HUGUR 7. ÁR, 1994-1995 s. 115-133

Jörgen Pind

Efnisleg táknkerfi

Tölvubylting undanfarinna áratuga hefur með margvíslegum hætti gripið inn í iðkun ýmissa fræðigreina. Þetta á ekki síst við í sálfræði þar sem „tölvulíkanið" hefur ýtt mjög við ímyndunarafli sálfræðinga og átti reyndar á sínum tíma snaran þátt í því að kollvarpa atferlishyggju Skinners. Í þessari grein er ætlunin að varpa nokkru ljósi á upptök gervigreindar og vekja athygli á þeirri staðreynd að sálfræðilegar vangaveltur voru mörgum brautryðjendum gervigreindar afar hugleiknar. Sömuleiðis verður vakin athygli á því að John Searle varð fráleitt manna fyrstur til þess að stíga fæti inn í Kínaherbergið. Það gerðu Newell og Simon þegar árið 1956 - þótt þeir hafi að vísu snúið þaðan með allt öðru hugarfari en Searle. Bak við kenninguna um „gamaldags gervigreind" búa hugmyndir um efnisleg táknkerfi. Fjallað verður um eðli þeirra og raktar nokkrar ástæður þess að vegur þeirra hefur dvínað að undanförnu.

Inngangur

Ein markverðasta nýbreytni í sögu vísindalegrar sálfræði á síðari árum eru þau nánu tengsl sem skapast hafa milli sálfræði og tölvufræða. Þessi tengsl hafa ekki síst lotið að kenningasmíði í sálfræði því tölvan hefur opnað nýjar leiðir til þess að *prófa* slíkar kenningar.[1]

Það er vitaskuld engin nýlunda að heilanum sé líkt við margvísleg tæki. Þannig voru samlíkingar heila og símstöðva á fyrri hluta þessarar aldar vinsælar. Slík líkingasmíð hefur hins vegar reynst hafa frekar takmarkað gildi. Reyndar hafa líkingarnar ekki einskorðast við tæki. Þannig eru þekkt þau ummæli bandarísku sálfræðingsins Tolmans: „[Heilinn] er miklu líkari kortaherbergi en gamaldags

1 Þessi grein var að mestu skrifuð meðan höfundur dvaldi í rannsóknaleyfi við MIT veturinn 1993-1994. Mér eru minnistæð þau orð sem einn sálfræði-prófessorinn þar, Whitman Richards að mig minnir, lét falla í fyrirlestri: „Hver er ástæða þess að sett var á laggirnar sálfræðideild við Tæknihaskólann í Massachusetts? Það getum við þakkað einu tæki, tölvunni." - Ég vil þakka Ólafi Páli Jónssyni einkar gagnlegar ábendingar við fyrri gerð þessarar greinar.

símaskiptiborði," sem hann varpaði fram í grein rétt eftir lok síðari heimsstyrjaldarinnar.

Tilkoma tölvunnar, sem virtist vera fær um að leysa margvísleg verkefni sem áður var talið að yrðu ekki leyst án vitsmuna, hefur orðið mörgum fræðimanninum hvatning til samlíkingar hugar og tölvu. Einna fyrstur til að velta þessum möguleika fyrir sér var breski stærðfræðingurinn Alan Turing eins og rakið er í grein Atla Harðarsonar í þessu hefti *Hugar*.

Turing (1950) hafði nokkra sérstöðu í viðhorfi sínu til samlíkingar hugar og tölvu því að hann einblíndi fyrst og fremst á samlíkingu *hugbúnaðar* og hugarstarfsemi frekar en samlíkingu *vélbúnaðar* og heilastarfsemi. Þessi samlíking átti upphaflega ekki miklu fylgi að fagna því flestum var samlíking tölvu og heila tamari eins og ráða má af því að á fyrstu dögum tölvunnar voru þær gjarnan nefndar „rafheilar".

Ungverski stærðfræðingurinn John von Neumann gerði þá samlíkingu að sérstöku umfjöllunarefni í þekktri bók *The Computer and the Brain* (von Neumann 1958). Þar vekur hann athygli á því að taugafrumur hegða sér á svipaðan hátt og stafrænar rásir, annað hvort senda þær boð eða ekki, og því megi réttilega lýsa heilanum, a.m.k. við fyrstu sýn, sem *stafrænu líffæri*. Eitt af því sem hægt er að áætla fyrir stafrænt líffæri (með hliðsjón af kennisetningum upplýsingafræðinnar) er rýmd líffærisins, hversu miklar upplýsingar það getur rúmað. John von Neuman áætlaði að í heilanum væru 10^{10} frumur og að hver þeirra gæti flutt um 14 taugaboð á sekúndu (lágmarkstími fyrir hvert taugaboð er um 70 þúsundustu úr sekúndu). Þá má áætla heildarílagið sem 14×10^{10} bita á sekúndu. Sé enn fremur gert ráð fyrir engri gleymsku í heilanum, sem von Neumann taldi allar líkur á, og að meðalmannsævi sé 60 ár (2×10^9 sekúndur) þá er auðvelt að áætla heildarumfang þeirra upplýsinga sem mannsheilinn meðtekur að meðaltali á einni mannsævi, $14 \times 10^{10} \times 2 \times 10^9 = 2{,}8 \times 10^{20}$ bitar.

Þetta þótti von Neumann nokkuð há tala og hæfilega miklu hærri en tölvutæknin þá réði við. En nú hefur tölvutækninni fleygt fram. 10^{20} bitar jafngilda $1{,}25 \times 10^{19}$ bætum eða 10^7 gígabætum (tíu milljónum gígabæta). Þegar öllu er á botninn hvolft er þetta þó kannski ekki svo há tala, því nú fást diskar er rúma 1 gígabæti í venjulega heimilistölvu.

Því er ekki að undra að þeir sem hafa lagt slíka útreikninga niður fyrir sér á seinni árum hafa fengið nokkuð aðra útkomu en von Neumann. Einn þeirra er stærðfræðingurinn Jacob T. Schwartz (1988) sem hefur nýlega reynt að reikna svipað dæmi miðað við nýjustu upplýsingar í taugalífeðlisfræði. Hann kemst að því að heilinn framkvæmi um 10^{21} aðgerðir á sekúndu (en ekki 14×10^{10} eins og von Neumann áætlaði).

En þessir útreikningar sýna líka að samlíking tölvu og heila að þessu leytinu er ekki sérlega *upplýsandi*. Væntanlega fýsir okkur flest að vita eitthvað um starfshætti taugafrumna, t.d. hvernig taugakerfið megnar að stilla saman tvær ólíkar myndir sem falla á sjónur augnanna þannig að úr verði ein mynd með dýpt. Magnreikningar af því tagi sem von Neumann iðkaði, hvort sem þeir eru réttir í smáatriðum eða ekki, varpa engu ljósi á slíkar spurningar.

Einn er þó sá þáttur í útreikningum von Neumans sem ástæða er til að staldra við því þar bendir hann á grundvallarmun sem er á tölvu og heila. Hann telur fullljóst að heilinn verði að framkvæma útreikninga sína með mikilli nákvæmni, slíkt sé eðli þeirra verkefna sem heilinn þurfi að leysa (skynjun, stjórnun hreyfinga o.s.frv.). Rannsóknir hafi hins vegar sýnt að heilinn skráir upplýsingar, t.d. um styrk áreita, með tíðni taugaboða. Þessi tíðni sé breytileg á bilinu 50 til 200 boð á sekúndu. Sú breidd í svörun felur hins vegar í sér afar litla nákvæmni. Af þessu dregur von Neumann þá ályktun að *heilinn beiti ónákvæmum aðferðum*. Slíkt getur reyndar haft sína kosti. Í útreikningum í tölvu má yfirleitt engu muna til að útkoman verði ekki rétt. Þessu er vitaskuld allt öðru vísi farið um heilann, hann getur starfað eðlilega þrátt fyrir ótrúlega breytileg ytri skilyrði. Niðurstaða von Neumans er því sú að vissulega beri taugafrumur svip stafrænna rása en starfshættir heilans séu hins vegar að verulegu leyti frábrugðnir starfsháttum tölvu.

Í ljósi þessa er ekki að undra að þegar menn fóru fyrst af alvöru að huga að möguleikum gervigreindar, að gæða tölvur skynsamlegu viti, varð þeim starsýnt á samlíkingu hugar og hugbúnaðar frekar en heila og vélbúnaðar. Sálfræðileg sjónarmið komu þar verulega við sögu.

Rætur gervigreindar

Áhugamönnum um sögu gervigreindar ber almennt saman um að rekja megi upphaf greinarinnar til ráðstefnu sem haldin var sumarið 1956 við Dartmouth-háskólann í Bandaríkjunum (McCorduck 1979; Gardner 1987; Crevier 1993). Aðalhvatamenn Dartmouth-ráðstefnunnar voru þeir John McCarthy og Marvin Minsky. Auk þeirra voru þar aðrir sem taldir voru líklegir til að geta lagt eitthvað af mörkum við mótun greinarinnar. Þeir reyndust ekki margir, þátttakendur fylltu rétt tuginn. Einn þeirra var Oliver Selfridge sem hafði gert merkar tilraunir til þess að lesa rithönd með vélrænum hætti og þannig skilgreint mörg af þeim vandamálum sem vélræn *mynsturgreining* (*pattern recognition*) varð að glíma við (Selfridge (1958), sbr. einnig Neisser (1967)). Mesta athygli vakti þó framlag Herberts Simons og Allans Newells sem komu frá RAND-rannsóknastofnuninni í Kaliforníu, því þeir höfðu einir þátttakenda í farteskinu tilbúið „skynugt forrit", forrit sem gat leitt út sannanir í rökfræði og stærðfræði. Nafn forritsins, *Rökfræðingurinn* (*Logic theorist*), var heldur ekki valið af neinni hógværð.

Þáttur Newells og Simons

Rannsóknir Newells og Simons hafa haft gagnger áhrif á þróun gervigreindar á undanförnum árum og þá ekki síður á þróun hugfræðinnar. Það vekur í raun nokkra furðu því hvorugur þeirra er menntaður sálfræðingur. Áður en áhugi Simons vaknaði á hugfræði hafði hann unnið merkt brautryðjendastarf í stjórnunarfræðum sem átti síðar eftir að færa honum Nóbelsverðlaunin í hagfræði (1978). Simon var prófessor í hagfræði við stjórnunardeild Carnegie Technical University í Pittsburgh þegar leiðir hans og Newells lágu saman við fyrrgreinda RAND-stofnun í Kaliforníu. Newell, sem hafði horfið frá stærðfræðinámi, glímdi þar við að búa til hermi af *heilli flugstöð* fyrir bandaríska flugherinn. Eins og Simon (1991) lýsir því í sjálfsævisögu sinni var það einkum aðferð Newells, og samstarfsmanns hans John Shaws, við að nota tölvu til að draga á skjá eftirhermur radarmynda sem vakti athygli hans á því að nota mætti tölvur til hvers kyns *táknunar*, ekki bara til þess að meðhöndla tölur. Úr þessum jarðvegi spratt síðan sú hugmynd að nota mætti tölvurnar

til þess að herma eftir hvers kyns flóknum ferlum, þ.á m. mannlegri hugsun. Þetta mun hafa verið árið 1954. Á RAND-stofnuninni tóku svo þeir þremenningar höndum saman við gerð skynugra forrita.

Fyrst sneru Newell, Shaw og Simon sér að skákforritun. Þeir ráku sig hins vegar fljótlega á að stöðumat í skák er býsna flókið fyrirbæri. Einkum vafðist fyrir þeim með hvaða hætti mætti forrita „skynjun" skákmanna á stöðum. Af þessu sést hvaða áherslu Newell og Simon lögðu á samsvörun forrita og mannlegrar hegðunar, í þessu tilviki skákmeistara. Vegna þessara vandkvæða beindist áhugi þeirra fljótt að „hreinum" hugarferlum, áþekkum þeim sem beitt er við rökhugsun. Fyrir valinu varð höfuðrit Russells og Whiteheads *Principia Mathematica* og einsettu þeir sér að útbúa forrit er gæti leitt út a.m.k. sumar þær sannanir sem er að finna í því riti. Við forritunina beittu þeir þeirri aðferð að Simon leiddi út sannanirnar, þrep fyrir þrep, og Newell og Shaw forrituðu. Newell og Shaw þurftu reyndar að leysa margvísleg vandamál varðandi forritunina sem snertu ekki *Rökfræðinginn* sem slíkan, heldur hvers kyns táknræna forritun (lista-forritun, sem þá var algerlega ókannað svið innan tölvufræðinnar).

Það var svo í ágúst árið 1956 sem fyrsta sönnunin birtist á prentara tölvunnar. Simon sendi Bertrand Russell sönnunina um hæl og Russell svaraði:

> Ég er himinlifandi að frétta að *Principiu Mathematicu* sé nú hægt að framkvæma með vélum. Ég vildi að við Whitehead hefðum vitað af þessum möguleika áður en við sóuðum 10 árum í að reikna þetta í höndunum. Ég er fyllilega tilbúinn til að trúa því að allt í afleiðslurökfræði megi framkvæma með vélum (Simon 1991, bls. 208).

Rökfræðingnum tókst að sanna 38 af fyrstu 52 setningunum í öðrum kafla *Principiu Mathematicu*.

En Simon taldi reyndar að þeir félagar hefðu afrekað snöggtum meira en að sýna fram á að hægt væri að leiða út sannanir í rökfræði með vélrænum hætti. Og hann heldur enn fast við þá skoðun:

> Hægt er að orða þetta á tæknilegri og jafnframt sjálfshælnari hátt. Við fundum upp forrit sem gat hugsað með táknum og leystum þar með hina æruverðugu gátu um samband hugar og líkama, skýrðum hvernig efnislegt kerfi getur haft eiginleika hugar (Simon 1991, bls. 190).

Von bráðar var athygli sálfræðinga vakin á tilraunum Newells, Shaws og Simons til að herma eftir mannlegum vitsmunum með aðstoð tölvu. Þegar árið 1958 skrifuðu þeir grein í *Psychological Review*, helsta tímariti fræðilegrar sálfræði, þar sem þeir greindu frá *Rökfræðingnum* og lögðu einkum áherslu á tengsl eigin rannsókna við klassískar rannsóknir í sálfræði hugsunar, aðallega rannsóknir Dunckers og de Groots (Newell, Shaw og Simon, 1958). Tveim árum síðar skrifuðu Miller, Galanter og Pribram bókina *Plans and the Structure of Behavior* þar sem lagt var til atlögu við atferlishyggju Skinners með aðferðafræði Newells og Simons (og að nokkru leyti málkunnáttufræði Chomskys) að vopni. Hugfræðin hafði fljótt fullnaðarsigur í þeirri baráttu eins og kunnugt er.

Í grein sinni lögðu þeir félagar áherslu á mikilvægi kenningar um upplýsingavinnslu (*information processing*) sem sjálfstæða tilraun til skýringar á starfsemi mannshugarins, óháða taugalíffeðlisfræði. Slík kenning um upplýsingavinnslu felur í sér þrjá þætti:

1. Stýrikerfi sem er sett saman úr nokkrum minniseiningum sem geyma táknrænar upplýsingar.
2. Tiltekinn fjölda grunnaðgerða sem meðhöndla upplýsingarnar í minniseiningunum. Eiginleikar þessara grunnaðgerða eru ekki til skoðunar, þeir eru gefnir, en þó er ekkert dularfullt við þessar grunnaðgerðir. Þær eru aðgerðir sem vélbúnaðurinn framkvæmir samkvæmt þekktum lögmálum eðlisfræðinnar.
3. Mengi af ótvíræðum reglum sem flétta grunnaðgerðirnar saman í forrit.

Síðan segir:

> Á þessu kenningaplani er hegðun lífveru skýrð með forriti sem byggir á frumstæðri upplýsingavinnslu sem leiðir af sér hegðunina.

Hér kemur greinilega fram að höfundarnir hafa engan áhuga á þeirri líkingu tölvu og mannsheila sem hafði ýtt svo við ímyndundarafli von Neumans. Þeir skipa sér á bekk með Turing og líkja mannlegum hugsanaferlum við gang forrits.

Newell og félagar héldu því fram í grein sinni að *Rökfræðingurinn* hegðaði sér oft svipað og fólk sem glímir við þrautir. Þessi hegðun

var þekkt af rannsóknum sálfræðinga og hafði fengið heiti á borð við „tilbúnað" (*set*) og innsæi. Tilbúnaður er skýrt sem vanabundin hneigð til að svara á einhvern tiltekinn hátt, og getur oft hamlað því að fólk finni lausn á viðfangsefnum sínum. Að mati Newells og félaga verður nákvæmlega sömu hegðunar vart hjá *Rökfræðingnum*, hann notar þær aðferðir sem hann ræður yfir í ákveðinni röð og tæmir möguleika hverrar aðferðar áður en aðrar eru reyndar.

Þetta svipmót í hegðun manns og tölvu höfðu Newell og Simon til marks um að þeir væru á réttri braut. Í þeirra huga lék aldrei neinn vafi á því að réttast væri að byggja forritunina á nákvæmri könnun á mannlegri hegðun. Þetta varð enn frekar áberandi í síðari rannsóknum þeirra sem birtust í bókinni *Human Problem Solving* árið 1972. Þar er að finna greinargerð um forrit (*þrautakónginn*, *General Problem Solver*) sem byggt var á nákvæmri könnun á „hugsanaklöddum" (*verbal protocols*) frá þátttakendum í margvíslegum sálfræðitilraunum.

Efnisleg táknkerfi

Sú gervigreind sem Newell og Simon urðu einna fyrstir til að þróa fékk síðar viðurnefnið „gamaldags gervigreind" (GOFAI - Good Old Fashioned AI) (Haugeland 1985). Að baki kenningum í gamaldags gervigreind býr ákveðin sýn á eðli vitsmuna og enn voru það Newell og Simon sem urðu til þess að skilgreina þessa sýn (Newell og Simon 1976).

Meginhugmynd þeirra snýst um *efnisleg táknkerfi* og þeir staðhæfa að efnisleg táknkerfi gegni sama hlutverki innan gervigreindar og landrekskenningin í jarðfræði, frumukenningin í líffræði og bakteríukenningin í læknisfræði.

Tákn verða á vegi okkar á hverjum degi og hæfileikum mannshugans til táknunar virðast lítil takmörk sett. Táknkerfi eru margvísleg. Tvö þeirra eru kannski öðrum mikilvægari, tákn talna í talnakerfi og tákn málhljóða í ritkerfum tungumála.

Flest táknkerfi eru með þeim hætti að samband tákns og táknmiðs, þ.e. hins táknaða, ræðst einvörðungu af hefð, ekki af eiginleikum táknsins eða svipmóti þess og táknmiðsins. Þetta er t.d. ein ástæða þess að tákna má tölur með margvíslegum hætti, svo sem með rómverskum tölum

XLV,

arabískum tugakerfistölum

45,

eða tvíundartölum

101101.

Táknun með orðum er með sama hætti, samband orðsins „hundur" við fyrirbærið ræðst einvörðungu af hefð eins og sést af því að í öðrum málum er fyrirbærið nefnt nöfnum eins og „dog" eða „chien".

Ekki eru þó öll táknkerfi með þessum hætti; umferðarmerki eru t.d. þannig gerð að menn eiga að geta ráðið af útliti þeirra hver merking þeirra er án þess nauðsynlega að hafa *lært* hana. Táknið á að vera gagnsætt, að *fela í sér merkinguna*. Umferðarmerki sem sýnir bíl renna til í hálku eða bleytu er þess háttar tákn. En þó er það ekki ótvírætt. Það gæti jú allt eins merkt að menn skyldu gæta sín á drukknum ökumönnum eða að menn hafi slysast inn á rallbraut.

Samband tákna í tölvum við táknmið helgast einvörðungu af hefð og samkomulagi tölvuframleiðenda. Þegar ég sit og rita þessi orð á tölvuskjáinn, styð t.d. á hnappinn „n", birtist jafnóðum stafurinn „n" á skjánum. En í innviðum tölvunnar er ekkert „n" heldur einvörðungu tala sem samsvarar „n", nánar tiltekið talan 110 í tugakerfi. Það er hins vegar ekkert sjálfsagt við að „n" skuli vera geymt með tölunni 110 í minni tölvunnar. Ástæða þess er einvörðungu sú að tölvan notar tiltekið stafróf, oft kennt við ASCII. Ef stafrófið væri annað, t.d. EBCDIC sem IBM notar á stærri tölvum sínum, væri „n" geymt með öðru tölugildi.

Upphaflega voru tölvur búnar til sem reiknivélar (eins og heiti þeirra ber með sér) en það leið ekki á löngu áður en menn komu auga á að þær mætti nota sem *táknvélar*. Þessi skilningur á eðli tölvunnar opnaði fyrir hvers kyns hagnýtingu sem menn sáu kannski ekki fyrir í upphafi (nema Turing, honum var ætíð ljóst að þessar vélar væru táknvélar). Af þessum sökum geta menn notað tölvurnar sem teiknivélar, til ritvinnslu, til að spá fyrir um veður og tefla skák, svo aðeins fátt eitt sé talið.

Á grundvelli þessa stigu Newell og Simon næsta skref, og staðhæfðu það sem var væntanlega skilningur allra þeirra sem sýsluðu við gervigreind á þessum árum:

Tilgátan um efnisleg táknkerfi: efnislegt táknkerfi býr yfir nauðsynlegum og nægjanlegum búnaði til almennra vitsmunalegra athafna (Newell og Simon, 1976).

Þessi tilgáta felur í sér tvennt. Annars vegar að öll „kerfi" (þar með mannshugurinn) sem búa yfir vitsmunum séu í eðli sínu efnisleg táknkerfi, hins vegar að gæða megi sérhvert efnislegt táknkerfi skynsamlegu viti. Af þessu leiðir að enginn eðlismunur er á mannsheila og tölvu sem er rétt forrituð (sbr. einnig Newell, Young og Polk (1993)).

Mikilvægi réttrar táknunar

Tákna má fyrirbæri, verkefni og þrautir með margvíslegu móti og oft skiptir miklu máli fyrir meðhöndlun þeirra hvaða táknun er valin. Eftirfarandi dæmi (Norman 1993) sýnir það einkar skýrt.

Ímyndum okkur leik sem gengur undir heitinu „15". Þátttakendur eru tveir sem skiptast á að velja tölu; sá sem er fyrri til að velja einhverjar þrjár tölur sem samanlagt eru 15 vinnur. Aðeins má velja á milli eftirfarandi talna: 1, 2, 3, 4, 5, 6, 7, 8 og 9. Hverja tölu er aðeins unnt að velja einu sinni.

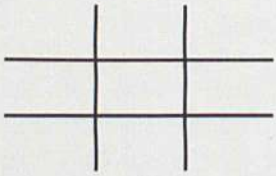
Ímyndum okkur því næst að leikurinn hafi spilast sem hér segir: Fyrri leikmaður, A, velur 7, seinni leikmaður, B, velur 3, A velur 2, B velur 6, A velur 5.

Hvaða tölu á B að velja þegar hér er komið sögu? Væntanlega vefst fyrir lesendum, öðrum en þeim sem eru sérlega stærðfræðilega sinnaðir, hvaða tölu B eigi að velja hér til að koma í veg fyrir að A vinni. Ástæðan er vitaskuld sú að leikmenn verða að leggja margar tölur á minnið og reikna summu þeirra. Ef tölurnar eru tvær er það auðvelt. Þannig vefst væntanlega ekki fyrir neinum að sjá að þegar A hefur valið 7 og 2 þarf hann 6 til viðbótar til að vinna leikinn (7+2+6=15). Af þessum sökum velur B 6 þegar þar er komið sögu. A velur 5 og þar sem hann hefur nú valið þrjár tölur eru ýmsar leiðir til að fá summuna 15 með einni tölu til viðbótar og leikmaður B verður að huga að öllum. Tölurnar sem A hefur valið eru 7, 2 og 5. Með einni tölu til viðbótar eru alls þrír möguleikar á að því að vinna leikinn:

$$7 + 2 + \mathbf{6} = 15$$
$$7 + 5 + \mathbf{3} = 15$$
$$5 + 2 + \mathbf{8} = 15$$

En þar eð tvær tölur eru þegar fráteknar, stendur í raun aðeins ein eftir og B ætti því að velja hana til að koma í veg fyrir að A vinni. Það er talan 8.

Beinum næst sjónum okkar að öðrum leik sem lesendur kannast væntanlega flestir við. Þessi leikur er leikinn með blaði og blýanti, með krít á töflu, með því að krota í sand eða öðrum svipuðum aðferðum. Leikurinn fer fram á leikborði sem er eins og hér er sýnt:



Tveir leikmenn skiptast á að merkja sér reiti, annar með X-um, hinn með O-um. Sá vinnur sem fyrstum tekst að merkja sér heila línu (lóðrétta, lárétta eða skálínu).

Setjum sem svo að eftirfarandi staða hafi komið upp í leik og B (sem merkir sér reiti með O) á leik - hvað á hann að gera?

	O	
	X	
X	X	O

Svarið „liggur í augum uppi" - B merkir sér reitinn efst til hægri.

X-O leikurinn er barnaleikur einn og börn eru fljót að læra hann til þeirrar hlítar að geta alltaf haldið jöfnu. Reikningsþrautin sem fyrr var nefnd er hins vegar býsna flókin og vefst væntanlega fyrir flestum að ná sambærilegri leikni í honum. En athyglisvert er að hér er í raun um *sama leikinn* að ræða, aðeins ólíka táknun hans. Hægt er að sannfæra sig um þetta með því að gefa reitunum í seinni leiknum tölugildi:

4	3	8
9	5	1
2	7	6

Lesendur geta sannreynt að staðan í seinni leiknum að ofan er í reynd sú sama og upp kom í talnaleiknum fyrr með því að finna tölugildi X-anna og O-anna með hliðsjón af þessari uppsetningu á leikborðinu. A hafði valið sér tölurnar 7, 2 og 5, en B 3 og 6. Rökréttur leikur fyrir B var að velja 8, reitinn efst í hægra horni.

Þetta dæmi sýnir einkar glögglega að huga verður að táknun þeirri sem notuð er hverju sinni við forritun. Táknanirnar eru formlega jafngildar en sálfræðilega eru þær það alls ekki. Gervigreindarfræðingum bar að vísu ekki saman um hvaða vægi ætti að veita sálfræðilegum röksemdum af þessu tagi við forritsgerð. Í huga Newells og Simons lék aldrei neinn vafi á því, velja ætti þá táknun sem væri sálfræðilega réttmætust.

Kínaherbergið

Það er kunnara en frá þurfi að segja að margir hafa orðið til þess að gagnrýna gervigreind og frá margvíslegum sjónarhólum. Óhætt mun þó að fullyrða að gagnrýni bandaríska heimspekingsins Johns Searles hafi vakið hvað mesta athygli. Hún var fyrst sett fram í frægri grein í tímaritinu *Behavioral and Brain Sciences* árið 1980, sem birtist í íslenskri þýðingu í þessu hefti Hugar.

Í þessari grein gerir Searle greinarmun á tvenns konar gervigreind, róttækri og hógværri. Gagnrýni hans snýst einvörðungu um hið róttæka afbrigði greinarinnar. Róttæk gervigreind felur í sér þrjár kennisetningar að mati Searles. Í fyrsta lagi gerir hún ráð fyrir að samband hugar og heila sé sama eðlis og samband hugbúnaðar og vélbúnaðar í tölvum. Í öðru lagi gerir kenningin ráð fyrir að hugsun feli einvörðungu í sér meðhöndlun tákna. Þessar tvær forsendur leiða síðan til þeirrar niðurstöðu að rétt forritaðar tölvur, tölvur sem keyra málskilningsforrit t.d., hugsi í raun og veru en séu ekki einberar hermivélar. Þar sem hugsun byggist einvörðungu á meðhöndlun tákna

og tölvur eru samkvæmt skilgreiningu táknvélar, þá sé óyggjandi að tölvur hugsi í raun og veru, rétt eins og menn.

Searle tekur sem dæmi um slíkt forrit eitt af málskilningsforritum Schanks (1975) sem að mati Schanks getur sett sig í spor viðskiptavina á hamborgarastöðum í Vesturheimi. Forritið fær sögu eins og þessa til að glíma við: „Jón fór á hamborgarastað. Hamborgarinn var viðbrenndur svo hann yfirgaf staðinn án þess að borga". Ef forritið getur svarað því til að líklega hafi Jón ekki borðað hamborgarann (en um það segir ekkert í sögunni) munu talsmenn róttækrar gervigreindar halda því fram að forritið hafi *skilið* söguna þar sem svör forritsins séu skynsamleg og viðeigandi. En hvernig ber tölvan sig að því að svara þessum spurningum? Jú, tölvan fylgir einvörðungu skipunum þess forrits sem hún er mötuð á og þar eð hún er formleg táknvél verður að álykta að skilningur feli einfaldlega í sér meðhöndlun tákna samkvæmt forskrift.[2]

Röksemdafærsla Searles leiðir hins vegar í ljós að hér er ekki allt sem sýnist. Hún er sem hér segir í lauslegri endursögn.

> Ímyndum okkar að ég sé læstur inni í herbergi og fyrir framan mig á borði er stafli af blöðum með kínverskum táknum. Ímyndum okkur enn frekar, sem er satt og rétt, að ég kunni enga kínversku og skilji því ekki táknin fyrir framan mig. Þegar hér er komið sögu fæ ég í hendur enn einn blaðastaflann með kínverskum táknum en líka blöð með upplýsingum um það hvernig sambandi táknanna í stöflunum tveim er háttað. Þessar reglur eru ritaðar á íslensku og ég skil þær. Þar stendur t.d. eitthvað á þá leið að sé mér rétt blað með krissi og krassi skuli ég leita að blaði með krassi og krussi í fyrri bunkanum. Finni ég það sé mér óhætt að láta af hendi eitt af blöðunum í öðrum bunkanum með krassi, krissi og krissi, en að öðrum kosti skuli ég bíða eftir næsta blaði að utan. Og viti menn, blöð taka nú að berast að utan hvert á fætur öðru og ég fletti og hræri í bunkunum tveim, skila stundum blöðum til baka, stundum ekki, en allt og sumt sem ég geri er að fylgja reglubókinni góðu.
>
> Nú háttar reyndar svo til, þótt það sé mér með öllu ókunnugt, að mennirnir utan herbergisins, sem hafa afhent mér blöðin og sjá um að taka við blöðum frá mér, hafa komið sér saman um að kalla fyrsta

2 Þessu hefur svo sem oft verið haldið fram. Nýlegt dæmi er t.d. í bók Bodens (1990) um „reiknifræði sköpunargáfunnar" þar sem hún segir: „[A] word in a language one understands is a mini-program". Ég hef gagnrýnt kenningar Bodens um sköpunargáfuna á öðrum vettvangi (Jörgen Pind 1994).

blaðabunkann sem ég fékk í hendur „handrit", annan bunkann „sögu", blöðin sem berast mér inn um rifuna nefna þeir „spurningar" og blöðin sem ég sendi þeim til baka kalla þeir „svör". Og leitum við aðstoðar einhvers sem er læs á kínversku gæti hann sagt okkur að úr „sögublöðunum" mætti lesa „Jón fór á hamborgarastað. Hamborgarinn hans var viðbrenndur svo hann yfirgaf staðinn án þess að borga", en úr fyrstu spurningunni „Borðaði Jón hamborgarann?" og úr fyrsta svarinu, „Nei".

Hér hefur Searle haft endaskipti á sambandi manns og tölvu í gervigreindarrannsóknum, hér er það maðurinn sem leikur tölvu, sem leikur mann. Hvað segir þessi saga okkur? Jú, segir Searle hún leiðir í ljós að tölva sem fylgir forskrift getur aldrei öðlast skilning í bókstaflegri merkingu þess orðs. „Mér tókst að leika hlutverk tölvunnar óaðfinnanlega en það er dagljóst að ég skildi ekki hvað fram fór. Ég gat að vísu meðhöndlað táknin rétt, fylgt forskriftinni og leikið hlutverk tölvunnar af stakri prýði. En ég *skildi* ekki kínverskuna, hvorki spurningarnar né svörin. Fyrir mér var þetta merkingarlaus meðhöndlun tákna." Skilningur felur í sér það sem Searle nefnir *íbyggni*, við skiljum mál af því að við vitum að orð og setningar í málinu eru *um* eitthvað, ekki bara formleg tákn. En geta vélar þá ekki hugsað? Öðru nær segir Searle, aðeins vélar geta hugsað, en nánar tiltekið einvörðungu þær vélar sem búa yfir íbyggni, eru gerðar úr taugafrumum og öðru því sem myndar heila manna og dýra. Hugsanlega geti jafnvel tölvur hugsað en *ekki* ef allt og sumt sem þær gera er að fylgja formlegri forskrift. Slíkar vélar geta ekki hugsað því þær skortir íbyggni.

Víkur þá sögunni aftur að þeim félögum Newell, Shaw og Simon. Eins og fyrr segir þurftu Newell og Shaw að glíma við margvísleg tæknileg vandamál við forritun *Rökfræðingsins*, ekki síst við þróun svonefndrar listaforritunar. Af þessum sökum drógst að forritið yrði nothæft. Simon og Newell létu það ekki aftra sér frá því að prófa forritið, og gerðu það með þeim skemmtilega hætti sem Simon lýsir í sjálfsævisögu sinni:

> Meðan við biðum eftir að forritun *Rökfræðingsins* lyki skrifuðum við Al [Newell] reglur einstakra forritshluta (undirforrita) á spjöld á ensku. Einnig útbjuggum við spjöld er greindu frá því sem varðveitt var í einstökum minniseiningum (kennisetningar rökfræðinnar). Eitt myrkt vetrarkvöld í janúar 1956 kölluðum við saman eiginkonu

mína og þrjú börn og nokkra stúdenta í framhaldsnámi í háskólabyggingunni. Hver viðstaddra fékk í hendur eitt af spjöldunum, þannig að hver þeirra varð í reynd hluti Rökfræðingsforritsins - undirforrit sem framkvæmdi sérstaka aðgerð, eða varð hluti af minni þess. Hver þátttakenda átti að framkvæma eigið undirforrit eða birta það sem í minni hans var skráð í hvert sinn sem kallað var á það af forritshluta sem var á næsta þrepi fyrir ofan og stjórnaði vinnslunni í það skiptið.

Þannig gátum við hermt eftir hegðun Rökfræðingsins með tölvu sem sett var saman úr mannfólki. Hér hermdi náttúran eftir listinni sem hermdi eftir náttúrunni. Þátttakendurnir báru engu meiri ábyrgð á eigin athöfnum en þrællinn ungi í *Menóni* Platóns en þeim tókst að sanna þær kennisetningar sem þeir fengu til meðferðar. Börnin okkar voru þá níu, ellefu og þrettán ára og minnast þessa atburðar enn greinilega (Simon 1991, bls. 206-207).

Þessi frásögn er allrar athygli verð því hér beittu þeir þremenningar nákvæmlega sömu aðferð og Searle lýsti rúmum tveim áratugum síðar í grein sinni. En ólíkt Kínaherbergi Searles var hér ekki um hreina hugarsmíð að ræða heldur raunverulega prófun forrits með því að láta fólk leika hlutverk einstakra undirforrita Rökfræðingsins. Newell og Simon drógu hins vegar allt aðrar ályktanir af tilraun sinni en Searle. Athygli þeirra beindist að *kerfinu í heild* en ekki að einstaklingunum sem þátt tóku í tilrauninni. Kerfið í heild leiðir út sannanir *Principiu*, ekki börn Simons eða aðrir sem þarna komu við sögu. Tilraunin er einnig að því leyti athyglisverð að hún sýnir einkar greinilega að tölvan sem slík er ekki gædd neinum sérstökum eiginleikum sem eru nauðsynlegir til að geta leyst verkefni af þessu tagi - það eina sem skiptir máli er að forskriftinni sé rétt fylgt, engu skiptir hvort börn eða rafrásir eiga hlut að máli.

Horfið frá efnislegum táknkerfum

Áberandi hefur orðið innan gervigreindar (og jafnframt í hugfræði) á síðustu árum að menn hafa í auknum mæli horfið frá kenningasmíði sem byggir á efnislegum táknkerfum eins og þeim sem hér hefur verið lýst. Áherslan hefur beinst frá samlíkingu hugar og hugbúnaðar en fræðimenn þess í stað beint sjónum sínum að samlíkingu heila og tölva (þó ekki í anda von Neumanns). Fyrir þessu eru margvíslegar ástæður. Ein er vafalítið sú að hin efnislegu táknkerfi hafa ætíð átt sér

óljósa tilvist í lífheimi. Allan Newell var ófeiminn við að viðurkenna þetta og lætur t.d. að því liggja á einum stað að maðurinn sé jafnvel eina lífveran sem sé viti borin:

> Fyrirbæri hugans hafa sprottið úr flóknum ferlum efnisheimsins og er með sláandi hætti að finna í okkur mannfólki og þó hugsanlega víðar (Newell 1980, bls. 138).

Í einni af síðustu greinum sínum er hann langtum afdráttarlausari og dregur skörp skil manna og dýra, svipað reyndar og Descartes (1637/1991) mörgum öldum fyrr:

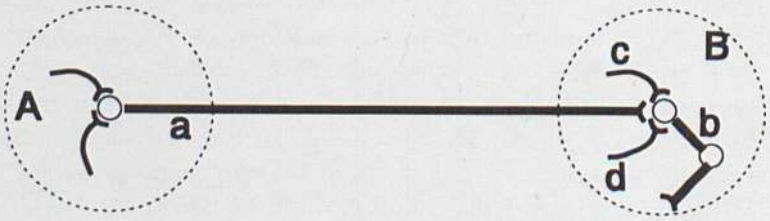
> Greind einkennir þann hátt sem menn hafa við að leysa úr verkefnum [...] Andstæða þess eru önnur kerfi - dýr, dúkkuhöfuð eða hvaðeina - sem búa alls ekki yfir þessum eiginleika og geta því ekki sýnt vitræna hegðun (*intelligent behavior*) (Newell, Young og Polk, 1993, bls. 39).

Jafnframt hefur Newell ætíð haldið því fram að gæða megi tölvur sams konar viti og prýðir mannfólkið. Því er ekki að neita að þessi kenning hefur farið fyrir brjóstið á mörgum sem vonlegt er. Á allra síðustu árum hefur áhrifa Darwinisma tekið að gæta á ný innan sálfræðinnar, þ.á m. í hugfræði (Barkow, Cosmides og Tooby 1992; Pinker 1994), og að sama skapi hefur áhugi á „líffræðilega trúverðugum" kenningum aukist. Þetta er vafalítið ein ástæða þess að *nethyggja* hefur á ný haldið innreið sína í gervigreind og sálfræði (Brooks 1991; Quinlan 1991) og efnisleg táknkerfi hafa látið undan síga nema þá helst í rannsóknum á „æðri hugarstarfsemi", viðfangsefnum eins og rökhugsun (Johnson-Laird, 1993).

Nethyggjan á sér reyndar margvíslegar rætur, bæði nýjar og gamlar. Að sumu leyti er nethyggjan angi af meiði *tengslahyggjunnar* sem hefur skipt sköpum fyrir sálfræði (James 1890) og hugarheimspeki undanfarinna alda. Þó eru öllu athyglisverðari þær rætur sem nær liggja, t.d í verkum kanadíska sálfræðingsins Donald O. Hebbs (1949) um nám í tauganetum þar sem reynt er að byggja sálfræðilega kenningasmíði á taugafræðilegum grunni. Hebb setti sér það markmið að skýra nám og skynjun með eiginleikum taugafrumna, nánar tiltekið með kenningu um það hvernig samspili taugafrumna væri háttað.

Nám og skynjun eru meðal elstu viðfangsefna sálfræðinnar og í augum raunhyggjumanna eins og Hebbs eru þessi viðfangsefni nátengd því skynjunin er afrakstur náms. Öll þau „lögmál skynjunar" sem rannsóknir sálfræðinga höfðu leitt í ljós, þ.á m. rannsóknir Gestalt-sálfræðinganna (Koffka 1935), voru talin afrakstur náms. Af þessum sökum var eðlilegt að þessu tvö fyrirbæri væru nátengd í rannsóknum og kenningasmíði.

Í kenningasmíði sinni einblíndi Hebb einkum á hlutverk taugamóta. Hebb gerði ráð fyrir því að endurtekin virkni í taugamótum (sem aftur mátti rekja til taugaboða um frumuna) *breytti* þröskuldi taugamótanna þannig að þau yrðu framvegis næmari fyrir áreitun. Af þessu leiðir auknar líkur á því að nærliggjandi frumur sýni sambærilega virkni, þ.e. fylgni í virkni þeirra eykst. Hebb gerði ráð fyrir því að slík endurtekin, samstillt virkni í hópi taugafrumna tengdi þær saman þannig að þær væru líklegar til að svara áreitum sem ein heild. Slíkar heildir nefndi Hebb „frumuklasa" (*cell assemblies*). Hebb taldi að leita mætti skýringa á lögmálum skynjunar og eðli náms í tilvist slíkra frumuklasa (mynd 1).



Mynd 1: Kenning Hebbs um tengslamyndun í taugakerfinu. A og B eru mengi taugafrumna (aðeins fáar frumur eru hér sýndar). Taugaboð berst frá A til B eftir a. Það veldur því að taugamengin A og B verða virk á sama tíma. Endurtekin áreitun taugasímans a leiðir til þess að þröskuldur fyrir taugaboð lækkar og því þarf stöðugt minni áreitun til að kalla fram virkni í B.

Ein megintakmörkun á kenningu Hebbs var sú að hann gerði aðeins ráð fyrir örvandi taugaboðum, ekki hamlandi og kemur kenning hans ekki nema að nokkru leyti heim við niðurstöður rannsókna sem hafa leitt í ljós bæði hamlandi og örvandi taugaboð. Peter Milner (1957) endurbætti kenningu Hebbs með því að gera ráð fyrir hamlandi boðum við hlið hinna örvandi. Nýrri rannsóknir hafa einnig sýnt að breytingar á virkni nærliggjandi taugafrumna virðast til muna flóknari en Hebb gerði ráð fyrir (Churchland og Sejnowski 1992, bls. 253-254).

Samantekt

Í þessari grein hefur verið vikið að afar mikilvægum þætti þeirra Newells og Simons í mótun gervigreindar á 6. tug þessarar aldar og einkum að þeirri hugmynd að eðlilegt sé að bera saman mannlega hugarstarfsemi og hugbúnað tölva þar sem hvoru tveggja feli í sér *efnisleg táknkerfi*. Sú kenning á ekki sama fylgi að fagna og fyrrum og greina má afturhvarf til eldri hugmynda í þeirri *nethyggju* sem nú nýtur vaxandi vinsælda. Nethyggjan tekur mið af starfshætti taugafrumna, eða öllu heldur frumuklasa svipuðum þeim sem Hebb varð einna fyrstur til að lýsa. Þessi viðhorfsbreyting innan sálfræðinnar er til marks um aukið vægi líffræðilegra sjónarmiða í nútímasálfræði sem að hluta skýrist af þeim gríðarlegu framförum sem orðið hafa í taugalíffeðlisfræði og þróunarlíffræði á síðustu árum. Nútíma nethyggja og gamaldags gervigreind eiga sér þó eitt sameiginlegt: Kenningasmíð snýst að mestu um gerð tölvulíkana. Áhrifa tölvubyltingarinnar á vísindalega sálfræði mun því vafalítið gæta áfram um ókomin ár. Og menn munu spyrja, eins og bandarískí málfræðingurinn Ray Jackendoff (1994) gerir í nýlegri bók: „Að hvaða leyti er maðurinn líkur öðrum dýrum en frábrugðinn tölvum?“

Heimildir

Barkow, J. H., L. Cosmides og J. Tooby. 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

Boden, M. A. 1990. *The Creative Mind*. New York: Basic Books.

Brooks, R. A. 1991. *Intelligence without Reason*. A. I. Memo 1293, Massachusetts Institute of Technology: Artificial Intelligence Laboratory, Cambridge, Massachusetts.

Churchland, P. S. og T. J. Sejnowski 1992. *The Computational Brain*. Cambridge, Massachusetts: MIT Press.

Crevier, D. 1993. *AI: The Tumultuous History of the Search for Artifical Intelligence*. New York: Basic Books.

Descartes, R. 1637/1991. *Orðræða um aðferð*. Lærdómsrit Bókmenntafélagsins. Reykjavík: Hið íslenska bókmenntafélag.

Gardner, H. 1987. *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea*. Cambridge, Massachusetts: MIT Press.

Hebb, D. O. 1949. *The Organization of Behavior: A Neuropsychological Theory*. New York: John Wiley.

Jackendoff, R. 1994. *Patterns in the Mind: Language and Human Nature*. New York: Basic Books.

James, W. 1890. *The Principles of Psychology, I-II*. New York: Henry Holt & Co.

Johnson-Laird, P. 1993. *Human and Machine Thinking*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Jörgen Pind. 1994. „Computational Creativity: What Place for Literature?" *Behavioral and Brain Sciences* **17**, 547-548

Koffka, K. 1935. *Principles of Gestalt Psychology*. London: Routledge & Kegan Paul.

McCorduck, P. 1979. *Machines Who Think*. San Francisco: W. H. Freeman.

Miller, G. A., E. Galanter og K. H. Pribram. 1960. *Plans and the Structure of Behavior*. New York: Henry Holt.

Milner, P. M. 1957. „The Cell Assembly: Mark II." *Psychological Review* **64**, 242-252.

Neisser, U. 1967. *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Newell, A. 1980. „Physical Symbol Systems". *Cognitive Science* **4**, 135-183.

Newell, A., J. C. Shaw og H. A. Simon. 1958. „Elements of a Theory of Problem Solving". *Psychological Review* **65**, 151-166.

Newell, A. og H. A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, N.J.: Prentice-Hall.

Newell, A. og H. A. Simon. 1976. „Computer Science as Empirical inquiry: Symbols and Search". *Communications of the ACM*, **19**(3), 113-126. Endurprentuð í Boden (1990, 105-132).

Newell, A., R. Young og T. Polk. 1993. „The Approach Through Symbols". Í D. Broadbent (ritstj.), *The Simulation of Human Intelligence*, bls. 33-70. Oxford: Blackwell.

Norman, D. A. 1993. *Things That Make us Smart: Defending Human Attributes in the Age of the Machine*. Reading, Massachusetts: Addison-Wesley.

Pinker, S. 1994. *The Language Instinct: How the Mind Creates Language*. New York: William Morrow.

Quinlan, P. 1991. *Connectionism and Psychology: A Psychological Perspective on New Connectionist Research*. Chicago: The University of Chicago Press.

Schank, R. 1975. *Conceptual Information Processing*. New York: Elsevier.

Schwartz, J. T. 1988. „The New Connectionism: Developing Relationships Between Neuroscience and Artificial Intelligence". *Daedalus*, **117**(1), 123-141.

Searle, J. 1980. „Minds, Brains, and Programs". *Behavioral and Brain Sciences* **3**, 417-458.

Selfridge, O. G. 1958. „Pandemonium: A Paradigm for Learning". Í *Mechanization of Thought Processes*, bls. 513-526. London: HMSO.

Simon, H. A. 1991. *Models of my Life*. New York: Basic Books.

Turing, A. M. 1950. „Computing Machinery and Intelligence". *Mind*, **LIX**, 433-460.

von Neumann, J. 1958. *The Computer and the Brain*. New Haven: Yale University Press.