Atli Harðarson

VÉLMENNI[1]

1. KAFLI: KENNING ALAN TURING

Árið 1950 birtist grein eftir Alan Turing í enska heimspekitímaritinu *Mind*. Greinin heitir "Computing Machinery and Intelligence". Það mætti kalla hana "Reikniverk og vitsmuni" á íslensku. Í þessari grein veltir Turing því fyrir sér hvort hægt sé að forrita tölvu þannig að hún fái mannsvit. Hann ræðir ýmis hugsanleg rök gegn þessari kenningu, hafnar þeim öllum og stingur upp á aðferð til að skera úr um hvort tölva geti hugsað eins og maður. Aðferðin er fólgin í því að láta vélina gangast undir próf, þar sem hún er lokuð inni í einu herbergi og maður inni í öðru. Prófdómarar skrifast svo á við manninn og tölvuna. Þeir mega fitja upp á hvaða umræðuefni sem er. Takist þeim ævinlega að finna út hvort er maðurinn þá hefur tölvan fallið á prófinu. Takist prófdómurunum þetta ekki þá hefur tölvan staðist prófið og þá er, að áliti Turing, engin ástæða til að ætla henni minna vit, eða minni andlega hæfileika, en manninum. Svona próf er kallað Turingpróf. Þótt til séu forrit sem halda uppi eðlilegum samræðum í smástund, og standast prófið ef prófdómurunum er gefinn nógu stuttur tími, vantar enn mikið á að tekist hafi að forrita tölvur þannig að þær geti spjallað tímunum saman um alla heima og geima eins og menn gera.

Kenningu Alans Turing í greininni um reiknivélar og vitsmuni má hluta sundur í tvennt:

a) Annars vegar taldi hann mögulegt að forrita tölvu þannig að hún standist Turingpróf.
b) Hins vegar áleit hann að tölva sem stenst Turingpróf skilji mannamál og hugsi í raun og veru eins og maður. Hugsun hennar og skilningur séu raunveruleg, ekki bara eftirlíking.

---

[1] Þessi ritgerð hefur verið nokkuð lengi í smíðum. Kristján Kristjánsson heimspekingur las eitt uppkast hennar og lagði til góð ráð. Þann 2. október 1994 las ég hluta hennar á fundi hjá Félagi áhugamanna um heimspeki og fékk nokkrar gagnlegar athugasemdir og ábendingar utan úr sal. Eftir það lásu Ólafur Páll Jónsson heimspekingur, Jörgen Pind sálfræðingur og Skúli Sigurðsson sagnfræðingur textann vandlega og bentu mér á ýmislegt sem betur mátti fara.

Sé þessi kenning rétt þá er fræðilega mögulegt að einhvern tíma takist að smíða vélmenni sem hafa svipaða vitsmuni og fólk af holdi og blóði.

Þótt hugmyndir upphafsmanna tölvufræðinnar hafi haft mikil áhrif innan mannvísindanna er langt frá því að sú tvíþætta kenning sem ég hef hér eignað Alan Turing sé almennt viðurkennd. Margir efast um að nokkru sinni takist að forrita tölvu þannig að hún standist Turingpróf, enda er þetta próf ansi erfitt. Til að átta okkur á hversu erfitt það er skulum við ímynda okkur að blindur maður gangist undir Turingpróf og falli ef prófdómarar geta þekkt hann frá manni með fulla sjón. Þeir mega spyrja alls konar spurninga um liti og lögun hluta, biðja hann að lýsa skýjunum á himninum og ræða um auglýsingar í sjónvarpi, kvikmyndir eða nýjustu strauma og stefnur í myndlist og sá blindi verður að standa sig jafn vel og hinn sem sér. Ef ykkur finnst þetta vonlaus leikur fyrir þann blinda hversu miklu vonlausari er hann þá ekki fyrir vél sem hefur svipaða vitsmuni eða svipaða greind og maður en allt öðru vísi skynfæri, allt öðru vísi möguleika á að hreyfa sig, mynda sambönd við aðra o.s.fr.? Til að standast prófið þarf vélin að vera miklu klárari en maðurinn.

Hér ætla ég að láta það liggja á milli hluta hvort hægt sé að forrita tölvu þannig að hún standist Turingpróf. Viðfangsefni mitt hér er sá hluti kenningarinnar sem segir að tölva sem hægt er að tala við um alla heima og geima sé, eða geti að minnsta kosti verið, vitsmunavera í þess orðs fyllstu merkingu, gædd raunverulegri hugsun og raunverulegum skilningi. Ég ætla semsagt að einbeita mér að þeim hluta kenningar Turing sem nefndur er í b lið hér að framan.

*

Á árunum eftir seinni heimstyrjöldina, þegar Alan Turing, John von Neumann og fleiri unnu að hönnun og smíði fyrstu tölvanna, þá lágu kenningar í þessum dúr í loftinu. Árið 1950 voru að vísu ekki til nema örfáar tölvur í veröldinni. En grundvallarhugmyndir tölvufræðinnar voru komnar fram. Þær höfðu strax áhrif í mörgum greinum mannvísinda og áttu meðal annars sinn þátt í því að hugfræði og bollaleggingar um innviði sálarlífsins tóku við af atferlishyggju í sálfræðinni og hugmyndir um formleg mál og rekjanlegar aðferðir urðu grundvöllur nýrra kenninga í málvísindum sem Noam Chomsky átti mestan þátt í að móta. Nokkrum árum áður (1943) höfðu Walter Pitts og Warren McCulloch bent á að net úr taugafrumum, eins og heili okkar er gerður úr, samsvarar á ýmsan hátt rökrásum af því tagi sem síðar voru notaðar til að smíða tölvur.[2] Í heimspeki var tvíhyggja í anda Descartes á undanhaldi og hughyggja 19. aldarinnar dauð að kalla. Rökfræðileg raunhyggja (logical positivism) stóð líka höllum fæti, en sú stefna hafði dæmt flestar frumspekilegar vangaveltur um samband sálar og líkama merkingarlausar.

Á sama tíma og sálfræðingar voru að verða leiðir á atferlisstefnunni og leituðu leiða til að kanna innviði mannshugans voru heimspekingar í leit að nýjum lausnum á vandamálinu um samband sálar og líkama[3] og líffræðingar að gæla við

---

[2] Þeir gera grein fyrir kenningum sínum í McCulloch & Pitts 1965. Um kenningar Pitts og McCulloch og fleiri frumkvöðla í stýri-, upplýsinga- og tölvufræðum má líka lesa í Singh 1966.

[3] Á árunum um og fyrir 1960 settu nokkrir heimspekingar fram kenningar í þá veru að samband hugar og heila sé sambærilegt við samband forrits og tölvu. Frægastur þessara heimspekinga er Hilary Putnam. Sjá Putnam 1960, 1967a og 1967b. Kenning Putnam um þetta efni er kölluð "functionalism".

samanburð á taugum og rökrásum. Spekingar í öllum þessum greinum litu til tölvutækninnar, stýrifræðinnar og rökfræðinnar og sóttu þangað hugtök og hugmyndir.[4]

Ritgerð Turing svaraði svo sannarlega kalli tímans. Það vantaði tilgátu um hvernig mannshugurinn virkar. Turing stakk upp á að líkja honum við tölvu og þegar um miðjan 6. áratuginn voru snjallir forritarar, eins og Allen Newell og Herbert A. Simon, farnir að gera tilraunir til að láta tölvur herma eftir mannlegum vitsmunum. Til varð ný fræðigrein, gervigreindarfræðin. En hún fjallar um aðferðir til að láta tölvur vinna verk sem menn vinna með því að beita hugsun, skilningi eða vitsmunum af einhverju tagi.[5]

Samhliða gervigreindarfræðum í anda Newell og Simon þróuðust rannsóknir á tauganetum sem héldu áfram á þeirri braut sem Pitts og McCulloch höfðu markað. Þótt þessi fræði ættu ýmislegt sameiginlegt byggðu þau um margt á ólíkum hugmyndum. Rannsóknir innan gervigreindarfræði snerust um að finna reglurnar sem mannlegt hugarstarf og hegðun fylgja og forrita tölvur til að fylgja sömu reglum. Rannsóknir af þessu tagi gera yfirleitt ráð fyrir að hugarstarf manna og hegðun fylgi reglum sem hægt er að koma orðum að og þessar reglur séu skráðar í heilabú þeirra á einhvers konar táknmáli líkt og forrit er skráð í minni tölvu.

Tilraunir til að láta tauganet herma eftir mannlegu hugarstarfi gera hins vegar ekki ráð fyrir því að reglur þess séu skráðar á táknmáli í kollinum á okkur neitt frekar en hagfræðin gerir ráð fyrir að lögmál efnahagslífsins séu skráð á markaðstorgum. Þessi lögmál eru ekki forskrift sem hagkerfið fylgir heldur tölfræðileg niðurstaða af hegðun margra einstaklinga.

Aðferðafræði gervigreindarfræðanna hefur lengst af verið í dúr við ríkjandi vinnubrögð við hugbúnaðargerð. Unnið er ofan frá og niður þannig að fyrst er skilgreint nákvæmlega hvað forrit á að gera og síðan eru smíðuð undirforrit sem annast einstaka verkþætti. Verkþættir undirforritanna eru svo samsettir úr undirforritum enn neðar í stigveldinu og þannig koll af kolli uns komið er niður í einfaldar aðgerðir sem eru byggðar inn í forritunarmálið sem vélin vinnur eftir. En þótt þess sé gætt að skilgreina verkið þannig að forritið skili sams konar niðurstöðum og það hugarstarf sem hermt er eftir þá er ekkert hirt um hvort hinar einföldu aðgerðir neðar í stigveldinu samsvara nokkru því sem gerist í mannshuganum. Segja má með nokkurri einföldun að rannsóknir á tauganetum hafi lengst af byggt á öndverðri aðferðafræði þar sem unnið er neðan frá og upp. Reynt er að herma eftir einföldustu þáttunum í starfi miðtaugakerfisins, nefnilega samskiptum einstakra taugafruma, og byggja á þeim flóknari ferli sem líkjast hugarstarfi lifandi vera.

Ýmis afbrigði hennar hafa verið vinsælt umfjöllunarefni þeirra sem fást við heimspeki hugans (philosophy of mind). Churchland 1988 er ágætur inngangur að þeim fræðum.

[4] Um þessa tölvubyltingu í mannvísindunum hefur margt verið ritað en það er óhætt að mæla sérstaklega með Gardner 1985 og Bolter 1984. Einnig er gerð grein fyrir áhrifum tölvufræða á sálarfræði í grein Jörgens Pind sem birtist í þessu tölublaði Hugar.

[5] Um gervigreindarfræði hefur margt verið ritað. Saga þeirra er vel og skilmerkilega sögð í Crevier, Daniel 1993. Stuttan og aðgengilegan inngang má finna í 6. kafla Churchland 1988. Minsky 1985 er skemmtileg tilraun til að lýsa mannshuganum frá sjónarhóli tölvufræðanna og tengja saman sálfræði og gervigreindarfræði.

Því miður hefur sáralítið verið skrifað um gervigreindarfræði á Íslensku. Þó má nefna greinar Jóns Torfa Jónssonar frá 1985 og 1992 og grein Þorsteins Gylfasonar frá 1985.

Á árunum upp úr 1950 fóru rannsóknir af þessum tveim gerðum af stað. Næstu 20 árin blómstruðu gervigreindarfræði í anda Newell og Simon en æ færri sinntu tauganetum. Síðan hafa hefðbundin gervigreindarfræði átt á brattan að sækja og áhugi á tauganetum aukist á ný. Nú eru þessar tvær rannsóknarhefðir að renna saman í eina.[6]

2. KAFLI: REIKNANLEG FÖLL, ALGÓRIÞMAR OG TÁKN

Árið 1936, 14 árum áður en hann skrifaði greinina um reiknivélar og vitsmuni, setti Turing fram skilgreiningu á reiknanleika eða reiknanlegu falli. Hann hugsaði sér vélar sem geta rennt í gegn um sig strimli, skrifað merki á hann, strokað þau út og lesið og breytt ástandi sínu eða stillingum á fáeina einfalda vegu eftir því hvaða merki þær nema á strimlinum. Ég ætla ekki að lýsa þessum vélum nánar. Þær eru kallaðar Turingvélar. Hver Turingvél getur unnið eftir einni aðferð.[7]

Það er hægt að lýsa þessum vélum og hegðun þeirra með stærðfræðilegum hætti og orða *aðferð til að herma eftir þeim ef slík lýsing er gefin*. Turingvél sem vinnur eftir þessari aðferð getur því hermt eftir öllum öðrum Turingvélum. Slík vél er kölluð altæk Turingvél (universal Turing machine). Turing gat sér þess til að altæk Turingvél geti reiknað öll reiknanleg föll. Með þessari tilgátu setti hann fram skilgreiningu á reiknanleika. Samkvæmt henni er reiknanlegt fall það sama og fall sem Turingvél getur reiknað.

Turing gat ekki sannað að vélarnar sem hann smíðaði í huga sér gætu reiknað allt sem stærðfræðingar hafa fundið upp eða munu finna upp aðferðir til að reikna. En hann gat sér þess til. Um svipað leyti settu Emil Post og Alonso Church, hvor í sínu lagi, fram skilgreiningar á reiknanleika sem síðar var sannað að væru jafngildar skilgreiningu Turing. Bæði af því að þessar skilgreiningar eru allar jafngildar og af því að engum hefur tekist að benda á fall sem menn gætu hugsanlega reiknað en fellur ekki undir þær má telja nær fullvíst að hér sé um að ræða réttar skilgreiningar á

---

6 Lýsingu á þessum tvenns konar rannsóknaraðferðum má finna í grein Jóns Torfa Jónassonar frá 1992. Heimspekilega greinargerð fyrir muninum á hefðbundnum gervigreindarfræðum og rannsóknum á tauganetum má finna í Clark 1990 og Dreyfus & Dreyfus 1988.

Með venjulegri tölvu er hægt að herma eftir tauganeti með því að forrita hermilíkan af því og sömuleiðis er fræðilega mögulegt að láta tauganet líkja eftir öllu því sem tölvur geta gert. En sum verk er hægt að vinna margfalt hraðar með dreifðri vinnslu eins og fram fer í tauganeti heldur en með einu miðverki eins og í venjulegri tölvu og aðferðirnar við að forrita tauganet kunna að opna möguleika á að láta vélar vinna verk sem er tæpast vinnandi vegur að forrita með hefðbundnum aðferðum.

Í Dennett 1991 setur heimspekingurinn Daniel Dennett fram þá tilgátu að mannshugurinn sé tauganet en meðvitundin og ýmislegt æðra hugarstarf sé til vegna þess að hluti af starfsemi þessa tauganets sé í því fólginn að herma eftir miðverki hliðstæðu því sem er í venjulegri tölvu.

7 Nánari lýsingu á Turingvélum og greinargerð fyrir kenningu Turing um reiknanleika má finna í Davis 1982 bls. 3 - 24. Umfjöllun Davis er nokkuð tæknileg. Styttri og heldur alþýðlegri útlistanir má finna í Cutland 1980 bls. 53 - 57, Singh 1966 bls. 184 - 204, Penrose 1989 bls. 35 - 57 og Kleene 1988. Um Turingvélar og kenningu Turing um reiknanleika er fjallað frá ýmsum hliðum í ritgerðasafni Herken frá 1988.

reiknanleika. Það má þá líka telja nær fullvíst að vélarnar sem Turing hugsaði sér, og vani er að kalla Turingvélar, geti reiknað allt sem hægt er að reikna.

Sú tilgáta að skilgreiningar þeirra Church, Post og Turing fangi eðli reiknanleikans, þannig að undir þær falli öll reiknanleg föll, er kölluð tilgáta Church og Turing. Ég geng að því sem vísu að hún sé sönn.

Tölvur geta gert allt sem altæk Turingvél getur svo það er hægt að láta tölvur framkvæma alla mögulega útreikninga.[8] Hvernig tengist þetta mannlegum vitsmunum? Vél sem getur unnið eftir hvaða reikniaðferð sem er getur ef til vill slegið fólki við í reikningi - en fólk getur gert svo ótal margt annað en að reikna. Er nokkur ástæða til að ætla að hægt sé að fá tölvur til að herma eftir annars konar hugarstarfi en því sem byggist á einhvers konar útreikningum?

Hér þurfum við að huga að því hvað átt er við með tali um reiknanleg föll. Fall tekur við táknum, þ.e. tölum, stöfum, orðum, merkjum eða munstrum af einhverju tagi og skilar táknum þannig að fyrir hvert mögulegt inntak er eitt og aðeins eitt úttak. Sem dæmi um fall má nefna annað veldi. Ef við setjum töluna 2 inn þá kemur talan 4 út og ef við setjum 3 inn þá kemur 9 út o.s.fr. Annað dæmi er fall sem tekur við orði og skilar orðinu "rétt" ef það er rétt stafsett íslenskt orð og orðinu "rangt" ef svo er ekki.

Að fall sé reiknanlegt þýðir að hægt sé orða endanlega og nákvæma aðferð til að finna hvaða úttaki það skilar fyrir hvert mögulegt inntak. Föllin sem ég tók sem dæmi eru bæði reiknanleg. En ekki eru öll föll reiknanleg. Fallið sem tekur við setningu og skilar orðinu "satt" ef hún er sönn og orðinu "ósatt" ef hún er ósönn, er til dæmis ekki reiknanlegt, því það er ekki til nein endanleg og nákvæm aðferð til að komast að sannleikanum í öllum málum.

Endanlegar og nákvæmar aðferðir til að vinna með tákn eru stundum kallaðar algóriþmar eða algrím. Fyrir hvert reiknanlegt fall eru til margir algóriþmar. Það eru til dæmis til margar aðferðir til að reikna annað veldi. Hverjum algóriþma samsvarar hins vegar bara eitt reiknanlegt fall. Vél sem getur reiknað öll reiknanleg föll getur unnið öll verk sem hægt er að vinna eftir algóriþma.

Það ætti nú að vera ljóst að vél sem getur reiknað öll reiknanleg föll getur gert ótal margt annað en að vinna venjulegan talnareikning. Hún getur unnið öll verk sem hægt er að vinna með því að beita endanlegum og nákvæmum aðferðum á tákn. Þar sem tölva getur reiknað öll reiknanleg föll getur hún til dæmis teflt skák, því skák er í því fólgin að möndla með tákn eftir aðferð eða reglu. Að vísu getur engin tölva teflt hina fullkomnu skák því aðferðin til þess er svo löng að það tæki jafnvel hraðvirkustu tölvu meira en milljón ár að ákveða hvern leik. Þegar sagt er að tölva geti reiknað öll reiknanleg föll þá verður að hafa þann fyrirvara á að sum föll er svo seinlegt að reikna að engin vél mundi endast til þess.

Það að tölvur geti möndlað með tákn á alla mögulega vegu þýðir ekki að þær geti gert hvað sem er. Það er til dæmis ekki víst að þær geti leikið körfubolta, því körfubolti er ekki í því fólginn að vinna með tákn. Hér er mikill munur á körfubolta og skák. Tölva sem keyrir skákforrit teflir raunverulega skák. Það eru til tölvuleikir

---

[8] Hér þarf að bæta við einum fyrirvara. Turing hugsaði sér að vélarnar sínar hefðu óendanlega langan tíma og óendanlega mikið pláss fyrir útreikninga á pappírsstrimlinum. Engin tölva endist endalaust og engin tölva hefur óendanlega stórt vinnsluminni. Það er því sama hvaða tölvu er bent á, ef við vitum endingartíma hennar og vinnslurými þá getum við fundið reiknisdæmi sem hún ræður ekki við.

sem herma eftir körfubolta en þeir láta tölvu ekki leika raunverulegan körfubolta. Skákmennirnir og reitirnir á skákborðinu eru tákn en körfuboltinn, körfuboltahringirnir og leikvöllurinn ekki.

Skákmennirnir eru tákn því það skiptir engu máli úr hvaða efni þeir eru, hvað þeir eru stórir eða hvernig þeir líta út. Það dugar að þeir þekkist í sundur og það sé ljóst eftir hvaða reglum þeir hreyfast. Ef einhverjir vilja nota brauðsneiðar fyrir reiti á skákborði, tómata fyrir peð og gulrætur fyrir biskupa þá er það allt í lagi. Fylgi þeir aðeins reglunum þá er leikurinn sem þeir leika fullgild alvöruskák. Það er líka fullgild alvöruskák þótt myndir á tölvuskjá komi í staðinn fyrir taflmenn úr tré. En mynd á tölvuskjá getur ekki komið í staðinn fyrir körfubolta. Það er ekki hægt að leika raunverulegan körfubolta með henni. Með þessu er ekki sagt að útilokað sé að smíða vélmenni, sem getur leikið körfubolta en það vélmenni verður að geta gert fleira en að möndla með tákn. Það verður að geta hoppað, hlaupið og gripið bolta.[9]

Það sem hér hefur verið sagt um skákmenn gildir um öll tákn. Það eina sem skiptir máli er að þau þekkist frá öðrum táknum í stafrófinu eða því safni tákna sem er í notkun.

*

Sú saga er sögð af hagfræðingnum Keynes að hann hafi eitt sinn verið spurður hvort hann hugsi fremur í myndum eða í orðum og hann hafi svarað: "Ég hugsa í hugsunum." Spurningin sem var borin fyrir Keynes minnir á að fólki er tamt að líta svo á að það hugsi í einhvers konar merkjum sem geta t.d. verið myndir, orð eða hljóð. Ein útgáfa af þessari hugmynd er að hugsun sé í því fólgin að möndla með tákn. Þessa kenningu ætla ég að kalla formhyggju um hugsun. Formhyggja (eða formalismi) á sér langa sögu[10] og birtist til dæmis í ýmsum myndum í heimspeki 17. aldar meðal annars í ritum Thomasar Hobbes sem taldi að hugsun væri ekki fólgin í neinu öðru en því að færa tákn fram og til baka í huganum.[11] Annar heimspekingur á 17. öld sem setti fram merkilegar kenningar um mannshugann var René Descartes. Hann þóttist hafa uppgötvað að mannshugurinn sé af allt öðru tagi en efnishlutir. Í riti sínu *Orðræðu um aðferð* segir hann:

> Af þessu varð mér vel ljóst, að ég var veruleiki og að allt eðli eða náttúra þessa veruleika var að hugsa og að þessi veruleiki var staðlaus í rúminu.[12]

Ef við bræðum saman þá kenningu Descartes að allt eðli manns sé að hugsa og þá kenningu Hobbes að hugsun sé í því fólgin að möndla með tákn þá fáum við út að allt mannlegt eðli sé fólgið í því að möndla með tákn. Ef þessi kenning er rétt þá er ekki mikill munur á mönnum og tölvum. Heimspekilegar forsendur fyrir kenningum

---

9 Þessi samanburður á körfubolta og skák er að nokkru fengin að láni úr Haugeland 1985.

10 Nokkrir kaflar úr þessari sögu eru sagðir í Pratt 1987.

11 Hobbes fjallar um þetta efni í *Leviathan* I:v, Hobbes 1962 bls. 81 o.áf.

12 Descartes 1991 bls. 99.

um mannshugann svipuðum þeim sem búa að baki gervigreindarfræðunum voru sem sagt að nokkru leyti komnar fram þegar á 17. öld.

Nú á 20. öld hafa verið settar fram ýmsar útgáfur af formhyggju um hugsun. Ein sú frægasta er í bókinni *The Language of Thought* eftir bandaríska heimspekinginn Jerry Fodor. Í þessari bók heldur Fodor því fram að heilinn eigi sér mál og upplýsingarnar sem hann geymir og vinnur með séu skráðar á þessu máli sem og forskriftirnar sem unnið er eftir. Þetta er auðvitað ekki sama mál og menn tala sín á milli og þarf ekki að líkjast því neitt frekar en vélamál tölvu líkist þeim merkjum sem hún notar til að hafa samband við menn eða aðrar vélar.

Ég þarf ekki að leggja neitt mat á þessar kenningar um mannlega hugsun til að fullyrða að ef hugsun er fólgin í því einu að möndla með tákn eftir aðferðum eða algóriþmum þá er hægt að gæða tölvur raunverulegri hugsun og þá verðum við að fallast á seinni partinn af kenningu Turing. En hvað ef hugsun felur í sér eitthvað annað en þetta, er þá útilokað að gæða tölvu raunverulegri hugsun? Nei, eins og ég útskýri í næsta kafla er málið ekki alveg svo einfalt.

3. KAFLI: HERMILÍKÖN

Tölva sem keyrir skákforrit teflir raunverulega skák. Tölva sem keyrir körfuboltaforrit leikur hins vegar ekki raunverulegan körfubolta heldur bara eftirlíkingu af körfubolta. Ástæðan er sú að skák er táknkerfi, þ.e. samsett úr táknum og reglum um uppröðun þeirra og tilfærslu, en körfubolti ekki. En tölva sem stjórnar skurðgröfu grefur raunverulegan skurð og skurður er ekki táknkerfi heldur farvegur fyrir vatn.

Þegar tölva leikur táknkerfi, eins og skák, þá er útkoman raunveruleg: Alvöru skák ef hún teflir; alvöru vísur ef hún setur saman ferskeytlur; alvöru útreikningar ef hún er að vinna með tölur; alvöru leiðréttingar ef hún er að leiðrétta stafsetningarvillur í texta. Þegar tölva hermir eftir einhverju sem er ekki táknkerfi eins og körfubolta eða skurðgröfustjóra þá er útkoman stundum ekta og stundum bara eftirlíking.

Ég er að velta fyrir mér spurningunni um hvort tölva sem stendur sig þokkalega á Turingprófi hugsi og skilji í alvöru eða búi aðeins yfir eftirlíkingu af hugsun og skilningi. Svarið veltur ef til vill á því að hvaða marki tölva getur líkt eftir öðrum fyrirbærum en táknkerfum.

Við getum lýst alls konar hlutum með orðum. Runa af táknum getur lýst býflugum og blómum, fellibyl og fárviðri eða atómum og öreindum. Það er hægt að mata tölvu á svona runu af táknum og láta hana herma eftir því sem lýst er. Til þess dugar þó ekki að lýsa bara ástandi hlutanna. Það þarf líka að lýsa hegðun þeirra, með formúlum, lögmálum eða almennum orðum. Að svo miklu leyti sem hægt er að orða nákvæmar hegðunarreglur má láta tölvu herma eftir hlut með því að breyta lýsingunni á ástandi hans sífellt í samræmi við hegðunarreglurnar.

Forrit sem láta tölvu herma með þessum hætti eftir hlut eða atburðarás kallast hermilíkön. Slík forrit gegna mikilvægu hlutverki í mörgum greinum vísinda. Oft er eina leiðin til að spá um hegðun fyrirbæra sú að keyra hermilíkan sem gengur hraðar en veruleikinn. Hermilíkön gagnast líka oft við að prófa tilgátur og kenningar. Tölva er þá forrituð til að herma eftir þeim reglum sem kenningin gerir ráð fyrir að viðfangsefnið fylgi og síðan er hegðun hermilíkansins borin saman við framvindu

hlutanna. Hermilíkön hafa meðal annars verið notuð töluvert í sálfræði til að prófa kenningar um skynjun, minni, nám o.fl. Fáir efast um að tölvur geti nýst með þessum hætti til að auka skilning manna á eigin sálarlífi.

Sem dæmi um notkun hermilíkans skulum við hugsa okkur að við höfum rannsakað flugu af einhverri tegund og komist að því að þegar hún flýgur um án þess að finna lykt af æti þá hafi hún sterkasta ljósgjafann alltaf hægra megin við sig.[13] Nú getum við forritað tölvu til herma eftir þessari hegðun flugunnar. Lýsingin á flugunni sem forritið vinnur með þarf að tiltaka staðsetningu hennar og stefnu og staðsetningu ljósgjafans. Forritið lætur tölvuna herma eftir flugunni að svo miklu leyti sem hún fylgir þessari hegðunarreglu. Til þess þarf það ekki að gera annað en reikna sífellt ný gildi á breyturnar sem geyma upplýsingar um stefnu hennar og staðsetningu. Þegar forritið er komið í gang getum við borið þær hreyfingar sem það sýnir saman við flug raunverulegra flugna og komist að því hvort kenningin sem forritið byggir á kemur heim við veruleikann.

Við getum svo búið til annað forrit sem hermir eftir vængjaslætti flugunnar og enn annað sem hermir eftir því hvernig ljós og hiti stjórna hraða hennar. Hvert þessara forrita lætur tölvu herma eftir einhverri einni reglu sem lesa má úr hegðun dýrsins. Við getum líka búið til stærra forrit sem hermir í senn eftir siglingatækni, vængjaslætti og hraðastillingum þess. Slíkt forrit ynni með margar breytur og reiknaði sífellt ný gildi á þær. En það er hæpið að hægt sé að búa til eitt forrit sem hermir eftir flugunni að öllu leyti því það má lesa ótal reglur úr hegðun hennar.

Þegar fengist er við táknkerfi skipta fáir eiginleikar máli. Í skák skiptir til dæmis máli um hvern mann hvar hann er, hvort hann er hvítur eða svartur og eftir hvaða reglum hann færist. Aðrir eiginleikar skipta engu máli. Þegar verið er að fást við náttúrufyrirbæri eins og flugu þá er hins vegar hæpið að hægt sé að benda á tæmandi safn eiginleika og segja að þetta séu allir þeir eiginleikar sem máli skipta.

Ef okkur er sagt að tölva hermi eftir flugu þá getum við alltaf spurt: Að hvaða leyti hún hermir eftir flugunni? Rétt svar við þessari spurningu getur aldrei verið: Að öllu leyti. Það er ekki einu sinni víst að hægt sé að láta tölvu herma eftir flugu í öllum aðalatriðum því það er kannski ekki til neinn endanlegur listi af aðalatriðum.

Við getum semsagt látið tölvu herma eftir sérhverri reglu sem við getum lesið úr hegðun flugna, manna eða annarra fyrirbæra. En það er ekki þar með sagt að við getum látið hana herma eftir þeim öllum í senn. Af þeirri setningu að ekki sé til nein regla eða aðferð sem menn geta unnið eftir en ómögulegt er að vél geti hermt eftir leiðir því ekki að hægt sé að smíða vél sem getur hermt eftir öllum reglum sem menn fylgja, eða gert að öllu leyti það sama og mannshugurinn.[14]

Þegar ég segi að hægt sé að láta vél herma eftir sérhverri reglu eða aðferð sem menn geta unnið eftir meina ég það eitt að þetta sé röklega mögulegt. Af þessu má

---

13 Mér skilst að sumar flugur hagi sér svona og fljúgi því nokkurn veginn beint úti í sólskininu en hringsóli kringum næstu ljósaperu ef þær komast inn í hús.

14 Þótt auðvelt sé að rugla þessum tveim setningum saman hafa þær alls ekki sömu merkingu og sú seinni er ekki rökleg afleiðing af þeirri fyrri. Hins vegar er sú fyrri rökleg afleiðing þeirrar seinni með sama hætti og setningin "Allir strákar elska einhverja stelpu" er rökleg afleiðing af setningunni "Til er stelpa sem allir strákar elska."

Eins og ég hef gert grein fyrir rennir tilgáta Church og Turing stoðum undir fyrri setninguna. En hún rennir engum stoðum undir þá síðari nema reglurnar séu endanlega margar.

alls ekki draga þá ályktun að þetta sé í mannlegu valdi. Kannski er sumt af því sem við getum gert svo flókið að aldrei takist að láta vél herma eftir því. Kannski er líka ómögulegt að smíða vél úr kísilflögum sem getur unnið sömu verk og mannshugurinn innan eðlilegra tímamarka. Það er hugsanlegt að öll hermilíkön af tilteknu hugarstarfi sem eru keyrð á vél úr rökrásum, eins og nú eru notaðar við tölvusmíð, séu óþolandi seinvirk.[15]

Hvort sem hugsun er táknkerfi eða ekki er hægt að láta tölvu herma eftir henni að svo miklu leyti sem henni verður lýst með táknum. Ef hún er táknkerfi þá má gera ráð fyrir að eftirlíkingin verði raunveruleg hugsun. Ef hugsun er ekki táknkerfi þá verður eftirlíkingin kannski raunveruleg hugsun og kannski ekki. Hvort hún verður veltur kannski að einhverju leyti á því hvað við erum til í að kalla "raunverulegt" í þessu efni. Við tölum um að flugvélar fljúgi. Við segjum hins vegar ekki að bátar syndi. Það sem báturinn gerir líkist þó sundi t.d. sela alveg jafnmikið og ferðir flugvélarinnar líkjast flugi fugla. Það eina sem kemur í veg fyrir að við eignum báti sundhæfileika er málvenja. Þessi málvenja gæti verið á annan veg án þess að hugmyndir okkar um báta, siglingar og sund breyttust að ráði. Þeir tímar koma ef til vill að það verði spurning um málvenju fremur en sálfræðilegan veruleika hvort rétt sé að eigna tölvum hugsun.

4. KAFLI: AÐ GERA EINS

Ýmsir þeir sem fjalla um gervigreindarfræði álíta augljóst að hægt sé að gæða tölvur hugsun og skilningi. Þeir segja sem svo að mannsheilinn hljóti að virka með einhverju móti, einhver lögmál hljóti að gilda um starfsemi hans. Hvort sem þessi lögmál kveða á um stafræna vinnu með tákn eða eitthvað annað þá hljóti að vera hægt að smíða forrit sem líkir eftir þeim, lætur tölvu einfaldlega vinna eins og heilinn. Hér er margs að gæta. Það er ekkert einfalt mál að gera grein fyrir því hvenær tveir hlutir vinna eins.

Hvað merkir það þegar við segjum að tveir hlutir, t.d. tvær vélar vinni eins eða fylgi sömu aðferð? Getum við til dæmis sagt að gamaldags reiknivél úr tannhjólum og nútíma rafmagnsreiknivél vinni eins? Við vitum að reiknivélarnar reikna sömu föll. En fara þær eins að því?

Það er vandalaust að setja fram samsemdarmið um föll, þ.e. reglu um hvenær fall sem við köllum a og fall sem við köllum b eru sama fallið.[16] Hins vegar er þrautin þyngri að setja fram samsemdarmið um aðferðir. Eftir því sem ég kemst næst er ómögulegt að setja fram samsemdarmið um aðferðir sem ekki er afstætt við hvernig aðferðunum er lýst.[17] Þetta má skýra með dæmi:

---

15 Um þetta sjá Dennett 1987 bls. 323 - 339.

16 a og b eru sama fall ef þau hafa sama grunnmengi G og (x)(x er stak í G -> a(x) = b(x))

17 Hér kunna algóriþmar eins og tölvufræðin fjallar um, þ.e. aðferðir til að vinna með tákn, þó að vera undanskildir. Ég hef ekki neitt algilt samsemdarmið um algóriþma á takteinum en ég sé ekki að tilvera þess sé á neinn hátt útilokuð.

Hugsum okkur tvö vélmenni, annað á hjólum og hitt með fætur. Hugsum okkur að vélmennin séu forritanleg á æðra forritunarmáli sem meðal annars hefur grunnaðgerðirnar *áfram gakk* og *hægri snú* þannig að sé þeim skipað *áfram gakk 10*, *hægri snú 45* þá gangi þau 10 metra áfram og beygi svo um 45 gráður til hægri. Nú getum við matað bæði vélmennin á sömu rununni af *áfram gakk* og *hægri snú* skipunum með þeim afleiðingum að þau ganga bæði sams konar krókaleið og miðað við þessar grunnskipanir getum við sagt að vélmennin fylgi sömu aðferð.

Áður en vélmennin framkvæma skipanirnar *áfram gakk* og *hægri snú* þýða þau þær á vélamál sem hafa einfaldari grunnaðgerðir. Annað vélmennið þýðir skipunina *áfram gakk 10* yfir í skipanir sem merkja *snúa öllum hjólum 10 hringi* og hitt yfir í einhverjar fótahreyfingar. Þegar búið er að orða aðferðirnar á vélamálunum eru þær ekki lengur sömu aðferðir, enda blasir það við að þessi tvö vélmenni nota ekki sömu aðferð til að hreyfa sig. Annað hreyfir sig með því að snúa hjólum hitt með því að lyfta fótum og beygja hné.

Ef til vill dugar þetta dæmi af vélmennunum. Ég ætla samt að bæta öðru við. Ef við lítum svo á að Flugleiðir og Norðurleið beiti grunnaðgerðunum: *hleypa farþegum inn í farartæki*, *stýra farartæki frá Reykjavík til Akureyrar* og *hleypa farþegum út úr farartæki* þá getum við sagt að þessi fyrirtæki noti sömu aðferð til að koma farþegum milli Reykjavíkur og Akureyrar. Miðað við þetta val á grunnaðgerðum fara þau eins að. En ef við lýsum því sem Flugleiðir gera með grunnaðgerðum eins og að *taka á loft* og *lenda* þá fara þessi fyrirtæki ekki eins að. Til að eigna þeim sömu aðferð þarf að velja mjög flóknar og almennt orðaðar grunnaðgerðir.

Í ljósi þessa getum við sett fram afstætt samsemdarmið um aðferðir svona: Ef hægt er að lýsa því sem hlutur a gerir og því sem hlutur b gerir með sömu grunnaðgerðum g1, g2, g3 o.s.fr. og báðir hlutirnir framkvæma sömu röð grunnaðgerða að gefnu sama inntaki eða sömu kringumstæðum þá fylgja a og b sömu aðferð miðað við þetta val á grunnaðgerðum.

Af þessu leiðir að það er yfirleitt tómt mál að tala um að tveir ólíkir hlutir vinni eins eða fylgi sömu aðferð í einhverjum algildum skilningi. Ólíkir hlutir geta aðeins unnið eins eða eftir sömu aðferð ef orðin "eins" og "sömu" eru skilin afstætt. Aðferðin er söm eða eins miðað við tiltekið val grunnaðgerða.

Þegar sagt er að tveir hlutir vinni eftir sömu aðferð má alltaf spyrja að hvaða marki? Er aðferðin bara söm rétt á yfirborðinu þannig að það þurfi að skipta verkinu niður í mjög flóknar og almennt orðaðar grunnaðferðir til að tala um að hlutirnir fari eins að eða eru aðferðirnar eins alveg niður úr eða eru þær eins á einhverju milliplani?

Heimspekingurinn Daniel Dennett stendur framarlega í flokki þeirra sem telja að mögulegt sé að gæða tölvur raunverulegu viti og skilningi. Í frægri ritgerð sem heitir "Cognitive Wheels: The Frame Problem of AI" kallar Dennett skýringar á mannlegu hugarstarfi sem stangast á við alla líffræði "hugsanahjól" og líkir þeim þannig við 300 ára gamlar skýringar á hreyfingum vöðva og beina sem gerðu ráð fyrir að mannslíkaminn væri vél með reimum, hjólum og öðrum búnaði sem stingur í stúf við alla líffræði. Í þessari grein segir Dennett að margir andmælendur gervigreindar séu sannfærðir um að öll dæmi um gervigreind hljóti að vera gírkassar fullir af hugsanahjólum og engu öðru en þau rök sem oftast eru færð fyrir þessu byggist á misskilningi á aðferðum gervigreindarfræðinnar. Hann segir að lýsingar á hermilíkani af einhverju vitsmunastarfi geti verið með ýmsu móti allt frá því að nota orðalag eins og haft er um fólk, niður í lýsingar á forriti á einhverju æðra máli og jafnvel enn

lengra niður að grunnaðgerðum sem eru byggðar inn í vélbúnaðinn. Síðan segir Dennett:

> Engum dettur í hug að líkanið samsvari sálfræðilegum og líffræðilegum veruleika *alveg niður úr*. Því er aðeins haldið fram að á einhverjum hærri stigum fyrir neðan stig fyrirbærafræðilegra lýsinga [below the phenomenological level] /.../ samsvari hermilíkanið því sem líkt er eftir, þ.e. vitsmunastarfi lifandi vera, t.d. manna.[18]

Áttum okkur aðeins á hvað þetta þýðir. Þetta þýðir væntanlega að tölvulíkan af hugarstarfi þurfi að vinna eins og mannshugurinn miðað við eitthvert safn grunnaðgerða. En nú hljótum við að spyrja hvaða safn grunnaðgerða? Á hvaða stigi þarf tölvan að vinna eins og heilinn til að hún hugsi og skilji í raun og veru?

Á hversdagsmáli lýsum við hugarstarfi manna með því að tala um tilfinningar, skoðanir, geðshræringar, skynjanir, reynslu, vilja, ætlun, skapgerð o.s.fr. Á máli líffeðlisfræðinnar er heilastarfseminni lýst með tali um taugafrumur, boðefni, rafstrauma eða eitthvað því um líkt. Hingað til hafa hvorki hversdagsleg hugtök né líffeðlisfræðileg dugað til að orða nein náttúrulögmál um atferli fólks og sálarlíf. En gervigreindarfræði og sálfræði sem byggir á tölvulíkönum gera yfirleitt ráð fyrir að til sé eitthvert stig á milli hversdagsmálsins og líffeðlisfræðinnar og með hugtökum á þessu stigi megi fanga eðli hugarstarfsins, skýra mannlega hegðun og orða lögmál sálarfræðinnar.

Það má líka lýsa hegðun tölvu bæði með hversdagslegu orðalagi og með orðalagi raftækninnar. Ef tölva er að tefla lýsum við því sem hún gerir með orðalagi eins og: "Hún er að reyna að valda biskupinn.", "Hún ætlar að hóta drottningunni í næsta leik." Sá sem kann skil á rökrásum og rafeindatækni getur lýst því sem gerist inni í tölvu með því að tala um straum, spennu, gikkrásir, rökhlið og því um líkt. En þegar tölvur eru annars vegar er til millistig milli hversdagsmálsins og rafmagnsfræðinnar. Það er hægt að lýsa hegðun vélarinnar með orðalagi hugbúnaðarfræða og tala um breytur, gagnaform, algóriþma, undirforrit og því um líkt. Með þessum hugtökum er hægt að útskýra hegðun tölvu fullkomlega, að minnsta kosti ef vélbúnaðurinn virkar rétt. Það er til dæmis hægt að útskýra hvers vegna hún velur þennan leik frekar en hinn ef hún er að tefla skák og slík útskýring er trúlega aðeins möguleg með hugtökum á þessu milliplani.

Eins og flestir talsmenn gervigreindarfræða álítur Daniel Dennett að til sé eitthvert ámóta millistig fyrir mannshugann eins og fyrir tölvuna. Það sé hægt að lýsa hugarstarfi manna með hugtökum sem eru einhvern veginn á milli hversdagsmáls og líffeðlisfræði og með þeim hugtökum megi útskýra hvernig mannshugurinn virkar. Ýmsir andmælendur gervigreindarfræði og hugfræði efast hins vegar um að þetta millistig sé til.

*

Talsmenn gervigreindarfræða væna andstæðinga sína stundum um dultrú eða hindurvitni. En það þarf enga dultrú til að efast um möguleikana á að gæða tölvur viti og skilningi. Forsendur slíks efa geta í fyrsta lagi verið kenningar í þá veru að milliplan eins og hér hefur verið rætt um sé ekki til. Ef til vill má svara svona

---

18 Dennett 1984 bls. 166.

mótbárum með því að benda á að sé ekki til neitt milliplan þá sé allt ént hægt að búa til hermilíkan af starfsemi taugafrumanna, eða smíða tauganet sem samsvarar mannsheila, og fanga þannig allt eðli hugsunarinnar. Ef menn geta orðað reglur sem taugafrumurnar fylgja þá er fræðilega mögulegt að herma eftir þeim þótt trúlega takist seint að búa til hermilíkan eða tauganet sem líkir í senn eftir öllum þeim þúsundum milljónum taugafruma sem mannsheilinn er samsettur úr.[19] En við megum ekki gleyma því að um slíkt líkan mætti spyrja að hvaða leyti það líki eftir taugafrumunum. Svarið við þessari spurningu getur aldrei verið: Að öllu leyti. Það er mögulegt að hugsun sé eins og fluga þannig að ekkert hermilíkan geti fangað allt eðli hennar án þess að samsvara "sálfræðilegum og líffræðilegum veruleika alveg niður úr".[20] Að svo miklu leyti sem mannlegu hugarstarfi er á þennan veg háttað getur sálfræði aldrei orðið annað en framlenging á líffræði og lífeðlisfræði.

Í öðru lagi geta efasemdir um kenningar Turing byggt á þeirri skoðun að skilningur og hugsun séu eiginleikar sem ekki er hægt að höndla með hermilíkani af hugsun neitt frekar en forrit sem hermir eftir slagveðri getur gert okkur rennblaut og feykt okkur um koll. Regn og vindar eru ekki hlutir af því tagi sem spretta fram við keyrslu forrits. Ef til vill má færa einhver rök fyrir því að hugsun og skilningur séu að þessu leyti eins og höfuðskepnurnar vatn og loft.[21] Mér vitanlega hefur það ekki verið gert og mér þykir ekki trúlegt að það sé hægt.

Þessar tvenns konar efasemdir gera ráð fyrir því að hugsun og skilningur séu á einhvern hátt bundin við efnafræðilega og eðlisfræðilega eiginleika taugakerfisins. Eftir því sem best er vitað veltur hugarstarf þó mest á boðskiptum milli taugafruma sem hægt er að láta rásir úr öðru efni líkja eftir.

Þriðju forsendurnar fyrir efasemdum um að hægt sé að gæða tölvur viti eða skilningi gætu verið rök gegn því að hugsun og skilningur hafi neitt eðli sem hægt er að höndla með lýsingum á grunnaðgerðum og hegðunarreglum sem lýsa heila-starfsemi einstaklinga. Slík rök gætu til dæmis vísað til þess að hugsun og skilningur eru afsprengi mannlegra samskipta og það er því mögulegt að ekkert sem gerist inni í heilanum dugi til þess, eitt og sér, að kveikja raunverulega hugsun. Til að skapa raunverulega hugsun gæti þurft að herma eftir heilu samfélagi, umhverfi þess og sögu.

---

19 Talið er að í mannsheila séu um það bil 10^{12} frumur og þar af um 10^{11} taugafrumur sem tengjast í net með um það bil 10^{15} taugamót. Hraðvirkar tölvur afkasta um 10^{9} einföldum aðgerðum á sekúndu eða um það bil 100 sinnum minna en miðtaugakerfi býflugu.

Með þeirri tækni sem nú er til vantar mikið á að hægt sé að smíða tölvu sem forrita má til að herma eftir hverri taugafrumu í miðtaugakerfi manna og væntanlega er langt þangað til tæknin kemst á það stig að þetta verði hægt þó ekki sé vegna annars en þess að rafeindabúnaður tölvu er um 100 milljón sinnum orkufrekari en heilafrumurnar í okkur. Tölvur eyða um 10^{-7} joule á hverja einfalda aðgerð en miðtaugakerfið um 10^{-15} joule. (Sjá Churchland & Sejnowski 1992 bls. 9.)

20 Í lokakafla Penrose 1989 setur stærðfræðingurinn Roger Penrose fram þá tilgátu að sumir þættir í hugarstarfi manna velti á skammtafræðilegum eiginleikum efnisins í heilanum og að þessir eiginleikar geri heilastarfsemina óútreiknanlega í ströngum skilningi þannig að útilokað sé að tölva eða Turingvél líki eftir henni. Ýmsir aðrir hafa gælt við svipaðar hugmyndir um að skýringa á æðra hugarstarfi sé að leita innan um leyndardóma öreindanna í lítt mótuðum viðbótum við skammtafræðina, sjá t.d. Hameroff 1994 og Nunn, Clarke og Blott 1994. Ég held að ekki sé á neinn hallað þótt ég segi að kenningar í þessum dúr byggist á einum saman vangaveltum og styðjist ekki við önnur rök en þau að furðuleg fyrirbæri hljóti að eiga sér langsóttar skýringar.

21 Searle virðist hallast að kenningum í þessum dúr. Sjá Searle 1980 og 1984.

Að svo miklu leyti sem þessar forsendur eru réttar getur sálfræði trúlega aldrei orðið aðgreind frá sögu og félagsfræði.

Það er óljóst að hve miklu leyti mannlegir vitsmunir eru háðir mannlegum skynfærum, sköpulagi og samfélagsgerð. Kannski getur enginn hugsað eins og maður nema hafa mannslíkama og alast upp sem barn í mannlegu samfélagi.[22] Sé þetta rétt þá er ef til vill vonlaust að tölva standist nokkurn tíma Turingpróf.[23] En þetta útilokar ekki að hægt sé að forrita tölvur þannig að þær hugsi, bendir aðeins til að hugsun þeirra verði að einhverju leyti öðru vísi en hugsun manna.

Að svo miklu leyti sem þessar þriðju forsendur nýtast til að sýna fram á að vél geti ekki hugsað eins og maður duga þær til að rökstyðja að vitsmunaverur sem kunna að vera til á öðrum hnöttum geti það ekki heldur, að minnsta kosti ekki ef líkamsbygging þeirra og samfélagsgerð víkur að ráði frá því sem gerist hér á jörð. Engum dettur samt í hug að hægt sé að sanna það með heimspekilegum rökum að íbúar annarra himintungla geti ekki hugsað.

Þær forsendur fyrir efasemdum um að hægt sé að gæða tölvu hugsun og skilningi sem ég hef rætt um eru ekkert yfirmáta sennilegar. En hvað sem annars má um þessar efasemdir segja þá er hægt að halda þeim fram án þess að bera á borð neitt dultrúarmoð.

*

Ég hef látið að því liggja að þótt mannshugurinn sé ekki táknkerfi þá kunni samt að vera mögulegt að gæða tölvu raunverulegri hugsun og raunverulegum skilningi. En hvernig þarf hugsun að vera til að hermilíkan geti fangað allt eðli hennar með þessum hætti? Við þessu hef ég því miður ekki svar. Ég get þó bent á að sumir hlutir og sum verk skilgreinast af hlutverki eða tilgangi og þegar þannig háttar er munurinn á veruleika og eftirlíkingu yfirleitt óljós eða enginn. Sem dæmi um svona hluti má nefna stól. Stóll skilgreinist af tilgangi sínum. Hann er til að sitja á og eftirlíking af stól er raunverulegur stóll ef það er hægt að sitja á henni. Það er ekkert vit í að segja um slíka eftirlíkingu að hún sé nánast óþekkjanleg frá alvöru stól en samt bara eftirlíking. Sem dæmi um hlut sem ekki skilgreinist af tilgangi sínum eða hlutverki má nefna ljósmynd. Það er alveg sama hvað við teiknum nákvæma eftirlíkingu af ljósmynd hún verður ekki alvöruljósmynd heldur bara eftirlíking og það er fullt vit í að spyrja hvort mynd sem enginn þekkir frá ljósmynd sé raunveruleg eða fölsuð.

Margvíslegir andlegir hæfileikar skilgreinast af tilgangi eða hlutverki. Sem dæmi má nefna fyndni. Ef vélmenni getur fengið fólk til að veltast um af hlátri þá hljótum við að eigna því alvöru fyndni. Það er einfaldlega ekkert vit að tala um eftirlíkingu af fyndni. En hvað með hugsun og skilning? Geta verið til eftirlíkingar af hugsun og skilningi?

---

22 Þetta er sennilegt ef maður reynir t.d. að hugsa sér að valtari aki yfir tærnar á sér. Hugsunin virðist í því fólgin að kreppa tærnar og bíta saman tönnum og þetta er ekki hægt að gera án þess að hafa tær og tennur.

23 Kenning Turing hefur verið gagnrýnd ítarlega á þessum forsendum af bandaríska heimspekingnum Hubert Dreyfus. Nýlegar útgáfur þessarar gagnrýni eru í Dreyfus & Dreyfus 1986 og 1988.

5. KAFLI: HUGSUN OG SKILNINGUR

Formhyggja um hugsun var komin fram þegar á 17. öld og hún á sér enn eldri rætur í nafnhyggju (nómínalisma) 14. aldar. Flestir þeir sem fást við gervigreindarfræði og telja að tölvufræðin sé lykill að nýjum uppgötvunum í sálarfræði aðhyllast einhvers konar formhyggju. Þeir frægu gervigreindarfræðingar Newell og Simon orða formhyggju sína svona:

> Rannsóknir í rökfræði og tölvufræði hafa sýnt að vitsmunir eru fólgnir í efnislegum táknkerfum. /.../
>
> Táknkerfi er safn af munstrum og ferlum. Ferlin geta framleitt munstur, eyðilagt þau og breytt þeim. Mikilvægasti eiginleiki munstranna er sá að þau geta staðið fyrir hluti, ferli eða önnur munstur og ef munstur stendur fyrir ferli er hægt að túlka það. Túlkun er í því fólgin að framkvæma viðkomandi ferli. Tveir merkustu flokkar táknkerfa sem okkur er kunnugt um eru menn og tölvur.[24]

Frægustu rök sem fram hafa komið gegn kenningu Turing eiga að sýna að formhyggja um hugsun geti ekki staðist. Þessi rök setti bandarískur heimspekingurinn John Searle fram í ritgerð sem hann birti árið 1980 og heitir "Minds, Brains and Programs".[25] Í þessari ritgerð rökstyður Searle þá niðurstöðu að þótt einhvern tíma verði kannski til tölvur sem geta hermt eftir fólki, talað eins og fólk og unnið sömu verk og fólk, þá geti þær ekki skilið mál eða hugsað eins og fólk. Að áliti Searle getur tölva sem hermir eftir hugsun ekki hugsað neitt frekar en tölva sem hermir eftir slagveðri getur feykt okkur um koll.

Það má endursegja rökfærslu Searle einhvern veginn svona:[26] Tölvur gera ekkert annað en að vinna með munstur og tákn eftir forriti, þ.e.a.s. reglu eða forskrift. Þetta geta menn líka gert og þannig leikið allar sömu kúnstir og tölvur. Searle getur t.d. hugsað sér að hann sitji inni í herbegi með leiðbeiningar eða forrit á sínu móðurmáli og bunka af spjöldum með undarlegum myndum eða táknum. Hann sér miða með skrýtnum myndum koma inn um lúgu eða glugga á herberginu og flettir upp í leiðbeiningunum hvernig brugðist skuli við og sér að þar stendur að réttu viðbrögðin við þessum myndum séu að láta spjöld með kriss-krass merki og kross-riss merki út úr herberginu. Skömmu seinna koma nýir miðar inn og Searle flettir upp í leiðbeiningunum hvernig bregðast skuli við og skilar réttum spjöldum út. Þannig líða dagarnir og Searle veit ekkert hverjir setja þessa miða inn eða hvers vegna. Hann hlýðir bara leiðbeiningunum og þær segja ekkert um hvað myndirnar eða merkin á spjöldunum þýða heldur lýsa aðeins útliti þeirra.

Utan við herbergið eru kínverjar og þeir stinga miðum með spurningum á kínversku inn um lúguna. Forritið sem Searle vinnur eftir lætur hann stinga út miðum

---

24 Newell og Simon 1976 bls. 130.

25 Ýmsar hugmyndir í þessari ritgerð eru betur útskýrðar í Searle 1984.

26 Sjá Searle 1980 bls. 68 til 72.

með réttum og eðlilega orðuðum svörum við þessum spurningum. Kínverjunum virðist sem herbergið, eða sá sem í því er, skilji kínversku. En Searle skilur ekki neitt. Veit ekki einu sinni að táknin á blöðunum sem hann fær eru spurningar og miðarnir sem hann réttir út eru svör.

Af þessari sögu dregur Searle þá ályktun að fyrst hann skilur ekki kínversku og leggur enga merkingu í táknin þótt hann fylgi svona leiðbeiningum þá muni forrit sem lætur tölvu haga sér eins og hún skilji mál og geti spjallað um alla heima og geima ekki gæða hana neinum raunverulegum skilningi. Searle rengir það ekki að hægt sé að láta tölvu herma eftir málskilningi og vitsmunum manna. En af þessari rökfærslu dregur hann þá ályktun að tölva sem þetta gerir skilji ekki neitt, hún láti bara sem hún skilji. Skilningur hennar sé ekki raunverulegur heldur bara eftirlíking. Hann hafnar semsagt þeirri kenningu Turing að tölva sem stenst Turingpróf geti verið gædd raunverulegum skilningi og raunverulegum vitsmunum og segir að svona eftirlíking af skilningi sé ekki raunverulegur skilningur neitt frekar en hermilíkan af slagveðri sé raunverulegt slagveður.

Searle telur að maðurinn í kínverska herberginu skilji ekki kínversku og að tölva geti aldrei skilið mál eða hugsanir vegna þess að gögn, eða munstur, sem unnið er með eftir formlegum reglum, eins og þeim sem maðurinn í kínverska herberginu og tölvur fylgja, hafa eingöngu form (syntax) en ekki neina merkingu (semantík), þau séu teikn en ekki tákn svo notað sé orðalag Þorsteins Gylfasonar.[27] Þau eru ekki um neitt, vísa ekki á neitt. Fyrir tölvunni og manninum í kínverska herberginu eru engin tengsl milli táknsins fyrir hund og hunda. Táknið er bara strik og deplar en kemur hundum ekkert við. Fyrir þeim sem getur hugsað í raun og veru hefur táknið hins vegar merkingu. Það eitt að færa munstur (eða aðra hluti) til eftir formlegum reglum getur, að áliti Searle, ekki gefið þeim merkingu eða látið þau vera um eitthvað eða vísa á eitthvað.

Rök Searle virðast við fyrstu sýn ósköp einföld. En þau tengjast þó ýmsum flóknum frumspekilegum vandamálum, ekki síst erfiðri gátu sem John Locke glímdi manna fyrstur við, svo ég viti, og má orða einhvern veginn svona: Þegar ég segi setningu hvað þarf ég þá að gera, annað en að hreyfa talfærin, til að ég meini eitthvað með henni? Svarið sem Locke gaf er á þá leið að ég þurfi að tengja orðin við hugmyndir eða endurskin af reynslu í huga mér. Eitt merkasta viðfangsefni heimspekinga á 20. öld hefur verið að skrifa neðanmálsgreinar við þetta svar Lockes.

Hér er ekki rúm til að gera grein fyrir heimspeki máls og merkingar eins og hún hefur þróast á þessari öld svo ég sný mér aftur að Searle. Hann gerir ráð fyrir að hægt sé að orða leiðbeiningar á ensku um hvernig svara skuli spurningum á kínversku án þess að þær innihaldi neinar upplýsingar um merkingu kínverskra orða. Þetta stenst ekki. Ég kann ekki kínversku og get því ekki tekið dæmi á því máli. En hugsum okkur að miðarnir innihaldi ekki kínversk tákn heldur íslensk orð og forritið sem Searle fylgir segi honum hvernig á að svara spurningum á íslensku.

Forritið getur ekki útskýrt hvernig á að svara spurningum á íslensku án þess að innihalda töluverðar upplýsingar um merkingu íslenkra orða. Hugsum okkur til dæmis að við látum eftirfarandi setningu inn um lúguna: "Jóna gat ekki farið í búðina því hún var lasin en Sigga gat ekki farið í sjoppuna því hún var lokuð." Leggjum síðan

---

[27] Sjá Þorsteinn Gylfason 1985. Svipuð andmæli gegn kenningu Turing má finna í Singh 1966 bls. 198 - 200.

spurninguna "Hver var lasin og hver var lokuð?" fyrir Searle. Leiðbeiningarnar sem hann fylgir verða að segja honum að þar sem orðið "hún" kemur fyrir í fyrra skiptið eigi það við Jónu en ekki búðina en þar sem það kemur fyrir í seinna skiptið eigi það við sjoppuna en ekki Siggu. Eina leiðin til að forritið upplýsi Searle um þetta er að það innihaldi almennan fróðleik eins og þann að búðir verði ekki lasnar og Siggur séu yfirleitt ekki lokaðar og þessar upplýsingar varða merkingu íslenskra orða.

Þessum athugasemdum mundi Searle trúlega svara með því að segja að þær upplýsingar um merkingu orða sem hægt er að byggja inn í forrit séu merkingarlausar því tölva geti ekki túlkað þær eða tengt við veruleikann. Sé tölva mötuð á upplýsingum um að orðið "Sigga" sé mannsnafn og mannsnafni geti fylgt umsögnin "er lasin" en síður umsögnin "er lokuð" þá getur hún tengt orðið sem við ritum "Sigga" við orðin sem við ritum "er lasin" en þetta gæðir hana ekki skilningi á því hvað orðin þýða.

Hér kann Searle að hafa nokkuð til síns máls. Til að gefa orðum merkingu og skilja hvað þau þýða dugar ekki bara að geta tengt þau öðrum orðum. Það þarf líka að tengja þau veruleikanum, eða einhvers konar endurskini af reynslu málnotenda, og það hefur maðurinn í kínverska herberginu sáralítil tök á að gera. En ef hann fylgir reglu sem segir honum að veifa einu kínversku tákni í hvert sinn sem tígrísdýr labbar fram hjá glugganum og öðru í hvert sinn pandabjörn kíkir inn um dyrnar þá er hann byrjaður að tengja orðin við veruleikann.

Eftir því sem ég best veit er það víxlverkun milli málnotenda og þess umhverfis sem þeir hrærast í sem gefur mannlegu máli og hugsun merkingu og ég get ekki betur séð en tákn sem tölva vinnur með geti öðlast merkingu með sama hætti. Vél getur tengt tákn við önnur tákn og hún getur líka tengt þau við boð frá umhverfinu og viðbrögð við þeim. Ef Searle heldur að við getum gefið orðum okkar og hugsunum merkingu með einhverjum öðrum hætti en þessum þá er ég hræddur um að honum skjátlist.[28]

Í greininni frá 1980 reynir Searle að svara mótbárum af þessu tagi. Hann ræðir þann möguleika að hann sé ekki lokaður inni í herbergi heldur komi í stað

---

28 Hér hef ég ýmsa af merkustu málspekingum aldarinnar eins og Ludwig Wittgenstein og Willard von Orman Quine á mínu bandi. Sjá Quine 1960 2. kafli og Quine 1969 bls. 29 o.áf. og Wittgenstein 1978 greinar 138 - 242.

Rök Quine sýna að ekkert í hegðun manna og atferli geti gefið orðum þeirra merkingu þannig að útilokað sé að tvær þýðingar sem stangast á geti verið jafnréttar. Af þessu dregur Quine þá ályktun að merking eða inntak af því tagi sem heimspekingar eins Frege, Brentano og margir fleiri hafa viljað eigna mannlegri hugsun sé ekki til.

Þau rök Wittgensteins sem ég vísa í eru kölluð einkamálsrökin. Sú túlkun þeirra sem ég styðst við er fengin frá Saul Kripke (sjá Kripke 1982). Samkvæmt henni sýna þessi rök að niðurstaða Quines er rétt. En Wittgenstein byggir ekki á heimspekilegri atferlishyggju eins og Quine heldur sýnir fram á að það sem gerist í huga mælanda getur ekkert frekar neglt fasta merkingu við orð hans heldur en hegðun sem er sýnileg öðrum.

Hafi þessir heiðursmenn rétt fyrir sér, eins og ég held að þeir geri, þá hefur mannlegt mál innihald eða merkingu með þeim eina hætti að setningar og tákn tengjast öðrum setningum og táknum og þetta kerfi tákna og setninga tengist reynslu manna og athöfnum á ýmsa vegu. Þessi losaralegu tengsl málsins við umheimin duga ekki til þess að negla sértæk fyrirbæri eins og yrðingar föst við orð okkar og hugsanir. En þau duga ágætlega til að við getum rætt við og notað málið á alla þá vegu sem við gerum.

Um túlkun Kripke á einkamálsrökunum má lesa í grein Eyjólfs Kjalars Emilssonar 1992. Öfugt við mig vill Eyjólfur hrekja þessi rök.

stjórntölvu vélmennis. Hann fær tákn á miðum inn um lúgu en nú koma sum þeirra frá skynjurum og nemum sem tengdir eru við vélmennið og gegna hlutverki skynfæra. Searle skilar táknum út um lúgu og nú lenda sum þeirra í klónum á vélum sem hreyfa liðamót vélmennisins.[29] Searle álítur augljóst að þetta breyti engu. En er eitthvað augljóst í þessu efni? Þegar hingað er komið minna rök Searle á rök sem þýski heimspekingurinn Leibniz setti fram gegn þeirri skoðun að vél geti búið yfir skynjun eða meðvitund. Þessi rök eru í 17. grein *Monadologie* sem Leibniz sendi frá sér árið 1714. Hann segir:

> Við verðum einnig að viðurkenna að það er ekki hægt að skýra skynjun og það sem á henni byggist með vélrænni skýringu, það er að segja með tilvísun til þess hvernig hlutir eru í laginu og hvernig þeir hreyfast. Gerum ráð fyrir að til sé vél sem er þannig byggð að hún láti hugsanir, tilfinningar og skynjanir verða til. Við getum ímyndað okkur að hún stækki, án þess að hlutföll hennar breytist svo við getum gengið inn í hana eins og við getum gengið inn í myllu. Þegar inn kæmi sæjum við hvernig hlutirnir ýta hver við öðrum en við sæjum ekkert það sem útskýrt gæti skynjun. Skýringar á skynjun verður því að sækja til einfaldra verunda en ekki samsettra hluta eða véla.[30]

Nú vill svo til að Searle hefur sjálfur hrakið rök Leibniz. Í bók sinni *Intentionality* segir hann:

> Það væri algerlega hliðstætt við rök Leibniz að halda því fram að H_2O sameindir geti aldrei útskýrt hvers vegna vatn er blautt. Hugsum okkur að við gætum gengið inn í kerfi sameindanna "eins og við getum gengið inn í myllu. Þegar inn kæmi sæjum við hvernig hlutirnir ýta hver við öðrum en við sæjum ekkert það sem útskýrt gæti" bleytu. Í báðum tilvikum værum við að horfa á vitlausa hæð í kerfinu. Áferð vatnsins finnst ekki með því að skoða einstakar sameindir og við verðum ekki vör við skynjanir eins og sjón eða þorsta með því að líta á einstakar taugafrumur eða taugamót.[31]

Þessi rök Searle minna okkur á að þegar komið er langt út fyrir hversdagslegan reynsluheim manna er ósköp lítið að marka hvað okkur finnst sennilegt og hvað við getum ímyndað okkur. Vísindalegar aðferðir verða að taka við af 'heilbrigðri skynsemi'.

Í raun og veru getum við ekki ímyndað okkur samverkun trilljóna vatnssameinda. Við getum heldur ekki ímyndað okkur útkomuna úr því að vél taki við milljónum merkja frá nemum og skynjurum, framkvæmi milljónir einfaldara aðgerða á munstrum og táknum sem láta hana vinna úr þessum merkjum og fletta upp í gagnabönkum sem fylltu heil bókasöfn ef þeir væru ritaðir á pappír og sendi milljónir merkja til úttakstækja. Þeir sem þykjast geta séð það í hendi sér að út úr þessu komi hvorki hugsun né skilningur eru á sama báti og þeir sem gátu séð það í hendi sér að

---

29 Sjá Searle 1980 bls. 76-7.

30 Leibniz 1973 bls. 181.

31 Searle 1983 bls. 268.

menn geti ekki verið komnir af öpum því afkvæmi apa séu ævinlega apar. Þeir gátu ekki ímyndað sér milljón ár.

Rök Searle sanna semsagt ekki neitt. En það er ekki þar með sagt að niðurstaða hans sé röng. Málið er enn opið.

6. KAFLI: NIÐURSTÖÐUR

Ég get hvorki fullyrt að Turing hafi haft rétt fyrir sér né að hann hafi haft rangt fyrir sér. Niðurstöður þessara bollalegginga eru ekki svo ótvíræðar. Mér er líklega óhætt að segja að svo framarlega sem formhyggja um hugsun og skilning er rétt kenning sé fræðilega mögulegt að gæða tölvur hugsun og skilningi, en ekki sé þar með sagt að mönnum muni nokkru sinni takast það. Það er efni í aðra ritgerð að fjalla um að hve miklu leyti formhyggja fangar eðli hugsunarinnar en mér finnst afar ósennilegt að hún geri það að öllu leyti. Það eru engin skörp skil milli hugsunar annars vegar og hvata, geðshræringa, skynjunar og athafna hins vegar. Allt þetta skarast og að svo miklu leyti sem hugsun skarast við hvatir, geðshræringar, skynjanir og athafnir felur hún í sér eitthvað meira en það eitt að sýsla með tákn. Það er ekki víst að þetta útiloki að hægt sé að forrita tölvur þannig að þær hugsi svipað og menn því það er hægt að láta tölvur herma eftir sérhverri reglu sem lesa má úr mannlegu hugarstarfi. Af þessu leiðir þó ekki að hægt sé að láta tölvu herma í senn eftir öllum reglum sem mannsheilinn fylgir eða haga sér að öllu leyti eins og maður.

Sé hægt að fanga allt eðli mannlegrar hugsunar og skýra alla mannlega hegðun með endanlegum fjölda af reglum um starfsemi heilans þá ætti tölva að geta haft mannsvit að svo miklu leyti sem hægt er að forrita hermilíkan af þessum reglum. Rök sem menn hafa fært fyrir því að svona reglur séu annað hvort ekki til eða að ekki sé hægt að forrita hermilíkan af þeim þannig að það vinni sömu verk og heilinn eru ekki mjög sannfærandi. Ýmist eru þau of óljós til að hægt sé meta þau af nákvæmni eða einfaldlega ógild eins og hin frægu rök Searle.

Enn sem komið er vita menn of lítið um sitt eigið hugarstarf til að geta sagt neitt ákveðið um að hve miklu leyti það felst í vinnu með tákn og að hve miklu leyti hægt er að láta tölvu fanga eðli þeirra þátta sem eru ekki fólgnir í vinnu með tákn.

Við vitum ekki hvernig mannsheili vinnur. Raunar vita menn ekki einu sinni hvernig miðtaugakerfi í frumstæðu dýri eins og flugu vinnur. En síðan Turing skrifaði greinina sína um reikniverk og vitsmuni hafa sérfræðingar í tölvufræði, sálfræði, málfræði, heimspeki, líffeðlisfræði og fleiri greinum unnið saman að því að rannsaka hugsun, skilning og skyld efni. Þessi samvinna hefur skilað sálfræðinni, og heimspekilegum fræðum um hugsun og skilning töluvert áleiðis og hver sem endanleg niðurstaða verður má fullyrða að tölvufræðin hefur nú þegar gerbreytt hugmyndum manna um sjálfa sig.

HEIMILDIR:

Boden, M. A. (ritstj.) 1990. *The Philosophy of Artificial Intelligence*, Oxford University Press.

Bolter, J. David. 1984. *Turing's Man*, North Carolina University Press.

Churchland, Patricia S. & Sejnowski, Terrence J. 1992. *The Computational Brain*, The MIT Press.

Churchland, Paul M. 1988. *Matter and Consciousness*, The MIT Press.

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

Cutland, N. J. 1980. *Computability: An Introduction to Recursive Function Theory*, Cambridge University Press.

Clark Andy. 1990. "Connectionism, Competence, and Explanation" í Boden 1990 bls. 281 - 308.

Davis Martin. 1982. *Computability and Unsolvability*, Dover.

Dennett, Daniel. 1984. "Cognitive Wheels: The Frame Problem of AI" í Hookway 1984 bls. 129 - 151. Endurpr. í Boden 1990 bls. 147 - 170.

Dennett, Daniel. 1987. *The Intentional Stance*, MIT Press.

Dennett, Daniel. 1991. *Consciousness Explained*, Little Brown & Co.

Descartes, René. [1637] 1991. *Orðræða um aðferð*, Hið íslenska bókmenntafélag.

Dreyfus, Hubert & Dreyfus, Stuart. 1986. *Mind over Machine*, Free Press.

Dreyfus, Hubert & Dreyfus, Stuart. 1988. "Making a Mind versus Modelling the Brain" í *Artificial Intelligence 117* nr. 1, vetur 1988. Endurpr. í Boden 1990 bls. 309 - 333.

Eyjólfur Kjalar Emilsson. 1992. "Sólin, hellirinn og hugsanir Guðs" í *Skírni haust 1992*, Hið íslenzka bókmenntafélag.

Fodor, Jerry A. 1975. *The Language of Thought*, Harvard University Press.

Gardner, Howard. 1985. *The Minds New Science: A History of the Cognitive Revolution*, Basic Books Inc.

Hameroff, Stuart. 1994. "Quantum Coherence in Microtubules: A Neural Basis for Emergent Consciousness" í *Journal of Consciousness Studies vol.1, no. 1*, Imprint Academic.

Haugeland, John. 1985. *Artificial Intelligence: The Very Idea*, MIT Press.

Herken, Rolf. 1988. *The Universal Turing Machine, A Half-Century Survey*, Oxford University Press.

Hobbes, Thomas. [1651] 1962. *Leviathan*, Collins/Fontana.

Hookway, C. (ritstj.) 1984. *Minds, Machines and Evolution*, Cambridge University Press.

Jón Torfi Jónasson. 1985. "Er vit í tölvuviti?" í *Stúdentablaðinu 4. tbl. 61. árg, júní 1985* bls. 13-16.

Jón Torfi Jónasson. 1992. "Hugur og heili" í Einar Logi Vignisson og Ólafur Páll Jónsson (ritstj.) 1992. *Af líkama og sál, sex erindi um mannshugann*, Einar Logi Vignisson og Ólafur Páll Jónsson.

Kleene, Stephen C. 1988. "Turing's Analysis of Computability, and Major Applications of It" í Herken 1988 bls. 17 - 54.

Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language*, Harvard University Press.

Leibniz, Gottfried Wilhelm. 1973. *Philosophical Writings*, J. M. Dent & Sons Ltd.

McCulloch, Warren S. & Pitts Walter H. 1965. "A Logical Calculus of the Ideas Immanent in Nervous Activity" í McCulloch, Warren S. 1965. *Embodiments of Mind*, MIT Press. Endurpr. í Boden 1990 bls. 22 - 39.

Minsky, Marvin. 1985. *The Society of Mind*, Simon & Schuster Inc.

Newell, Allen og Simon, Herbert A. 1976. "Computer Science as Empirical Enquiry" í *Communications of the Association for Computing Machinery 19*. Endurpr. í Boden 1990 bls. 105 - 132.

Nunn, C. M. H., Clarke, C. J. S & Blott, B. H. 1994. "Collapse of Quantum Field may Affect Brain Function" í *Journal of Consciousness Studies vol. 1, no. 1*, Imprint Academic.

Penrose, Roger. 1989. *The Emperor's New Mind*, Oxford University Press.

Pratt, Vernon. 1987. *Thinking Machines: The Evolution of Artificial Intelligence*, Basil Blackwell.

Putnam, Hilary. 1960. "Minds and Machines" í Putnam 1975 bls. 362 - 385.

Putnam, Hilary. 1967a. "The Mental Life of some Machines" í Putnam 1975 bls. 408 - 428.

Putnam, Hilary. 1967b. "The Nature of Mental States" í Putnam 1975 bls. 429 - 440.

Putnam, Hilary. 1975. *Mind, Language and Reality*, Cambridge University Press.

Quine, Willard von Orman. 1960. *Word and Object*, MIT Press.

Quine, Willard von Orman. 1969. *Ontological Relativity and Other Essays*, Columbia University Press.

Searle, John R. 1980. "Minds, Brains and Programs" í *The Behavioral and Brain Sciences 3* Cambridge University Press. Endurpr. í Boden 1990 bls. 67 - 88.

Searle, John R. 1983. *Intentionality*, Cambridge University Press.

Searle, John R. 1984. *Minds, Brains and Science*, Harvard University Press.

Singh, Jagjit. 1966. *Great Ideas in Information Theory, Language and Cybernetics*, Dover Books.

Turing Alan. 1950. "Computing Machinery and Intelligence" í *Mind LIX*, nr. 2236. Endurpr. í Boden 1990 bls. 40 - 66.

Wittgenstein, Ludwig. 1978. *Philosophical Investigations*, Basil Blackwell.

Þorsteinn Gylfason. 1985. "Teikn og tákn" í *Stúdentablaðinu 4. tbl. 61. árg, júní 1985* bls. 17-19.