim preprostim strojem izvedemo vsako nalogo, za katero lahko jasno navedemo korake, ki so potrebni za izpolnitev naloge. Turingov stroj lahko programiramo tudi tako, da bo oponašal katerikoli drug Turingov stroj. Takemu oponaševalcu pravimo univerzalni Turingov stroj.

Turingove ideje so porodile misel o stroju, ki bi samo s svojo formalno strukturo posnemal delovanje duha. Sposobnost izvajanja katerekoli dovolj dobro določene spoznavne naloge je v principu določena s formalnimi lastnostmi, materialna realizacija ni pomembna. Tako teoretsko izhodišče je temelj klasični kognitivni znanosti. Toda Turingov stroj je vendarle le

abstrakten, matematičen konstrukt in tako narejen stroj bi bil veliko prepočasen za kakršnekoli zanimive naloge. Dejstvo, da je mogoče narediti stroj, ki samo s pomočjo zapisa v dvojiškem sistemu (0, 1) izvaja različne programe, pa je spodbudilo znanstvenike k izdelavi elektronskih digitalnih računalnikov.

Prav razvoj elektronskih digitalnih računalnikov, univerzalnih Turingovih strojev, ki so veliko hitrejši in jih lažje programiramo, je omogočil izdelavo kognitivnih modelov in njihovo eksperimentalno preverjanje. Turinga samega je zelo prevzela ideja, da bi bilo mogoče tako programirati računalnik, da bi zanj lahko rekli, da misli. V zelo odmevnem članku "Stroji, ki računajo, in inteligenca" (1950) je predlagal preizkus, igro posnemanja. Računalnik bi odgovarjal (pisno) na zastavljena vprašanja. Če bi tako dobro oponašal človeško obnašanje, da spraševalec ne bi mogel ugotoviti, ali je vprašani človek ali računalnik, potem bi računalnik opravil "Turingov test". To bi bilo za Turinga dovolj, da bi računalniku lahko pripisali inteligentnost. Tako Turingovo stališče ni bilo soglasno sprejeto in znanstveniki in filozofi so ga napadli na različne načine. Omenimo naj le ugovor kognitivistov, da je test preveč behavioristično in da samo posnemanje inteligentnega obnašanja še ni zadosten kriterij za inteligenco. Ta ugovor je uperjen proti testu kot takemu, vendar ne zanika možnosti, da bi računalniki lahko bili inteligentni, čemur nasprotujejo drugi kritiki (npr. Dreyfus in Searle).

Za razvoj kognitivnega modeliranja je bilo zelo pomembno tudi delo Mc Cullocha in Pittsa, ki sta leta 1943 pokazala, da lahko operacije živčnih celic, nevronov, in njihove povezave z drugimi nevroni modeliramo s pojmi logike. Nevrone si lahko zamislimo kot stavke, stanja posameznih nevronov, ki so lahko ali vzburjeni ali ne (lastnost vse ali nič), pa primerjamo z resničnostno vrednostjo stavka. V takem modelu si lahko zamislimo, da je aktivni nevron, ki prižge drug nevron, kot stavek, iz katerega v neki logični sekvenci sledi drug stavek ($p \rightarrow q$). McCulloch in Pitts sta pokazala, da je možno vse, kar lahko izčrpno in nedvoumno opišemo, realizirati z ustrezno končno nevronsko mrežo. Pojem Turingovega stroja tako gleda v dve smeri – k nevronski mreži, ki je sestavljena iz velikega števila med seboj povezanih nevronov, in k računalniku, ki lahko realizira vse tiste procese, ki jih lahko nedvoumno opišemo. Ta dva pogleda nista bila vedno enako močno zastopana in dejansko kažeta na dva pristopa, o katerih bo govor v nadaljevanju.

KOGNITIVIZEM

Dosežki, ki smo jih omenili, so sicer odprli vrata novemu pogledu, vendar sami po sebi še ne nudijo teorije, ki bi razla-

gala naravo duha. Filozofski poskus, da bi razložili kritičen del, tj. način, na katerega prepoznavamo in razvrščamo mentalna stanja, predstavlja funkcionalistična teorija duha.

Funkcionalizem je filozofska teorija, po kateri duševna (mentalna) stanja razvrščamo glede na njihovo funkcionalno vlogo, ki dejansko predstavlja njihovo vzročno vlogo. Vsako mentalno stanje in proces ima svojo vzročno vlogo znotraj sklopa vzročnih relacij. Te vključujejo vzročne dražljaje iz okolice organizma, medsebojne vzročne relacije z drugimi mentalnimi stanji in procesi v organizmu, in vzročne učinke, ki se kažejo v obnašanju.

Pomembni trditvi, ki sledita iz funkcionalističnega pristopa, sta:

1. mentalna stanja imajo vzročno vlogo,
2. določanje mentalnih stanj in procesov ni odvisno od materialne realizacije (nevrofizioloških stanj in procesov).

S prvo trditvijo se funkcionalisti odmikajo od behavioristič-nega stališča, po katerem se razlaga zvede izključno na dražljaje iz neposredne okolice (vzročni vhodni podatki) in na obnašanje (vzročni izhodni podatki). Omogoča jim razlago, v kateri se sklicujejo na mentalna stanja, kot so prepričanja, želje, namere itd. S tem zadržijo zdravorazumske pojme in jih skušajo umestiti v nov znanstven okvir. Druga trditev pa pomeni, da ne sprejemajo teorije identitete, po kateri tipsko izenačujemo mentalna in nevrofiziološka stanja. Nevrofiziološka raven ni ustrezna raven za določanje funkcionalne organizacije organizma, saj tipe mentalnih stanj razvrščamo na osnovi njihovih vzročnih vlog in ne na osnovi njihovega fizičnega primerjanja. Za primer funkcionalisti velikokrat omenjajo bolečino. Da bi za neko mentalno stanje lahko rekli, da je bolečina, je dovolj, da določimo njegovo vzročno vlogo, ki je izogibanje ali odstranitev izvora poškodbe organizma. Pri tem bolečino pojmujemo kot bolečino zato, ker ima tako funkcionalno vlogo, ne pa zaradi nepoznanega nevralnega stanja. Za funkcionaliste so posamezna mentalna stanja, ki imajo isto vzročno vlogo, le primerki določenega tipa mentalnega stanja. Isti tip je tako lahko implementiran v različnih nevralnih stanjih.

Teoretske postavke funkcionalizma so vodile k novemu metodološkemu pristopu v psihologiji, h kognitivizmu, ki je nekakšno jedro kognitivne znanosti in je prevladoval do vzpona konekcionizma konec osemdesetih let.

Kognitivistični pristop temelji na tehle predpostavkah:

1. Za razlago obnašanja organizma je potreben notranji kognitivni sistem, ki posreduje med vhodnimi podatki iz okolice in izhodnimi podatki iz organizma (tj. obnašanjem).

2. Funkcija kognitivnega sistema je obdelovanje informacij kot simbolne komputacije nad mentalnimi reprezentacijami.

Osnovni pojem je pojem mentalne reprezentacije. Kognitivni znanstvenik namreč meni, da vsako inteligentno obnašanje predpostavlja zmožnost ustrezne predstavitve sveta in da razlaga obnašanja brez te predpostavke ni mogoča. Po analogiji z računalnikom naredi kognitivist še naslednji korak. Postavi hipotezo, da so te reprezentacije fizično realizirane v obliki simbolne kode v možganih, tako kot so v stroju. S simbolno komputacijo rešuje problem, kako pokazati, da so mentalna stanja (prepričanje, želja, namera...) ne samo fizično mogoča, ampak tudi dejansko povzročajo obnašanje. Simbol ima namreč tako fizične kot semantične (pomenske) vrednosti in komputacije so operacije na simbolih, ki so omejene s temi semantičnimi vrednostmi. Vemo, da digitalni računalnik deluje le glede na fizično obliko simbolov in da nima dostopa do njihovih semantičnih vrednosti. Kljub temu so operacije omejene s semantiko, kajti vsako pomensko razliko, ki je pomembna za program, programer zakodira v sintaksi simbolnega jezika. Pravimo, da v računalniku sintaksa zrcali semantiko. Kognitivizem trdi, da nam taka vzporednost kaže, kako je inteligenca fizično mogoča, in postavlja hipotezo o računalniku kot mehanskem modelu misli. Najbolj znana zagovornika hipoteze o fizičnem simbolnem sistemu sta Simon in Newell, ki sta te teoretske postavke in rezultate, ki sta jih dobila s psihološkimi eksperimenti, uporabila v različnih računalniških modelih mišljenja (npr. program "Human Problem Solver", 1972).

Za pravilno razumevanje analogije med računalnikom in duhom je pomembno, da ugotovimo, na kateri ravni je hipoteza postavljena. Kognitivisti govore o treh ravneh opisa duha/možganov: semantični, sintaktični in ravni strukturalne arhitekture. Za razlago delovanja duha sta pomembni semantična in sintaktična raven, saj duh predstavlja množico mentalnih reprezentacij, nad katerimi potekajo sintaktično določene reprezentacije. Raven strukturalne arhitekture, raven možganov, po mnenju kognitivistov ni pomembna za psihološko razlago in je domena nevroznanosti. Tudi če bi lahko odprli glavo in pogledali v možgane, v njih ne bi našli majhnih simbolov. Kajti čeprav je sintaktična raven fizično realizirana, je ne moremo reducirati na fizično raven. Namesto o redukciji kognitivisti govore o implementaciji, podobno kot je računalniški program implementiran v strojni opremi. Kognitivizem tako poleg ravni fizike in nevrobiologije postavlja še dve posebni ravni – sintaktično in semantično ali reprezentacijsko, ki ju ne moremo reducirati. Iskanje ustreznih psiholoških razlag je podobno iskanju ustreznih programov, ki jih duh/možgani izvaja. Psihologijo torej zanima raven funkcionalne in ne strukturalne arhitekture.

Analogija med računalnikom in duhom je koristna tudi zato, ker omogoča empirično preverjanje resničnosti psiholoških teorij. Temeljni kriterij za sprejetje ali zavrnitev hipotez je možnost njene računalniške realizacije, izdelave modela. Prav zato je umetna inteligenca, veja računalništva, ki se ukvarja z vprašanjem, kako konstruirati (programirati) stroj, da bo kazal inteligentno obnašanje, ena osrednjih disciplin kognitivne znanosti. Raziskovanja so prinesla zanimive rezultate, tako na tehničnem področju, kjer ni bil cilj modeliranje spoznavnih funkcij (npr. ekspertni sistemi, roboti, program za igranje šaha), kot tudi pri izdelavi kognitivnih modelov (npr. reševanje problemov, vizualno procesiranje).

Po začetnih obetajočih rezultatih in optimističnih napovedih pa je umetna inteligenca zašla v težave. Programi, zasnovani kot eksplicitna množica navodil za manipuliranje s simboli, so se izkazali za preveč toge. Klasični simbolni modeli, ki so se dokaj dobro izkazali pri modeliranju višjih miselnih procesov, so le slabo pokrivali take spoznavne funkcije, kot so prepoznavanje vidnih ali slušnih vzorcev, kategorizacija in gibanje v prostoru. Kriza, ki je zajela umetno inteligenco, je spodbudila iskanje drugačnih rešitev. Nekateri so ostali pri simbolnih modelih, a so jih skušali na različne načine (npr. z dodajanjem utežnih parametrov k pravilom) narediti bolj prožne. V sredini osemdesetih let pa se je začel uveljavljati raziskovalni pristop, ki je za zgled jemal delovanje možganov in črpal iz začetnega raziskovanja nevronskih mrež.

KONEKCIONIZEM

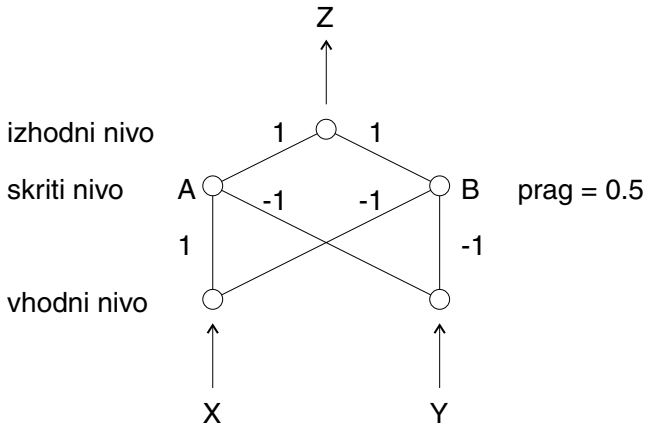
Konekcionistični pristop (nevronske mreže, paralelno distribuirano procesiranje) so navdihnila spoznanja nevroznanosti o možganih in ti modeli vključujejo pomembne značilnosti možganske arhitekture. Pri tem se ne ukvarjajo podrobno z delovanjem nevrona in nevronskih procesov, temveč skušajo zajeti delovanje možganov na bolj abstraktni ravni. Začetnika takih raziskovanj sta že omenjena McCulloch in Pitts. V 60-tih je Rosenblatt raziskoval dvonivojske mreže (Perceptron), ki pa so bile še dokaj omejene v nalogah, ki so jih lahko izračunale (npr. računanje funkcije izključujoči ali je bila pretrd oreh). Z odkritjem učnih algoritmov za večnivojske mreže v 80-tih letih so bile odpravljene nekatere omejitve in s tem odprta pot hitrejšemu razvoju. Proučevanje mrež pa tudi v vmesnem obdobju ni zamrlo, čeprav je bilo v senci klasične umetne inteligence. Naj omenimo le avtoasociativne mreže Kohonena in Hopfieldov vnos idej s področja fizike, predvsem statistične mehanike.

Osnovna značilnost konekcionističnega modela je, da je sestavljen iz preprostih, neinteligentnih enot (idealizirani nevroni), ki so medsebojno povezane. Vsaka enota ima določeno aktivacijsko vrednost, ki jo preko vezi, ki so različno močne, posreduje drugim enotam in s tem pripomore k povečanju ali zmanjšanju vrednosti teh drugih enot. Cel proces se odvija vzporedno in ne potrebuje nobenega osrednjega dela za nadzor.

Konekcionistični modeli (mreže) imajo lahko različne arhitekture. Ločimo jih glede na topologijo – koliko nivojev imajo (edega, dva ali več) in kako so enote med seboj povezane (enosmerno ali dvosmerno), glede na to, katero funkcijo aktivacije in katero izhodno funkcijo izberemo in glede na izbiro učnega pravila.

S pomočjo učnega pravila sistem postopno spreminja moč povezav. Proces učenja lahko poteka nenadzorovano ali pa nadzorovano. V prvi skupini so najbolj razširjeni algoritmi učenja, ki temeljijo na pravilu, ki ga je za učenje že leta 1949 predlagal D. Hebb. Po tem pravilu se med dvema enotama, ki sta istočasno aktivni, poveča moč povezave, v vseh drugih primerih pa se zmanjša. Na ta način se sistem samoorganizira. V drugo skupino spadajo delta pravila. Učni primeri so tu podani z vhodnimi vzorci in pravilnimi izhodnimi vzorci. Učenje poteka tako, da se v vsakem koraku izračuna izhod pri danem vhodu in razlika med dejanskim izhodom in željenim izhodom. Uteži povezav se potem zmanjšajo ali povečajo sorazmerno z razliko.

Natančneje se o možnih konfiguracijah konekcionističnih sistemov lahko poučimo v knjigi Rumelharta, McClellanda in PDP skupine raziskovalcev (1986), ki je postala nekakšna biblija konekcionističnega modeliranja, v slovenščini pa npr. v knjigi dr. Dobnikarja (1990). Tu si bomo ogledali le primer, kako trinivojska mreža izračunava funkcijo izključujoči ali.



Slika 1: Trinivojska mreža za "izključujoči ali"

X	Y	vhod v A	vhod v B	A	B	Z
0	0	0x1+0x(-1)=0	0x(-1)+0x1=0	0	0	0x1+0x1=0
1	0	1x1+0x(-1)=1	1x(-1)+0x1=-1	1	0	1x1+0x1=1
0	1	0x1+1x(-1)=-1	0x(-1)+1x1=1	0	1	0x1+1x1=1
1	1	1x1+1x(-1)=0	1x(-1)+1x1=0	0	0	0x1+0x1=0

Pri načrtovanju konekcionističnega modela ne začenjamo s simboli in pravili, ampak s preprostimi komponentami, ki so med seboj dinamično povezane. Vsaka enota deluje le v svojem lokalnem okolju, a vendar prihaja do neke vrste globalnega sodelovanja. To se pojavi spontano, takrat, ko vse sodelujoče enote dosežejo medsebojno zadovoljivo stanje. Prehod od lokalnih pravil do globalne usklajenosti in s tem pojav novih globalnih lastnosti je glavna značilnost dinamičnih mrež. Celoten proces poteka od spodaj navzgor in ne od zgoraj navzdol kot pri klasičnih modelih.

Kadar načrtujemo konekcionistično mrežo kot kognitivni model, moramo predstaviti pojme, ki so pomembni za to področje. To lahko naredimo na dva načina. Tako, da vsaki enoti pripišemo en pojem ali pa da je reprezentacija vsakega pojma porazdeljena po več enotah. Pri mrežah s porazdeljenimi reprezentacijami enote predstavljajo entitete, ki jih le težko opišemo v naravnem jeziku, in predstavljajo nekakšne mikroznačilnosti. Šele vzorcu kot celoti lahko pripišemo pomen. Take reprezentacije so torej holistične in odvisne od vsakokratnega konteksta, saj so prilagoditev sistema na zahteve okolja in so neposredna posledica učenja. Odslikujejo nalogo in podatke, na da bi bil za to potreben zunanji nadzor. Posledica tega je, da imajo taki modeli nekatere lastnosti, ki jih lahko najdemo tudi pri človeku, ne pa pri klasičnem modelu. Tako na primer poškodba nekaj enot ali vezi v splošnem še ne pomeni resnejše ovire za delovanje sistema, medtem ko izguba elementa v klasičnem modelu pomeni izgubo celotne informacije, ki jo je ta element nosil. Druga taka lastnost je posploševanje. Model, ki se je naučil določene naloge, bo na nov primer s tega področja odgovoril v skladu z znanjem, ki je implicitno spravljeno v povezavah. Novi vhodni vzorec bo uporabil obstoječe vezi in z njihovo pomočjo sestavil nov odgovor. To dejstvo lahko uporabimo tudi pri modeliranju spomina, kjer je mogoče, da sistem samo iz delnega vzorca ponovno sestavi celoto.

Omenjene značilnosti konekcionističnih sistemov so ugodne predvsem za modeliranje spoznavnih procesov, ki temeljijo na prepoznavanju vzorcev, za učenje s pomočjo primerov, za učenje veščin, torej povsod tam, kjer ne moremo natančno podati pravil in opisati postopka. Posledica tega je, da je

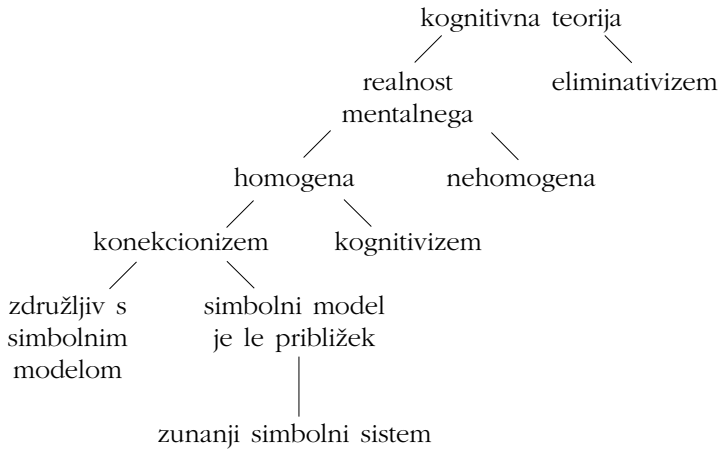
namesto reševanja problemov in logičnega sklepanja paradigmatsko raziskovanje postalo prepoznavanje vzorcev in učenje.

KONEKCIONIZEM, KOGNITIVIZEM IN NEVROZNANOST

Kot smo že omenili, se je za opisani način modeliranja uveljavilo več izrazov. Izraz nevronska mreža že z imenom opozarja na podobnost z možgani in velikokrat nevronske mreže res služijo modeliranju posameznih funkcionalnih delov možganov (npr. vizualni korteks). Če s svojimi raziskavami ostajajo neposredno pri možganih in se pri tem ne spuščajo v teoretiziranje, ki bi presegalo strukturalno modeliranje, ostajajo na področju nevroznanosti.

V psihološki in filozofski literaturi pa največkrat srečamo izraz konekcionizem, ki je včasih ime za način modeliranja (ime izhaja iz angleške besede connection – povezava), včasih pa označuje nov pristop k razlagi kognitivnih fenomenov. Ob tem se postavlja vprašanje, v kakšnem odnosu je konekcionizem do drugih takih poskusov, na eni strani do kognitivizma in klasičnih simbolnih modelov in na drugi do nevroznanosti.

Shematsko bi kognitivne teorije razdelili takole:



Pri oblikovanju kognitivne teorije moramo najprej zavzeti stališče do realnosti mentalnih stanj in procesov. Na eni strani imamo eliminativistične teorije, ki pravijo, da so naša psihološka stanja (prepričanja, namere,...) prazna, da se ne nanašajo na nič (npr. izraz "bolečina" je prazen, tako kot izraz "čarovnica"). Eliminativistični materialist trdi, da raziskovanja v nevroznanosti niso pokazala ujemanja med procesi v možganih in mentalnimi procesi. To je po njegovem mnenju razlog, da nadomestimo

mentalistični besednjak z govorom o stanjih možganov, in v ta namen predlaga razvoj znanstvene psihologije, ki se ne bo naslanjala na intencionalistično zdravorazumsko psihologijo. Dandanes najbolj znana zagovornika takega stališča, Paul in Patricia Churchland, menita, da je kognitivna nevroznanost najboljše upanje za razvoj znanosti o duhu. Če je stališče eliminativističnega materializma pravilno, potem moramo vse moči v kognitivni znanosti usmeriti na nevroznanost, saj lahko le ona nudi ustrezno razlago.

Scenarij, po katerem se moramo odpovedati osnovnim pojmom, s pomočjo katerih usmerjamo naše delovanje, je sicer možen, a se zdi malo verjeten. Na tem mestu se ne bomo spuščali v natančnejšo argumentacijo, raje si bomo pogledali, kakšne so alternative, če smo realisti v pogledu mentalnega.

Kognitivizem in konekcionizem sta dva različna pristopa, vendar ni nujno, da sta teoriji tekmici, saj imata lahko različni domeni raziskovanja. Takšno stališče vodi do nehomogene kognitivne teorije. Na eni strani imamo klasično simbolno paradigmo, ki se ukvarja s tem, kako ljudje razmišljamo s pomočjo pojmov, na drugi strani pa intuicijo in nižje spoznavne procese (zaznavanje in prepoznavanje vzorcev), ki jih lahko ustrezneje razlagamo s konekcionizmom. To se zdi na prvi pogled zelo sprejemljivo, saj se večina modelov res ujema s to delitvijo. Predpostavlja pa, da je duh razvil dva različna sistema, enega za obravnavanje zavestnega pojmovnega mišljenja in enega za intuicijo. Pri tem se takoj zastavi vprašanje, kako povezati rezultate pojmovnega mišljenja z vrstami intuitivnega mišljenja, ki ga razlaga konekcionizem. Ker na to vprašanje zaenkrat ni zadovoljivega odgovora, poglejmo še možnost, da konekcionizem in klasične kognitivne teorije razlagajo isto stvar. Potem se moramo odločiti za eno izmed možnosti, ali kognitivizem ali konekcionizem.

Kognitivizem kot starejša teorija se je čutil izzvanega od konekcionizma. Najbolj izčrpno sta v prid klasičnih modelov argumentirala Fodor in Pylyshyn (1988). Njuna izhodiščna teza je, da je tako kot jezik tudi misel sistematična in produktivna. Produktivnost se nanaša na zmožnost, da vedno lahko proizvedemo nove stavke ali nove misli, sistematičnost pa se nanaša na dejstvo, da če smo zmožni misliti določen stavek, smo avtomatično zmožni misliti še različne z njim povezane stavke. Na primer, če mislimo stavek “Maja ljubi Marka”, potem gotovo lahko mislimo tudi stavek “Marko ljubi Majo”. Simbolni sistem lahko zagotovi sistematičnost in produktivnost jezika, ker privzame notranji jezik misli. Notranje reprezentacije imajo kompozicionalno sintakso in semantiko. Če poznam pomene osnovnih gradnikov in sintaktična pravila, po katerih jih sestavljam, potem poznam tudi pomen celote. Konekcionizem pa sistematičnosti in produktivnosti ne more zagotoviti, saj pora-

zdeljene reprezentacije niso sintaktično strukturirane in tako ne omogočajo, da bi na njih uporabili strukturno občutljiva pravila. Porazdeljene reprezentacije kot vzorci v mrežah lahko predstavljajo le prisotnost ali odsotnost lastnosti ali objektov, ne pa relacij med njimi. Po mnenju Fodorja in Pylyshyna konekcionistični modeli zato ne morejo biti modeli duha. Lahko so le implementacija klasičnih modelov, ki pa kot implementacija niso pomembni za kognitivno razlago.

Kako lahko konekcionist odgovori Fodorju in Pylyshynu? Ena možnost je, da pokaže, kako je možno razviti simbolni sistem v konekcionistični arhitekturi. Z lokalistično mrežo, torej tako, kjer je ena enota ena reprezentacija, je to možno, kot sta pokazala Tauretzky in Hinton (1988). S tem ohranimo produktivnost in sistematičnost, ne moremo pa se otresti ugovora, da je to implementacija. Konekcionist ta ugovor sprejme, vendar hkrati izpodbija kognitivistično tezo, da implementacija ni pomembna za kognitivno raven. Konekcionistični modeli, ki veliko bolj upoštevajo funkcionalne lastnosti možganov, lahko omejujejo razred uresničljivih simbolnih modelov in tako dodajo nov kriterij za ustreznost modela. Nevroznanost tako dobi selekcijsko vlogo pri oblikovanju psiholoških teorij. Ta strategija je še blizu kognitivizmu, tako da se nanjo nanaša tudi večina ugovorov, ki se nanašajo na klasične modele, in ne prinaša revolucionarnih sprememb.

Bolj radikalen odnos do klasične kognitivne znanosti ponuja pristop, ki trdi, da je za nekatere procese simbolni sistem dovolj dober približek za človekov kognitivni sistem, da pa je to vendarle približek in da je natančen opis možen le na podsimbolni ravni, ki ustreza konekcionističnim modelom. Zagovornik take usmeritve, Smolensky (1988), primerja odnos med klasičnimi in konekcionističnimi modeli z odnosom med klasično mehaniko in kvantno fiziko, torej med makro in mikro teorijo.

Raven opisa, ki je primerna za konekcionistične modele, je med nevrološko in pojmovno. Smolensky zanika, da bi bili konekcionistični modeli nevronski modeli, čeprav so si sintaktično podobni – oboje vodijo splošni principi kompleksnega dinamičnega sistema in podobnost med enotami in nevroni ter povezavami in sinapsami je očitna. Večja pa je razlika v semantiki, saj lahko le malo povemo o odnosu med vzorci aktivnosti nad enotami, ki predstavljajo reprezentacijo spoznavne domene, in reprezentacijami nad nevroni v možganih. Podsimbolni modeli višjih procesov se ne morejo naslanjati na metodologijo "poglej, kako to delajo možgani", saj nas ta ne bi pripeljala daleč. Semantično so konekcionistični modeli bliže pojmovni ravni kot pa nevralni.

Na pojmovni ravni Smolensky loči dve vrsti obnašanja, zavestno uporabo pravil in intuicijo. Za prvo je značilno, da izraža znanje v taki obliki, da je splošno dostopno in ga lahko

uporabljajo tudi neizkušeni ljudje. Tako zavestno delovanje pri posamezniku lahko simuliramo na navideznem stroju, imenovanem zavestni interpret pravil. Za program, ki ga predstavlja kulturno znanje, so primerne jezikovne formulacije, saj omogočajo splošnost in zanesljivost izvajanja. To področje, kjer ljudje zavestno in sekvenčno sledijo pravilom, je mogoče dokaj uspešno modelirati s klasičnimi modeli. Druga vrsta obnašanja, ki obsega intuicijo, pridobivanje veščin in sploh individualno znanje, pa ni odvisna od zavestne uporabe pravil. Navidezen stroj – intuitivni procesor in procesi, ki tečejo na njem, so odgovorni za obnašanje živali in velik del obnašanja človeka, npr. zaznavanje, gibalne veščine. Klasično modeliranje je sledilo hipotezi, da imajo tudi programi na intuitivnem procesorju podobno semantiko in sintakso kot zavestni interpret pravil. Smolensky je tak pristop zavrnil in postavil hipotezo, da je kognicija konekcionistični dinamični proces, ki ne omogoča popolne in natančne pojmovne ravni opisa.

Če je klasična kognitivna teorija dober približek za procese, ki zahtevajo zavestno pojmovno mišljenje, mora nova podsimbolna teorija pokazati, kakšen je razvoj, ki je do tega pripeljal. Poleg tega mora še odgovoriti na Fodorjev in Pylyshynov argument in pokazati, kako je v takem sistemu možno sistematično in produktivno obnašanje. Tako Smolensky kot drugi raziskovalci se trudijo izdelati modele, ki bi kazali te lastnosti, obeta predvsem razvoj mrež, ki s povratno povezavo omogočajo informacijo o predhodnem ciklu ("recurrent networks": Elman, 1990; Pollack, 1990).

Zanimivo hipotezo o tem, kako v sistemu pridemo do sistematičnega obnašanja, sta podala Bechtel in Abrahamsen (1991). Njun predlog je pristop, v katerem formalne simbole potegnemo iz sistema in jih postavimo v okolje. Kar obstaja znotraj sistema, niso notranje reprezentacije teh zunanjih simbolov, ampak zmožnost sistema, da iz njih razbere informacijo in potem proizvede simbol, ki se pokorava sintaktičnim pravilom. V primeru človeka je tak zunanji simbolni sistem jezik.

Konekcionistični modeli poleg tega, da so biološko bolj verjetni, zbujajo tudi upanje, da so zmožni odgovoriti na problem notranje intencionalnosti, ki ga klasični simbolni modeli niso uspešno rešili in na katerega je opozarjal J. Searle z znanim miselnim eksperimentom "Kitajska soba". Če si zamislimo robota, ki bi bil s senzorji povezan z okoljem, potem bi notranje reprezentacije, ki bi spontano nastale v robotu kot odgovor na robotovo prilagajanje in učenje, bile povezane s svetom, bile bi o nečem v svetu. Veliko vprašanje, ki ga mora konekcionizem rešiti, pa je, kako lahko sistem te reprezentacije kasneje uporablja v drugih nalogah, ki niso povezane s procesom, v katerem je reprezentacija nastala.

ZAKLJUČEK

V kognitivni znanosti danes prevladujeta predvsem dva pristopa za pojasnitev kognitivnih fenomenov, kognitivizem in konekcionizem. Znanstveniki obeh pristopov skušajo spoznavne funkcije simulirati s pomočjo računalnika. Kognitivist se pri tem naslanja na formalno logiko, konekcionist pa na matematiko dinamičnih sistemov. Kognitivistova teza je, da je za razlago obnašanja potrebna posebna kognitivna raven, raven mentalnih reprezentacij. Te imajo notranjo sintaktično strukturo in tako omogočajo, da se semantične lastnosti reprezentacij kažejo v sintaktičnih in s tem fizičnih lastnostih ter tako igrajo vzročno vlogo pri generiranju obnašanja. Porazdeljene reprezentacije konekcionističnega sistema nimajo take strukture, saj semantični vsebini izraza ne ustreza struktura, ki na transparenten način kaže sestavne dele izraza. Zato se postavlja vprašanje, kako interpretirati konekcionistične modele. Podali smo le skico alternativ, od katerih pa bi vsaka zahtevala podrobnejšo analizo, ki bi vključevala dognanja vseh disciplin kognitivne znanosti.

Olga Markič, mlada raziskovalka in asistentka na Oddelku za filozofijo Filozofske fakultete v Ljubljani.

LITERATURA

BECHTEL, W. (1988): Philosophy of Mind: An Overview for Cognitive Science, Lawrence Erlbaum Associates, Hillsdale, NJ.

BECHTEL, W. in ABRAHAMSEN, A. (1991): Connectionism and the Mind, Basil Blackwell, Oxford.

CHURCHLAND, P. M., (1989): A Neurocomputational Perspective: The Nature of Mind and the Structure of Science,MIT Press, Cambridge, MA.

DOBNIKAR, A. (1990): Nevronske mreže, Didakta, Radovljica.

ELMAN, J. L. (1990): "Finding structure in time", Cognitive Science 14, 179-211.

FODOR, J. A. in PYLYSHYN, Z. W. (1988): "Connectionism and Cognitive Architecture: A Critical Analysis", Cognition 28, 3-71.

GARDNER, H. (1987): The Mind's New Science: A History of Cognitive Revolution, Basic Books, New York.

HAUGELAND, J. (1985), Artificial Intelligence: The very idea, MIT Press, Cambridge, MA.

McCULLOCH, W. S. IN PITTS, W. H. (1943): "A Logical Calculus of the Ideas Immanent in Nervous Activity", v M. Boden (ur.) The Philosophy of Artificial Intelligence, Oxford University Press, Oxford.

POLLACK, J. (1990): "Recursive Distributed representations", Artificial Intelligence 46, 77-105.

RUMELHART, D. E., McCLELLAND, J. L, and the PDP Research Group (1986): Parallel Distributed Processing: Explorations in Microfeatures of Cognition, vol. 1&2, MIT Press, Cambridge.

SEARLE, J. R. (1980): "Duhovi, možgani in programi", v: D. Dennett in D. Hofstadter (ur.), Oko Duha, (1990), Mladinska knjiga, Ljubljana.

SMOLENSKY, P. (1988): "On the Proper Treatment of Connectionism", Behavioral and Brain Sciences 11, 1-74.

TAURETZKY, D. S. in HINTON, G. E. (1988): "A Distributed Connectionist Production System", Cognitive Science, 12, 423-466.

TURING, A. (1950): "Stroji, ki računajo, in inteligenca", v: D. Dennett in D. Hofstadter (ur.), Oko Duha, (1990), Mladinska knjiga, Ljubljana.