

Kristinn R. Þórisson

Vélvitund, meðvitund og sjálfsvitund í kjötvélum og vélmennum

Allar lífverur eru gæddar hæfileikum til að bregðast við áreiti frá umhverfi sínu. Hjá mannfólki er taugakerfið ábyrgt fyrir hæfileikum á þessu sviði – og þá helst sá hluti sem er nefndur „heili“ eða „hugur“, eftir því hvort verið er að tala um líffærið sjálft (heilann) eða það sem líffærið gerir (hugann). Mörgum hafa þótt eiginleikar þessa lífræna fyrirbæris göldrum líkastir. Á tækni- og vísindaöld, þegar menn hafa gengið á tunglinu, kafað ofan í dýpstu hyldýpi sjávar og sent gervihnetti út fyrir endimörk sólkerfisins gerir starfsemi þessa litla líffæris, sem aðeins er 20 sentímetrar í þvermál, almenning jafnt sem sérfræðinga enn orðlausa.¹

Hugsun er eitt flóknasta náttúrulega fyrirbærið sem menn hafa reynt að skilja. Það er engin tilviljun að þekking okkar á einfaldari fyrirbærum, eins og til dæmis efni, er mun lengra komin. Hugsun er birtingarmynd efnisins í sinni flóknustu mynd – það hefur safnast saman á mjög sérstakan hátt og myndað líffræðileg form, gen, sem í kjölfarið móta kerfi,

1 Orðleysi okkar yfir þau mörgu fyrirbæri sem heilinn kemur í kring með starfsemi sinni hefur ekkert með það að gera að *það er hugurinn sjálfur* sem er ábyrgur fyrir skorti á orðum – það hefur enga dýpri merkingu að orsakasambengið virðist endurkvæmt (e. *recursive*). Endurkvæmni hefur aðeins sérstaka merkingu í huga stærðfræðinga og forritara sem finnast þeir komast nær einverjum djúpstæðum eiginleikum alheimsins þegar þeir nýta sér slíka tækni.

heila, sem aftur stjórnar hegðun þessara líffræðilegu fyrirbæra á mjög margbrotinn en reglubundinn máta. Við skiljum hvorugt ferlið nema að mjög litlu leyti en framfarir hafa verið örar undanfarna áratugi. Tölvur eru þar í aðalhlutverki.

Eitt af því sem veldur mönnum heilabrotum er sú afleiðing, fylgifyrribæri eða orsök hugsunar sem við köllum „meðvitund“. Enginn veit í raun hvað meðvitund er en samt vita það allir. Þegar ég geng um götur Reykjavíkur og horfi á umhverfið gerist eitthvað sem ég ætla að gerist ekki í grjóthnullungum. Við getum kallað þetta fyrirbæri „upplifun“, „meðvitund“, eða „meðvitaða upplifun“. Ég læt engan segja mér hvort ég hef þessar upplifanir eða ekki, ég veit sjálfur að ég hef þær – ég *er meðvitaður um mínar eigin upplifanir*. Að sama skapi eru allir aðrir meðvitaðir um sínar eigin upplifanir. Vandamálið er að enginn hefur aðgang að meðvitund eða upplifun annarra – við verðum að *ímynda* okkur að hún sé til.

Sumir vísindamenn og heimspekingar hafa reynt að „losa sig við“ meðvitund með því að skilgreina hana sem hið sama og starfrænir eiginleikar heilans/hugans, það er hvernig hann starfar. Með þessu eru þeir í raun að segja að þegar við höfum útskýrt starfsemi heilans séum við jafnframt búin að útskýra meðvitund – algjörlega. Meðvitund er hvorki meira né minna en starfsemi heilans, samkvæmt þessari hugmynd. Ég tek undir orð Davids Chalmers (1996) þegar hann svarar þessu á eftirfarandi hátt:

Allar tilraunir til að skilgreina meðvitund á starfrænan (e. *functional*) máta eru annaðhvort tilraun til að tala sig út úr vandamálinu eða leikur að skilgreiningum. Við gætum alveg eins skilgreint „heimsfrið“ sem „samloku með skinku“. Mun auðveldara væri að koma á heimsfriði, en árangurinn væri innantómur.²

Er hægt að rannsaka fyrirbæri eins og meðvitund á vísinda-

2 „To analyze consciousness in terms of some functional notion is either to change the subject or to define away the problem. One might as well define ‘world peace’ as ‘a ham sandwich.’ Achieving world peace becomes much easier, but it is a hollow achievement.“ (Chalmers 1996:105).

legan hátt? Vestræn vísindi hafa náð frábærum árangri í að skilja heiminn og við höfum mikla ástæðu til að vænta þess að með áframhaldandi beitingu hinnar vísindalegu aðferðar³ og sleitulausri vinnu munum við smám saman öðlast skilning á flestum þeim fyrirbærum sem enn eru óútskýrð. Það eru hins vegar nokkur atriði varðandi meðvitundina sem fá okkur til að staldra við og íhuga væntingar um árangur af rannsóknum okkar á þessu athyglisverða fyrirbæri.

Vit-vísindi

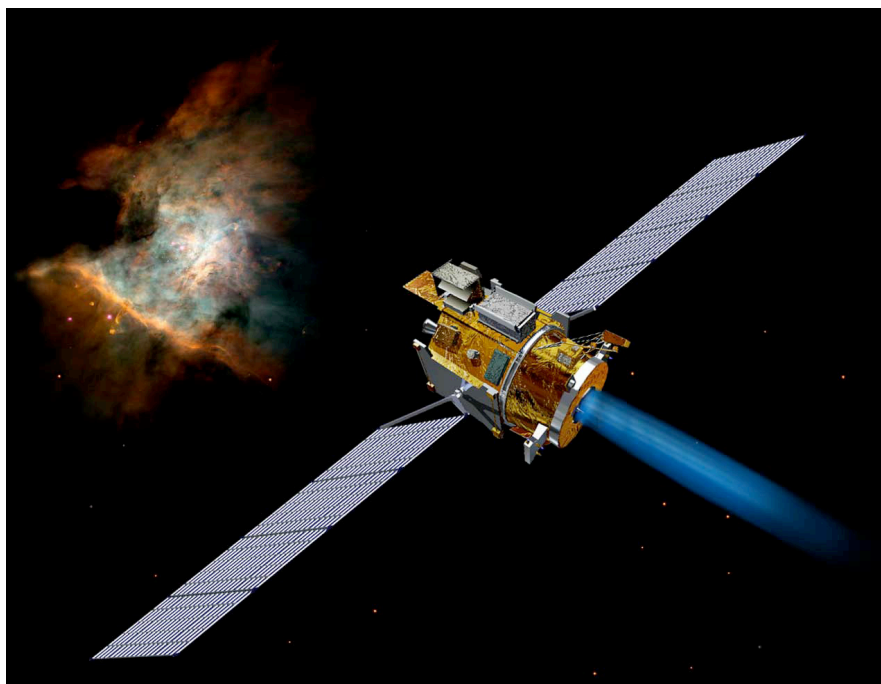
Í þessari grein sýni ég fram á að meðvitund er flókið fyrirbæri, en umfram allt samsett fyrirbæri. Það sem ég á við með þessu er að meðvitund er ekki „atómískt“ eða ósundurgreinanlegt fyrirbæri heldur er hún sett saman úr mörgum þáttum. Til að rannsaka margbrotin fyrirbæri sem samsett eru úr mjög mörgum þáttum verður að beita sérstökum rannsóknar-aðferðum.

Ein gagnlegasta aðferðin sem vísindin hafa þróað til að skilja fyrirbæri í náttúrunni er líkanasmíð. Við smíðina nota vísindamenn blöndu af tilraunastarfsemi og „agaðri hugar-leikfimi“ – hugmyndir eru settar fram í formi vísindalegra til-gátna og prófaðar á tilraunastofu. Í eðlisfræði tókst, á um það bil 2000 árum, að búa til kenningar um hegðun efnis sem hægt er að setja fram á stærðfræðilegu formi. Formúlur lýsa fyrirbærum sem virtust óyfirséanlega flókin, til dæmis gangi himintunglanna. Eðlisfræðin – sem var felld undir heimspeki lengst framan af, eins og reyndar öll vestræn vísindi – komst að því að skipulagi (sumra dauðra) hluta var hægt að lýsa með (tiltölulega) einföldum formúlum.

Menn dreymir auðvitað um að gera hið sama fyrir önnur náttúrufyrirbæri sem eru skemmra á veg komin, svo sem viðfangsefni líffræði, hagfræði,⁴ vistkerfi og greind. Engar líkur

3 Hér er ekki um neina eina aðferð að ræða heldur samþættingu mismunandi aðferða, aðferðafræði og hegðunarmynstra, sem saman gefa okkur betri og haldbærari niðurstöður en aðrar nálganir á viðfangsefnið.

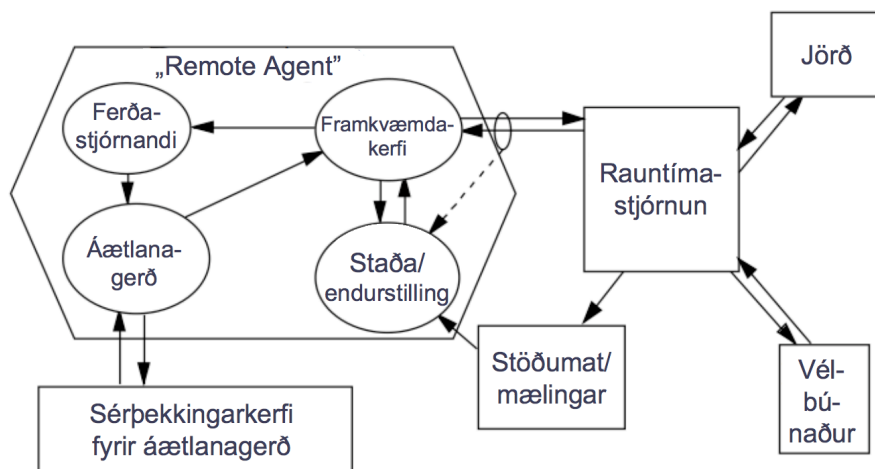
4 Sumir flokka hagkerfi ekki sem náttúrufyrirbæri, en auðvitað er það afurð hugsunar og hegðunar fjölda einstaklinga sem allir eru afurð náttúrunnar.



Mynd 1. Deep Space One geimfar NASA sem sent var út í geiminn 1996 með gervigreind innanborðs. Geimfarið hafði skilning á sjálfu sér sem gerði því kleift að fást við algjörlega ófyrirséð vandamál.

eru hins vegar á því að það gerist, og fyrir því eru tvær megin-ástæður. Önnur er sú að innan þessara sviða er *gífurlegur fjöldi orsakasambanda milli fjölmargra fyrirbæra af mismunandi toga*. Hin er það vandamál sem vitvísindin eiga við að etja að sam-
anburðartilraunum er einungis hægt að beita að takmörkuðu leyti; stór inngrip í starfandi heila sem gætu leitt til djúps skilnings á starfsemi hans eru í flestum tilvikum talin siðlaus og ólöglegt að framkvæma þau á lifandi, hugsandi mannver-
um þar sem ekki er hægt að tryggja nægilega vel gegn óaft-
urkallanlegum skemmdum. Hliðstætt vandamál er að finna í stjarnedlisfræði en þar er útilokað að stilla upp stjarnþokum að eigin geðþótta til að skýra spurningar varðandi svarthol, svo dæmi sé tekið. Tilraunastarfsemi á þessum sviðum er því takmörkuð. Í gervigreindarfræðum, sálfræði og vitvísindum

Simon (1969) kallaði vísindi sem fást við slíkar „annars stigs náttúruafræð-
ir“ „vísindi gervifyrirkæra“ (e. *sciences of the artificial*), en þar notaði hann
einmitt orðið „gervi“ – eins og það er notað í hugtakinu „gervigreind“ – í
merkingunni „gert af mönnum“.



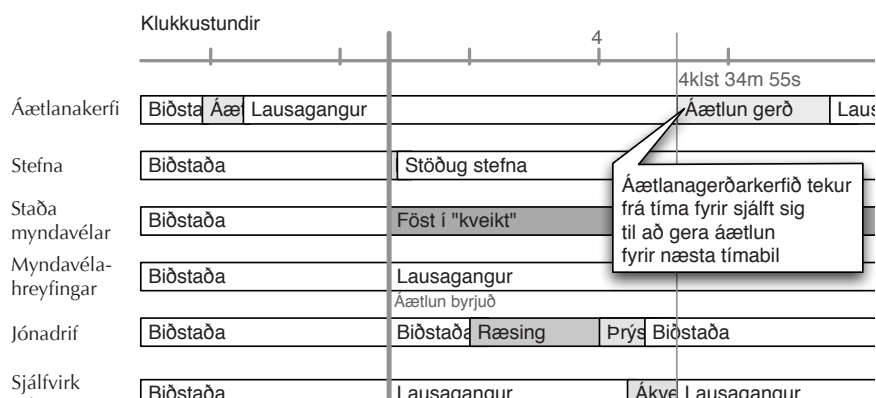
Mynd 2. Gervigreindarkerfi Deep Space One sem nefnt var Remote Agent (Muscoletta o.fl. 1998).

freistast margir til óhóflegrar og oft óígrundaðrar hugarleikfimi, sem gerir lítið annað en að þjálfa huga þeirra sem hana stunda.⁵ Lausnin á báðum vandamálunum: Tölvulíkön.⁶

Tölvur og vélmenni: Vísindaleg verkfæri

Tölvur er verkfæri sem gerir okkur kleift að vinna með margbrotin líkön af hugar- og heilastarfsemi. Með tölvulíkönunum getum við prófað svo til allar hugmyndir okkar um hvernig orsakasamhengi í heilanum gæti verið háttað. Við getum smíðað líkön af því hvernig ályktanir eru dregnar, hvernig

- 5 Í þessu samhengi heyrir því stundum fleygt að vísindalegar rannsóknir á hugsun eða meðvitund séu ekki mögulegar því þar sé hugurinn að reyna að skilja sjálfan sig og að slíkt hljóti að leiða til vandræða sem koma í veg fyrir að það gangi upp – til dæmis vegna endurkvæmni (e. *recursion*), gífurlegs flækjustigs eða röklegra mótsagna. Þetta er náttúrulega bara orðaleikur; þegar mannshugurinn skoðar hegðun heilans – hér er rétt að minna á að hugurinn er það sem heilinn gerir, þ.e. „heili í gangi“ – er heilinn og starfsemi hans nákvæmlega eins og hver annar hlutur/starfsemi sem skoðað er með vísindalegum aðferðum (hugsanlega að meðvitund undanskilinni). Eina augljósa afleiðing þess að hugurinn er að rannsaka sjálfan sig eru svona ranghugmyndir um hvers vegna slíkt á ekki að vera hægt.
- 6 Þótt tölvur séu öflugasta verkfærið til að skoða slík fyrirbæri nú er ekki loku fyrir það skotið að önnur betri verkfæri líti dagsins ljós. Þau munu hins vegar ekki breyta eðli viðfangsefnisins, það er að greind er margbrotið náttúrufyrirbæri sem aldrei verður hægt að lýsa í mjög stuttu máli.



Mynd 3. Deep Space One skilur sjálfan sig nægilega til að vita að það tekur hann tíma að gera áætlunir. Hér sést hvar (4h, 34m, 55s) hann hefur tekið frá tíma til að gera áætlun fram í tímann. (Skjámynd af hermilíkani Deep Space One teymisins.)

flóknu samspili líkama og hugsunar er háttað og hvernig taugafrumur tengjast til að framkalla það sem við köllum hugsun. Tölvulíkon geta til dæmis hjálpað okkur að skilja hvernig hugur okkar skilur fyrirbæri eins og tíma, sem í fyrstu virðist kannski augljóst, af því að við eigum engra annarra kosta völ en að lifa í honum. Við frekari ígrundun kemur þó í ljós að tíminn er mjög furðulegt og snúið fyrirbæri.

Tölvur í líki róbóta með sjálfstæða ákvarðanatöku eru sérstaklega vel til þess fallnar að nota í samanburði við náttúrulega greind fyrirbæri. Við skulum skoða tvo róbóta – annars vegar Deep Space One gervihnöttinn sem NASA sendi út í geiminn 24. október 1998 og hins vegar sýndarvélmennið S1 sem er í þróun í Gervigreindarsetri Háskólans í Reykjavík – sem grunn til að fjalla nánar um þessi mál. Bæði vélmennin kalla ég *vitvélar* þar sem þau geta skynjað flókið umhverfi sitt og tekið skynsamlegar ákvarðanir upp á eigin spýtur út frá markmiðum sínum hverju sinni.

Sá hluti Deep Space One sem skiptir mestu máli hér er hugbúnaðarkerfi sem gerði honum kleift að „skilja sjálfan sig“ að einhverju leyti (Muscoletta o.fl. 1998) og tengsl sín við umhverfið. Kerfið, sem kallaðist Remote Agent, var meðal annars samsett úr gervitauganeti sem stöðugt mat ástand kerfisins, þekkingarkerfi (e. *expert system*) sem skilur

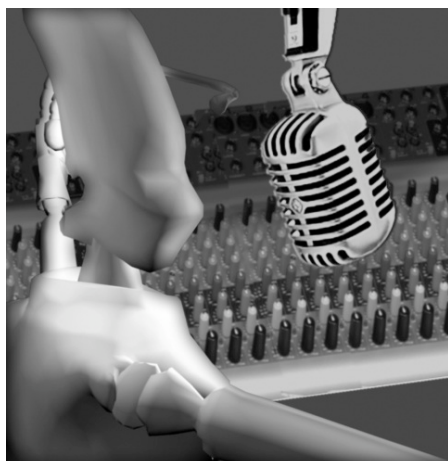
hvernig „líkami“ (vélbúnaður) Deep Space One er samsettur, áætlanakerfi sem stjórnar ferðinni, og framkvæmdakerfi sem sér um að framkvæma áætlanir sem hinar einingarnar búa til. Þekkingarkerfi geimfarsins var svo öflugt – hafði svo nákvæma lýsingu á samsetningu og getu vélbúnaðarins – að það gat dregið ályktanir um hvernig væri best að fást við bilanir, eins og til að mynda þegar rofi á orkugjafa virkaði ekki sem skyldi. Við þær aðstæður „hugsaði“ Deep Space One sem svo: „Takkar eru mekanísk fyrirbæri; slíkir hlutir geta „staðið á sér“; ef eitthvað stendur á sér má reyna að losa það með því að „hrista það“; best að prufa að „hrista“ rofann og sjá hvort hann losnar.“⁷ Viti menn, gervigreindarkerfið gat leyst vandann. Þarna sjáum við dæmi um gervigreint kerfi sem hefur *sjálfsmynd* (þekkingu á sjálfu sér) sem tengist *raunveruleikanum* (efnislegum fyrirbærum sem tengd eru því sjálfu) – Deep Space One er sem sagt gætt vissri tegund af *sjálfsvitund*.

S1 er sýndarvélmenni sem verið er að þróa við Gervigreindarsetur Háskólans í Reykjavík (CADIA). Vélmennið getur haft samskipti við umheiminn í gegnum rödd og heyrn – hann getur hlustað á og skilið talað mál, og talað sjálfur. S1 stjórnar útvarpsþætti og gerir áætlanir um lagaval, spilun og kynningar. Hann getur líka tekið viðtöl, spurt spurninga og gert athugasemdir við svörin. S1 skilur tíma betur en flest önnur sýndarvélmenni sem hafa verið smíðuð, en til þess er hann búinn skynjurum og ákvarðanatökukerfum sem geta brugðist við tali og mannlegum samskiptum á mörgum tíma- kvörðum, frá fáum millisekúndum upp í margar mínútur.

Kjötvélar, vitvélar og tíminn

Skilningur okkar á núinu virðist nátengdur skilningi okkar á okkur sjálfum og tilvist okkar. Þá virðist núíð sérstaklega tengt því hvernig við upplifum raunveruleikann: Við upp-

7 Þekkingargrunur Deep Space One var forritaður af vísindamönnum NASA, en þekkingin um eðli mekanískra fyrirbæra og þekkingin um takka eru ekki tengd á neinn sérstakan hátt í upphafi – sú tenging er alfarið ályktun Deep Space One þekkingarkerfisins, gerð út frá almennri þekkingu þess á því hvernig hlutir og náttúrulögmál haga sér.



Mynd 4. S1 (Superhumanoid One): Gervigreindur útvarpsmaður við Gervigreindarsetur Háskólans í Reykjavík (CADIA).

lifum raunveruleikann alltaf *núna*. Eins og Metzinger (2005) orðar það: „Meðvituð upplifun þýðir að vera *í núinu*“⁸ – ég er að skrifa þessi orð *nú*. Þú ert að lesa orðið „nú“ *nú*. „Þegar ég geri áætlanir um fjarlæga framtíð mína er ávallt ljóst að ég er að gera þær áætlanir *nú*.“⁹ Ef þessi hluti tilveru minnar væri fjarlægður – það er upplifun mín um að vera *ég í núinu* – yrði upplifun mín af tilvist minni engin¹⁰ og upplifun eins og hún leggur sig myndi hverfa; þar sem tími er órjúfanlegur hluti af því sem við köllum raunveruleika myndi ég hætta að vera meðvitaður um umhverfið og þess vegna hætta að vera til gagnvart sjálfum mér, það er að upplifa meðvitað.

Tími er því nátengdur meðvituðum upplifunum. Hvernig skilja S1 og Deep Space One tíma? Af illri nauðsyn skilja þeir hann á marga vegu því án skilnings á tíma gætu þeir ekki athafnað sig í umhverfinu. Nánar tiltekið: Báðar vitvélarnar þurfa að skilja hvað hugtakið „nú“ þýðir. Út frá því þurfa þær jafnframt að skilja framtíðina, að því leyti hvernig þær eiga að haga sér gagnvart ókomnum atburðum, og fortíðina, það er hvað hefur þegar gerst. Án skilnings á núinu geta þessar vélar því gert lítið af viti.

8 „To consciously experience means to *be in a present*.“ (Metzinger 2005:9).

9 „...when making plans about my own distant future ... it is always clear that I am making these plans *now*...“ (Metzinger 2005:21).

10 Eftir Metzinger (2005:21): „My own existence possesses temporal immediacy – a sense of being in touch with myself in an absolutely direct and non-mediated way... If it was possible to substract the phenomenal content now at issue, I would simply cease to exist on the level of subjective experience.“

Grunnur þess að skilja núið er að taka inn upplýsingar (gegnum skynfæri) um það hvað er að gerast í umhverfinu. Með því að sjóða saman ýmsar upplýsingar frá skynfærum má setja fram tilgátur um stöðu raunveruleikans á hverjum tímapunkti. Út frá þeim tilgátum má taka ákvarðanir um hvað skuli gera og hvað skuli ekki gera. Deep Space One reynir að skilja hvar það er statt í geimnum, hver staða ýmissa kerfa er og hvenær það eigi að kveikja og slökkva á ýmsum mælitækjum, sendum og orkugjöfum.¹¹ S1 er með skilning á raunveruleika útvarpsþátta, þar með talið öllu sem getur gerst í viðtali, svo sem að það dragist á langinn, að viðmælandinn vilji fá orðið, að viðmælandinn sé búinn að tala of lengi og þar fram eftir götunum: S1 hefur í raun afar fagaðan skilning á því hvenær viðmælandinn vill fá orðið, hvenær er viðeigandi að spyrja næstu spurningar, og svo framvegis.

Það sem vert er að skoða nánar er sú staðreynd að það tekur tíma að safna upplýsingum og vinna úr þeim. Hvort sem um er að ræða hljóðnema á vélmanni eða eyra á manni þá eru skynfæri mælitæki sem bregðast við eðlisfræðilegum áreitum og breyta þeim yfir á annað form: rafboð. Þessi breyting tekur tíma, þannig að þegar veran veit hvað gerðist í umhverfinu er það þegar búið að gerast. Frá því að viðmælandi segir ákveðið orð líður allt að því hálf sekúnda þar til S1 veit hvaða orð hann sagði. Seinkunin á vitneskjunni um hvort viðmælandinn þegir eða talar er mun styttri, en samt aldrei styttri en 30 millisekúndur. Vélmannið upplifir því raunveruleikann alltaf *eftir að hann gerist* – ákvörðun um að eitthvað sé að gerast „núna“ er alltaf á eftir raunveruleikanum. Í vélmanni með margar tegundir skynjara getur það tekið talsverðan tíma að „sjóða saman“ upplýsingar til túlkunar frá tveimur eða fleiri skynjurum og við það lengist tíminn enn frekar.

Til að fást við seinkunina þarf vitvélin að muna hvenær gögnunum var safnað. S1 tímastimplar gögnin þegar þeim er safnað, setur saman gögn frá mörgum skynfærum, skoðar

11 Þeir sem hafa áhuga að skyggjast inn í hugarstarfsemi Deep Space One Remote Agent geta kíkt á eftirfarandi Java-smáforrit: <http://ic.arc.nasa.gov/projects/remote-agent/applet/TimelineApplet.html>

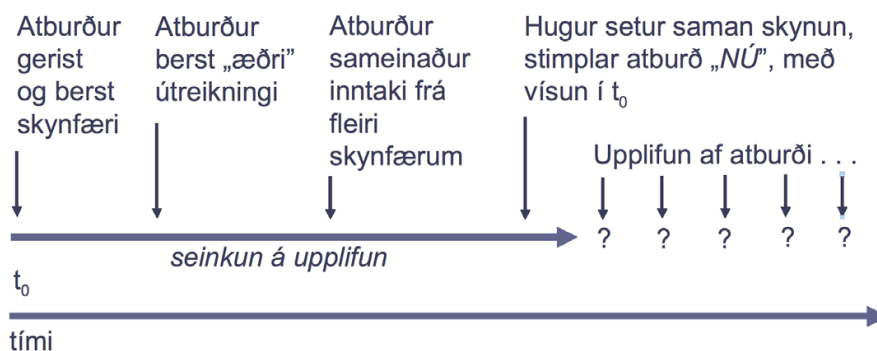
svo klukkuna aftur, dregur frá mismuninn og segir „þetta var að gerast rétt í þessu“. S1 hreinlega afturreiknar hvenær atvik gerðist og merkir það með tímastimpli sem samsvarar þeim tíma sem það gerðist í raun og veru.¹²

En hvað með muninn á vélrænni úrvinnslu, eins og þeirri sem fer fram í örgjörvunum sem gefa Deep Space One og S1 þekkingu á tíma, og því sem við köllum „hugsun“ – þeirri starfsemi sem fer fram í heilanum á dýrum og mönnum?

Minni og skynjun eru líka margþætt fyrirbæri í mönnum. Tökum sem dæmi rýmdarskynjun. Mörg dýr þurfa að geta áttað sig á úr hvaða átt hljóð berst og eru útbúin nemum sem nema hreyfingar á lofti á vissu tíðnisviði; eyru okkar eru gerð fyrir 16 til 20.000 sveiflur á sekúndu. Flest erum við með tvö eyru sem vinna saman þannig að hreyfing á lofti sem á upptök sín hæggra eða vinstra megin við höfuð okkar berst í hvort eyra á mismunandi tíma. Heilinn notar ýmsar upplýsingar, sem hann fær úr hreyfingu loftsins, bæði til þess að flokka hljóðið og til að ákvarða hvort það komi frá vinstri eða hægri. Þegar hljóð fellur á hæggra eyra er því fyrst umbreytt í taugarafboð. Greining á hljóðinu byrjar strax við þá umbreytingu: Taugafrumur bregðast við á mismunandi hátt og senda rafboð á mismunandi hátt eftir því hversu sterkar vissar gerðir tíðni eru í hljóðinu. (Útreikningarnir kalla jafnframt á samanburð við hljóð sem geymd eru, á taugaboðaformi, í langtímaminni lífverunnar sem er að hlusta; þetta er meðal annars forsenda þess að við getum flokkað hljóð sem „píanó“, „mannsrödd“ og svo framvegis.) Hljóðið berst áfram á hljóðhraða, og eftir nokkrar míkrósekúndur lendir það svo á vinstra eyra. Nú fyrst byrjar vandinn: Er þetta sama hljóðið og barst hæggra eyra stuttu áður eða eitthvert annað hljóð? Nokkrar míkrósekúndur eru of stuttur tími til að flokka það sem píanó, mannsrödd eða eitthvað annað, þannig að notast verður við aðrar aðferðir. Hvernig bregst heilinn við því? Þar sem hljóðið er komið á taugaboðaform er kominn grundvöllur til að bera saman tíðnidreifinguna í vinstra eyra

12 Reyndar getur slíkur afturreikningur aldrei orðið meira en nálgun að milli- eða míkrósekúndum, en það er óþarfa smáatriði fyrir þessa umræðu.

Vélvitund, meðvitund og sjálfsvitund í kjötvélum og vélmennum

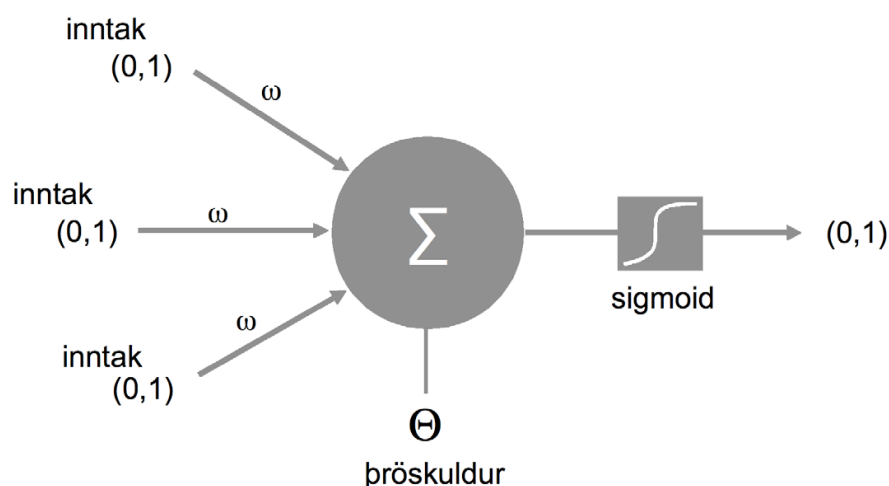


Mynd 5. Tímaseinkun á meðvitaðri upplifun.

við þá úr hægra eyra. Gögn úr hægra eyra eru geymd í hljóðminni til skamms tíma og borin saman við þau sem koma úr því vinstra, stuttu seinna. Flokkun á hljóðtegund er of flókin til að fara nánar út í hana hér; en til að reikna út hvort hljóðið komi frá vinstri eða hægri notar heilinn (meðal annars) einfaldan frádrátt: tímamismuninn milli þess þegar (sama) hljóðið lenti á hægra og vinstra eyra; útkoman gefur okkur beina túlkun yfir í gráðurnar til vinstri eða hægri, frá miðlínu höfuðsins, sem hljóðið barst úr.¹³

Snúum okkur aftur að skilningi og túlkun á *núinu* í dæmunum hér að ofan. Mannsheilinn er ekkert öðruvísi en hugbúnaður vélmennisins þegar kemur að seinkun úrvinnslu á skynjuðum gögnum. Á mynd 5 sést hvernig tíminn líður frá því að atburður gerist. Spurningin er: Hvenær verða skynjanir okkar að upplifun? Það getur ekki orðið fyrr en upplýsingarnar eru á því formi að þær hafi þýðingu fyrir vélmennið eða manneskjuna. Það tekur tíma fyrir sjónfrumurnar að bregðast við ljósbreytingunum og sjóða saman upplýsingarnar frá mismunandi stöðum í útjaðri sjónsviðsins, og senda þær upplýsingar síðan áfram í afturhluta heilans þar sem milljónir frumna taka við þeim. Rannsóknir benda til að meðvituð upplifun um atvik sé á bilinu 30–250 millisekúndum á eftir

¹³ Heilinn nýtir ekki sömu tímastimplunaraðferð og S1 eða Deep Space One til að ákveða úr hvaða átt hljóðið berst en auðvelt er að sýna fram á að aðgerðin í báðum tilvikum er algjörlega sambærileg – og ekki aðeins sambærileg heldur *nægjanleg* til að framkvæma verkið. Með öðrum orðum, heilinn er að *reikna*.



Mynd 6. Reiknilíkan af taugafrumu. Margbrotnari útfærslur eru til sem eru betri nálgun að lifandi taugafrumu, en þær fylgja þó allar þessari grunnhönnun.

raunveruleikanum (Klein 2002, Bergheim o.fl. 1996, Libet o.fl. 1979).

Þar sem upplifun á sér alltaf stað eftir að atburður hefur orðið hlýtur kerfið líka að þurfa minni til að geyma þær upplýsingar sem þarf til að gera upplifun mögulega.

Við komumst að þeirri niðurstöðu, eftir að hafa grandskoðað fjölda slíkra ferla, að heilinn er einfaldlega kjötvél.¹⁴ Hann beitir vissum eiginleikum lífræns efnis til reikniaðgerða. Inntakið er í formi skynáreita, geymslan og útreikningarnir eru í formi taugaboða, sem framkvæmd eru í ýmsum hlutum heilans (Hackett & Kaas 2004). Allir útreikningar taugafrumna, til dæmis sá sem gerir þeim kleift að segja til um hvort hljóð í öðru eyranu sé það sama og í hinu, er gerður á grundvelli samlagningar og frádráttar, þessa sem við lærðum í grunnskóla, $6 - 2 = 4$ (Chalmers 1994). Á mynd 6 sjáum við einmitt líkan af taugafrumu sem notað er í líkanasmíðum í heilarannsóknnum. Hægt er að sýna fram á að það er nóg að svara spurningum um skynáreiti, eins og þeim sem nefndar voru hér að ofan, með hópi af slíkum gervitaugungum, tengdum á vissan hátt og þjálfuðum með réttu áreiti. Engar

¹⁴ Þýðing á *meat computer* – hugtakið er fengið að láni frá Marvin Minsky (persónuleg samskipti).

ástæður eru til að ætla að önnur grundvallarkerfi þurfi til að lýsa starfsemi heilans að fullu.¹⁵

Vissulega er heilinn flókinn en starfsemi hans grundvallast á reikniðgerðum, sem þýðir að hugsun er útreikningur – flókinn útreikningur. Skynjun og minni eru *aðgerðir* sem unnar eru á atvikum í umhverfinu sem breytt er í rafrænar upplýsingar með hjálp skynfæra. Ef heilinn er ákveðin tegund af vél þýðir það að við ættum að geta búið okkur til eintak af honum inni á gervigreindarsetri.

Vélvitund, meðvitund og upplifun

Nú getum við farið að svara spurningum varðandi meðvitund og upplifun. Við sjáum að forsendur upplifunar af skynáreitum eru margþættar og að upplifun er því nauðsynlega *samsett fyrirbæri* í öllum vitverum. Til *nauðsynlegra* hluta meðvitaðrar upplifunar telst *skynjun*, sem er samsett úr tímaskilningi, skilningarvitum (sjón, heyrn, snertiskyn) og annarri hugarstarfsemi; *minni*, sem geymir hina ýmsu þætti til samþættingar; *ýmis önnur úrvinnsla* sem þarf til að styðja þessi fyrirbæri, svo sem langtímaminni, skammtímaminni (hljóðminni, sjónminni), og svo framvegis. Við getum þannig leitt líkur að því að það að vera meðvitaður um fyrirbæri í umhverfinu (eða innri fyrirbæri eins og tilfinningar) sé háð þessum aðgerðum, það er að skynjun, minni og annað slíkt, séu *nauðsynlegar forsendur* meðvitundar.

En hvað sem því líður stendur eftir ákveðið vandamál: spurningin um það hvort S1 og Deep Space One upplifi nokkurn skapaðan hlut meðvitað – það er hvort þeir hafi meðvitaða upplifun. Fáir myndu halda því fram¹⁶ að Deep

15 Endanlegt líkan af taugungi verður að sjálfsgöðu útfært í nánari smáatriðum, t.d. með kerfi til að taka á móti upplýsingum í formi annarra boðefna en taugaboðefna, svo sem adrenalíns, gegndræpi taugafrumuhimnunnar, o.s.frv. Slíkt teldist þó aðeins nánari útfærsla á þessu grunnlíkani.

16 Auðvitað eru það engin rök hversu margir trúa á eina eða aðra skýringu áður en sýnt er fram á að hún sé rétt eða röng. Hins vegar þykir mér líklegt að skilgreining okkar á fyrirbærinu „meðvitund“ myndi þurfa að breytast mjög mikið ef við færum að halda því fram að venjulegt efni, eins og steinar, málning og tölvur, finni til sársauka. Slíkar „alvitundarkenningar“ (e. *pan-*

Space One finni til sársauka þegar við tökum loftnet úr sambandi, eða að S1 finni til spennu eða kvíða rétt áður en bein útsending hefst, eða þegar hljóðneminn er tekinn úr sambandi. Flestir munu eflaust telja það ólíklegt – svo ólíklegt að það sé í raun nánast óhugsandi. Það virðist þurfa fleira til að meðvituð upplifun, eins og hún birtist okkur öllum, verði til.

Við getum talað um *vélræna upplifun*, sem er samþætting skynáreita og flokkun þeirra til að taka ákvarðanir út frá, en slík *vélvitund* svipar ekki mikið til þeirrar meðvituðu upplifunar sem við mannverurnar erum aðnjótandi. Bæði S1 og Deep Space One eru með samþættingu skynjana, minni og túlkun á skynjunum – allt *nauðsynlegar* forsendur meðvitundar – en það virðist eitthvað vanta. Þetta eru nauðsynlegar en ekki *nægjanlegar* forsendur. Kjötvélin hefur eitthvað sem nútíma vitvélar skortir. Einhver eðlismunur virðist vera á tölvum nútímans og kjötvél mannshugans hvað þetta snertir. Hver er hann?

Magn breytir eðli

Sumir hafa lagt til að hæfileikinn til að hugsa um eigin hugsun (e. *meta-cognition*) sé grundvöllur (mannlegrar) meðvitundar. Við gætum reynt að bæta því við listann hér að ofan yfir nauðsynlegar forsendur meðvitaðrar upplifunar, en það væru mistök. Deep Space One hefur þegar slíka hæfileika¹⁷ og það er nokkuð augljóst að þótt við myndum bæta þann hæfileika og gera hann enn öflugri myndi það ekki hafa nein áhrif á eða breyta eðli Deep Space One. Upplifun okkar á litum, til dæmis appelsínugulum lit appelsínunnar, er oft meðvituð upplifun, og slík upplifun virðist ekkert tengjast hugsun um hugsun.

Ein athyglisverð hugmynd sem vert er að skoða hér er hugmyndin um að magn breyti eðli. Ef við hugsum um foss

psychism) verða ekki skoðaðar hér.

- 17 Eins og sést á mynd 3 þar sem Deep Space One áætlar tíma fyrir sjálft sig til að áætla tíma fyrir ferðina („The planner schedules time for itself to plan activity for the next period of time“).

sem samsettur er úr milljörðum dropa, sem hver um sig er samsettur úr milljörðum vatnsatóma, þá er ljóst að bæði fossinn og droparnir eru blautir – en er hvert og eitt vatnsatóm blautt? Hvar er bleytan geymd? Á sviði gervigreindar varð atvik tengt þessari spurningu 1997 þegar Deep Blue, skáktölva sem smíðuð var af vísindamönnum hjá I.B.M., vann Gary Kasparov í heimsmeistaraeinvígi í skák, en Kasparov var þáverandi heimsmeistari. Kasparov hafði áður sagt um Deep Blue að „magn hefði breytt eðli“¹⁸ (Campbell 1996): Sá mikli fjöldi reglna sem tölvan gat beitt fyrir hvern leik og sá gífurlegi fjöldi mögulegra leikja sem hún gat skoðað áður en hún lék – um 50 milljarðar skákfærslna á þeim þremur mínútum sem skákmaður hefur að meðaltali fyrir hvern leik í einvígi – hafði breytt eðli tölvunnar á þann veg að Kasparov fannst sem hann væri að spila gegn nýrri *tegund* greindar.

Er hugsanlegt að hið sama gæti átt við um meðvitund – að meðvitund spretti fram þegar nægilegt magn réttra hluta er tengt saman, til dæmis taugafrumur í tauganetum? Ef svo er þá væri það augljóslega ekki einungis fjöldi taugafrumna, heldur fjöldi taugafrumna sem tengdar eru á ákveðinn hátt (uppbygging, skipulag). Eins og komið hefur fram þá yrðu þær að vera tengdar til að gera samþættingu og túlkun skynjana og minnis mögulega. Þetta er mjög sannfærandi röksemdafærsla þegar greind er annars vegar: Stöðugar framfarir í gervigreind á síðustu áratugum hafa sýnt að eftir því sem reiknigeta eykst og hugbúnaðurinn verður margbrotnari þá eykst geta vélanna til að hegða sér á þann hátt sem við myndum kalla „greindan“.

Hvað varðar meðvitund er „magn-breytir-eðli“-kenningin hins vegar ekki mjög sannfærandi. Við getum reyndar ekki útilokað það fyrr en við höfum betri gögn í höndunum, en það virðist ólíklegt að aukinn fjöldi reikniaðgerða í Deep Blue, Deep Space One eða S1 muni á nokkurn hátt gera þeim kleift að til dæmis finna til. Að minnsta kosti ekki ef um er að ræða reikniaðgerðir eins þær sem við þekkjum og

18 „Quantity had become quality“ (Campbell 1996; <http://mitpress.mit.edu/e-books/Hal/chap5/five4.html>).

höfum þegar gefið þessum gervitverum í einhverjum mæli. Það virðist alveg sama hvað við bætum við mörgum reikni-
aðgerðum, gervitauganetum, vélrænni ályktanahæfni, loðinni
rökfræði (e. *fuzzy logic*), þekkingarkerfum og öðru slíku, stað-
reyndin virðist áfram sú að *það er ekki líkt neinu* að vera þessi
vélmenni. Þar finnst engin tilfinning, enginn sársauki, *engin
meðvituð upplifun*. Hvernig er að vera Deep Blue? S1? Deep
Space One? Svarið er einfalt – *það er ekki eins og neitt*.

Þótt meðvituð upplifun sé *augljóslega til* – upplifun á litum,
sársauka, ánægju og sjálfinu – virðist hún einhvern veginn
vera annars eðlis en þau fyrirbæri sem vísindaleg efnishygga
gerir okkur kleift að skilja, að minnsta kosti eins og er. Eitt
af grundvallarvandamálunum varðandi meðvitaða upplifun
er að hún er staðbundin (e. *local*), það er, lokuð inni í huga
þess sem verður fyrir ákveðinni upplifun – það sem einn aðili
upplifir er engin leið fyrir annan aðila að upplifa á sama hátt,
nema með því að vera *sá sem upplifir*. (Um leið og það gerist
er hann ekki lengur „annar aðili“.)

Við höfum komist ágætlega áleiðis með að skerpa á því hvað
nákvæmlega er átt við með hugtakinu *meðvituð upplifun*, með
því að draga sálræn og líffræðileg fyrirbæri út úr umræðunni,
en enn vantar mikið upp á skilning á *nægjanlegum forsendum*
meðvitaðrar upplifunar. Lykilatriði fyrir skilning á meðvitaðri
upplifun virðist tengjast vandamálinu um lokaða, staðbundna
vitneskju. Vegna þess að meðvituð upplifun er staðbundin og
„innilokuð“ í þeim sem verður fyrir henni, vitum við ekki,
strangt til tekið, hvort S1 finnur til eða ekki. Hann býr ekki
yfir hæfileikanum til að ræða við okkur um það. Við komum
aftur að þessu vandamáli hér á eftir.

Þýðir þetta að við eigum að gefast upp á að reyna að skilja
fyrirbærið? Alls ekki. Þýðir þetta að við verðum að leita ann-
arra aðferða en vísindalegra til að skoða heiminn? Ekki um
sinn að minnsta kosti, þótt auðvitað sé alltaf gott að þróa ný
verkfæri, hvort sem þau eru efnisleg eða hugmyndafræðileg,
til að skilja heiminn betur. Ég ætla nú að ræða um það
hvernig við getum nálgast fyrirbærið á næstu árum og áratug-
um og hvaða niðurstöðu við getum vænst.

Víxlverkunarkenningin

Við skulum nú skoða kenningu um meðvitaða upplifun sem ég kalla „víxlverkunarkenninguna“.¹⁹ Kenningin er algóritmísk²⁰ efniskenning og er ólík mörgum öðrum kenningum að því leyti að hún gengur út á samspil margra þátta í stað þess að reyna að finna „hina einu sönnu orsök“ meðvitundar.

Kenningin gengur í meginatriðum út á víxlverkun milli fyrirbæra og orsakatengsla sem verða til á mismunandi flækjustigum í náttúrunni.²¹ Ef við skoðum lagskiptinguna (mynd 7, næstu síðu) sem á sér stað í heilanum sjáum við að hægt er að tala um áhrif og orsakasambönd af ólíkum stærðargráðum, á ólíkum tímakvörðum og á mismunandi flækjustigum. Í mannsheilanum er að finna um 500 sérstæðar þyrpingar af taugafrumum, hópar þar sem frumurnar eru af ákveðinni gerð; í hverjum hópi eru að meðaltali 5 tegundir taugafrumna (sem oft mynda minni hópa) – það eru því um 2.500 mismunandi tegundir af taugafrumum í heilanum (Swanson 2001). Taugafrumur tengjast hver annarri á vissan hátt þannig að aðskildir hópar taugafrumna myndast, hver með sínar aðferðir til útreikninga og sérstæðu tengingar við aðra hópa. Hópar mynda svo stærri hópa, sem hver sér um vissa hluti hugsunar, svo sem sjón, heyrn, hreyfingu, tal, mál-skilning, jafnvægissskyn, tilfinningar og svo framvegis.

Samkvæmt víxlverkunarkenningunni verður meðvitund til við víxlverkun milli reikniaðgerða á mörgum flækjustigum (í mismunandi lögum í huganum/heilanum) og efnis. Þetta er stærðfræðileg eða algóritmísk efniskenning: Hún gengur út frá því að *sérstök algóritmísk ferli séu nauðsynleg* – en ekki

19 Víxlverkunarhugmyndin á sér fleiri en eina birtingarmynd og er ættuð frá fleiri en einum höfund; því er réttast að kalla þetta kenninga*ffjölskyldu*. Útfærsla kenningarinnar eins og hún birtist hér er þó alfarið mín eigin.

20 Orðin algrím, algóritmi og reikniaðgerð eiga hér við sama fyrirbærið, það er Turing-reiknanlegar aðgerðir (sjá til dæmis <http://plato.stanford.edu/entries/computability/>).

21 Hér er meðal annars átt við orsakasamhengi eins og þau sem Emmeche o.fl. (2000) kalla „medium downwards causation“ og „weak downwards causation“, það er að segja ekki er gert ráð fyrir þörf á nýjum efnisfyrirbærum til að útskýra meðvitund – hún er ófyrirséður eiginleiki (e. *emergent property*) efnisheimsins á ekki ósvipaðan hátt og líf.

Stærð	Tími	Lag
Líkamur/ líkami/ líkams- hlutar	Ár	Hegðun
	Mánuðir	Markmið
20 sm	Klukkustundir	Sálrænar einingar, gróf fúnksjón (minni, skynun, ..)
12 sm	Mínútur	Hugrænar einingar, fín fúnksjón (skammtímaminni, hljóðminni, ..)
8 sm	Desísekúndur	Gróf heilasvæði
5 sm	Sentísekúndur	Smærri heilasvæði
Sentímetri	100 msek	Taugafrumusúperklasar
100 mm	50 msek	Taugafrumuklasar
10 míkrómetrar	10 msek	Taugafruma
Míkrómetri	4 millísekúndur	Taugafrumutengingar, rafboð
20 nanómetrar	Millísekúnda	Taugaboðefni ○ ○ ○

Færri einingategundir
Einfaldari einingar
Einfaldari tengingar
Stærri stöðurými
Aukið flækjustig

Mynd 7. Fjöldmörg flækjustig heilans/hugans. Þótt hvert hækkandi stig sé í raun afleiðing margbrotinna orsakatengsla neðri stiga myndast oft ný orsakatengsl sem ekki var hægt að spá fyrir um út frá þeim fyrir neðan. Líf er gott dæmi: Það er bein afleiðing efnaskipta en er um leið umhverfi sem takmarkar hvaða efnaskipti eru möguleg. Slík ný orsakatengsl geta haft áhrif á bæði efri og neðri stig og flækir það málið enn frekar.

nægjanleg – forsenda fyrir meðvitund, og ferlin hafi jafnframt víxlverkun hvert við annað og við sérstaka uppbyggingu efnis. Efnið sjálft – efnasambönd, ákveðin frumefni, og/eða margbrotið skipulag þess – hefur áhrif á útfærslu og keyrslu algóritmanna. Víxlverkun efnis og algóritma hefur þau hliðaráhrif að meðvitund verður til – sá eiginleiki efnis að *það er eins og eitthvað að vera* það efni/kerfi.

Samkvæmt kenningunni er meðvitund (1) ferli; ferlið er (2) algóritmískt í eðli sínu, en þrátt fyrir að vera á mörgum flækjustigum eru algóritmarnir ekki nægilegir einir sér, heldur þurfa þeir (3) viss tengsl við efnið. Rökestyðja má þennan síðasta lið með ýmsum hætti. Ef nauðsynlegar forsendur með-

vitundar væru einungis bundnar við ákveðin reikniferli væri hægt að gæða nútímatölvur – eða í raun hvaða efnisfyrirbæri sem er – meðvitund, svo fremi sem það framkvæmdi reikni-aðgerðirnar. Slík niðurstaða hljómar ósannfærandi: Algrím er bara mynstur. Framkvæmd á algríminu, og framköllun mynstursins, virðist ekki líkleg sem nægjanleg forsenda meðvitundar. Gætum við skapað meðvitund í maurapúfu ef við þjálfuðum maurana í samskiptum sem herma nákvæmlega eftir því *mynstri* sem taugarnar framkvæma sín á milli í heil-anum? Ef okkur tækist að útbúa risastóran vefstól úr viði sem framkvæmdi mynstur hugsunar í formi teppis, myndi vefstóllinn þá hafa meðvitaða upplifun? Það þykir mér afar ólíklegt. Ef við hins vegar gerum ráð fyrir að eitthvað meira þurfi til en ákveðin reikniferli höfum við um leið viðurkennt að framsetning – það er efnisleg uppsetning og aðferðin sem notuð er við framkvæmd reikniferlanna – skipti máli.

Við munum ekki skilgreina tengsl algóritmanna við efnið að svo stöddu og látum okkur nægja að einfalda málið (í bili, að minnsta kosti) með því að útiloka að skammtafræðileg fyrirbæri séu nauðsynlegur þáttur í meðvitund. Við getum líka, að svo stöddu, kastað fram þeirri tilgátu að taugafrumur hafi eitthvað með meðvitund að gera og þá hugsanlega að þær seu hluti af nauðsynlegri formgerð reikni-aðgerðanna – það virðist að minnsta kosti ágætis byrjunarpunktur.

Eins og margar aðrar kenningar gerir víxlverkunarkenn-ingin ráð fyrir að meðvitund sé ekki nauðsynleg fyrir hugsun heldur aðeins að hún verði til við sérstakar aðstæður og sé (að mestu) hliðarafurð hugsanaferla, eins og náttúran hefur skapað þau. Eins og Gonzalez o.fl. (2004) orðuðu það:

Meðvitund er ekki nauðsynleg forsenda hegðunar og hugsunar heldur ófyrirséð afurð sjálfsskipulagðrar hegð-unar sem samþættir skynjun, ákvarðanatöku, tilfinningar og minni og býr til heildstætt upplýsingaflæði.²²

22 „Consciousness is not a pre-condition for action and thought, but rather an emergent product of self-organized action, which integrates sensory experiences, volition, emotion and memory, producing a unified informational field.“ (Gonzalez o.fl. 2004).

Helstu rökin fyrir þessari ályktun eru þau að við getum að öllum líkindum smíðað vélmanni sem haga sér algjörlega eins og manneskjur að öllu leyti – til dæmis úr nægilega fullkomnu líkani af heilanum byggðu á gervitauganetum – án þess að meðvitund upplifun þurfi nokkuð að koma þar við sögu. Þetta er að sjálfsgöðu tilgáta, en í meðferð margra góðra heimspekinga og vitvélavísindamanna hefur hún fengið sterkan meðbyr.²³ Þetta þýðir þó ekki að meðvitund skipti nauðsynlega engu máli í hugsun eins og hún fyrirfinnst í náttúrulegum vitvélum (heilum) – hún gæti haft áhrif á hugsun með sams konar gagnvirkni og víxlverkunarkenningin byggir á. En við gerum þó ráð fyrir að þetta sé raunvísindaleg spurning sem ekki sé hægt að svara að svo stöddu.²⁴

Er víxlverkunarkenningin öðruvísi en aðrar kenningar um meðvitund? Hún er það meðal annars vegna þess að hún segir reikniferli nauðsynlegan hluta meðvitundar. Margar kenningar, til að mynda sumar sem byggja á skammtafræði, gera það ekki (sjá t.d. Hameroff o.fl. 1996). Mikilvægara er þó að hún forðast smættunarvandamál þeirra kenninga sem einblína á skammtafræði eða önnur einstök náttúruleg fyrirbæri (veljið ykkar uppáhald) með því að vísa til víxlverkunar og samlegðaráhrifa hennar. Hún er líka sér á báti vegna þess að hún byggir beint á því gífurlega flækjustigi sem þarf til að hugsun eigi sér stað. Þetta er líklega sá þáttur sem skilur hana eindregnast frá öðrum kenningum. Samkvæmt víxlverkunarkenningunni er meðvitund því *ófyrirséð afurð* (e. *emergent property*):

1. ákveðins reikni-skipulags/formgerðar (forrits);
2. keyrð á kjötvél;
3. með sérstaka efnislega skipulagseiginleika;

²³ Sjá t.d. umfjöllun Chalmers (1996) um uppvakninga (e. *zombies*).

²⁴ Hér gæti verið um vissa mótsögn að ræða: Ef meðvitund hefur áhrif á hugsanaferlin er hún þá ekki nauðsynleg fyrir þá tegund hugsunar sem fyrirfinnst í mönnum? Jú, en það er hugsanlega hægt að nota annars konar kerfi til að fá fram sömu orsakaáhrif í vélmennum eins og meðvitund hefur í mönnum. Jafnframt er því ósvarað hversu mikil áhrif meðvitund hefur á hegðun og hugsun.

4. Þessir eiginleikar hafa sérstök tengsl við reikniaðgerðirnar;
5. á mismunandi flækjustigum.

Hún segir jafnframt að einhverjir efniseiginleikar heilans skipta máli um það hvort meðvituð upplifun birtist eða ekki,²⁵ en hversu margir og hverjir þeirra það eru þarf að skoða með vísindalegum tilraunum: Það er einmitt eðli kenninga að beina athyglinni í vissa átt, rannsóknnum inn á vissar brautir, hjálpa til við að spyrja spurninga og leiðbeina þannig við hönnun tilrauna. Eins og allar góðar vísindalegar kenningar er víxlverkunarkenningin sett fram þannig að hægt sé að afsanna hana með tilraunum. Þótt það gæti reyndar reynst erfitt með þeirri tækni sem við búum við í dag er möguleikinn þó ekki langt undan.

Tilraunir á kjötvélum og vitvélum

Þegar skoða á flókin fyrirbæri verður að vanda valið á aðferðafræðinni. Sú aðferð sem reynst hefur hvað best til að skilja margbrotin fyrirbæri er að smíða líkön og bera hegðun þeirra saman við hegðun fyrirbæranna sem þau eru líkön af. Það eru 100 milljarðar af taugafrumum í hverjum einasta mannsheila – hundraðfaldur fjöldi fjölmennustu þjóðar heims, Kínverja. Hver þessara taugafrumna hefur milli 1.000 og 10.000 tengingar við aðrar taugafrumur; fjöldi tenginga í heilanum er því að meðaltali kringum 600 þúsund milljarðar. Vegna náinna tengsla meðvitundar við mannsheilann/hugann er rétt að flokka hana líka sem flókið kerfi, hvort sem okkur líkar betur eða verr. Aðferðafræðin sem við veljum til að skoða meðvitund ætti því að vera mjög svipuð þeirri sem við notum til að skoða heilann. Ég ætla nú að lýsa aðferðafræði og tilraunaniði sem er sérstaklega hannað til að fást við meðvitund. Ég

²⁵ Kenningin er í raun tvískipt: Við gætum fjarlægt tilgátuna um tengsl algrímanna við efnið án þess að hreyfa við hugmyndinni um víxlverkun (hins vegar væri ekki hægt að fjarlægja kröfuna um reikniferli á mörgum flækjustigum – þau eru kjarninn í henni). Þetta er mál sem þyrfti meiri umfjöllun en rúmast í þessari grein.

tel að þetta tilraunasnið geti komið okkur talsvert áleiðis í skilningi á meðvitaðri upplifun og jafnframt að þessi nálgun sé besta leiðin til að nálgast nauðsynlegar og nægjanlegar forsendur þess að meðvitund komi fram.

Aðferðin byggir á munnlegri lýsingu á meðvitundum upplifunum. Ég geri ráð fyrir að hægt sé að hafa meðvitaða upplifun án þess að sá sem hana hefur geti lýst henni með orðum (mannskepnan er eina dýrið sem getur það²⁶). Hins vegar skiptir tungumálið miklu máli fyrir sýn okkar á meðvitundina. Í raun er talað mál önnur af aðeins tveimur ástæðum fyrir því að nokkur skyldi ætla að aðrir en hann sjálfur hafi meðvitaðar upplifanir (hin ástæðan er samræmið í uppbyggingu líf- og sálfræði einstaklinga af sömu tegund). Munnleg frásögn verður þess vegna fyrir valinu sem lykill í aðferðafræði framtíðarinnar við að kanna meðvitund.²⁷

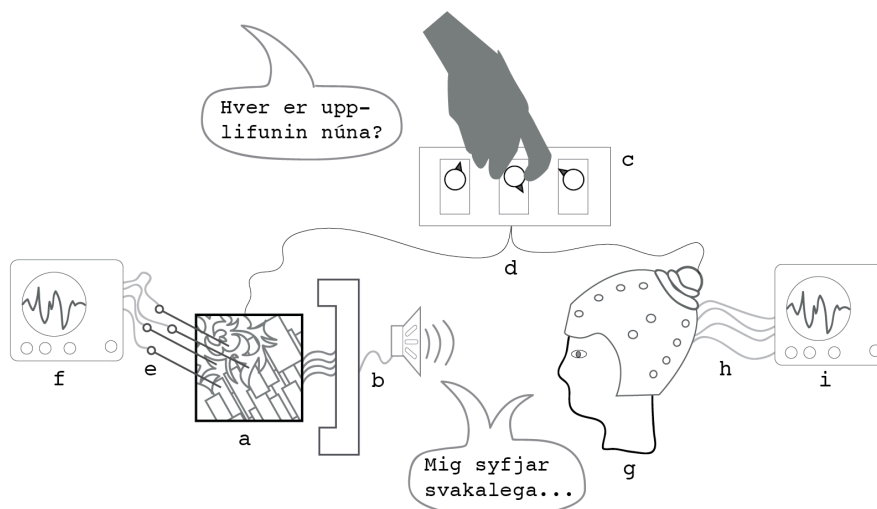
Við byrjum á að smíða kerfi sem getur lýst innri stöðu sinni munnlega. Kerfinu, svokallaðri *vélvitundarrás*, þarf að svipa til þess lífkerfis sem við vitum þegar að er meðvitað, það er mannsheila, bæði hvað varðar uppbyggingu og starfsemi.²⁸ Talkerfi vélvitundarrásarinnar verður að vera mjög fullkomið, byggt á nákvæmu líkani af mannlegum talhæfileikum, þar sem minnsti mismunur gæti kastað rýrð á niðurstöður rannsóknanna. Við þyrftum jafnframt að gæða kerfið langtíma- og skammtímaminni – hvort tveggja er nauðsynleg forsenda meðvitundar. Kerfið þarf að vera talsvert málglatt og búa yfir svipuðum málhæfileikum og skilningi á raunveruleikanum og við mennirnir – og hugsanlega einhverri almennri skynsemi.²⁹ (Hundaskynsemi myndi ekki duga, við gætum varla treyst hundi til að segja satt, og þar sem við þekkjum enga talandi hunda myndi það jafnframt gera okkur erfiðara fyrir að túlka niðurstöðurnar.) Samkvæt víxlverkunarkenn-

26 Rösk mætti færa fyrir því að vél sárðs hunds veiti sambærilegar upplýsingar um upplifun hans eins og lýsing með orðum, þótt að sjálfsögðu séu það ekki nærri eins haldbærar upplýsingar. Jafnframt eru þær bundnar við sársauka.

27 Slíkt hefur þegar verið gert, sjá til dæmis Libet o.fl. (1979).

28 Hér er átt við hugtök sem útleggjast á ensku *structural properties* og *functional properties*.

29 Við gerum þó ekki ráð fyrir að almenn skynsemi eða málgeta sé nauðsynleg til að efni öðlist meðvitaða upplifun.



Mynd 8. Tilraunaaðstaða til að rannsaka meðvitaða upplifun. Kjarni kerfisins er (a) vélvitundarrás sem hönnuð er til að framkalla meðvitaða upplifun, ásamt (b) rás sem gerir því kleift að tala, (c, d) stjórnþæki til að hafa áhrif á starfsemi þess, (e, f) mælitæki sem lýsa lífeðlisfræðilegri starfsemi þess, og (g) sjálfboðaliði með útbúnað til að (d) hafa áhrif á og (h, i) mæla heilastarfsemi hans.

ingunni verðum við jafnframt að huga vel að því hvernig þessi reikniferli – málgeta, þekking, minni, athygli – eru framsett; einhverjir, þó líklega ekki allir, reiknieiginleikar vélvitundarrásarinnar þurfa að vera framkvæmdir á sérstakan efnislegan hátt til að framkalla meðvitaða upplifun.

Fyrsta útgáfa tilraunasniðsins gengur út á að nota vélvitundarrásina eina og sér. Tilraunasniðið er sýnt á mynd 8. Við byrjum á að skoða vinstri hluta myndarinnar. Tækjabúnaðurinn byggir á (a) vélvitundarrás sem við ætlum að nota til að herma eftir lykilatriðum tengdum meðvitund í mannsheilanum/huganum og reyna að framkalla meðvitaða upplifun í, (b) tækni fyrir vélvitundarrásina til að lýsa innri stöðu sinni með orðum – og þar með meðvitaðri upplifun, ef hún á sér stað, (c, d) stjórnþækjabúnaði sem gerir vísindamanni kleift að hafa áhrif á starfsemi margbrotinna kerfa svo sem vélvitundarrása og heila, (e, f) mælitækjum sem lýsa efnislegri stöðu og hegðun vélvitundarrásarinnar.

Framkvæmd tilraunarinnar felst í því að hafa kerfisbundin áhrif á starfsemi vélvitundarrásarinnar og fylgjast gaumgæfi-

lega með (1) munnlegri frásögn þess og (2) efnislegri stöðu og virkni. Tilgátur um áhrif ýmissa breytinga, efnis, áreitis, og svo framvegis á upplifun kerfisins eru þá grundvöllur fyrir breytingum á vélvitundarrásinni. Við getum til dæmis skoðað áhrif deyfilyfja á upplifun sársauka, áfengis og ofskynjunarlyfja á sjón og heyrn, og svo framvegis. Þar sem slík áhrif eru efnisleg væri æskilegt að þeim væri náð fram með því að hafa vissan hluta vélvitundarrásarinnar lífrænan. Þá væri hægt að bera saman starfsemi tveggja kerfa þar sem annað væri að hluta framkvæmt af lífrænu efni en hitt væri allt framkvæmt af tölvulíkani.

Í næstu, og fullkomnari, útgáfu af tilraunasniðinu myndum við bæta við (g) manneskju sem tengd er sama stjórnþækjabúnaði og vélvitundarrásin (d) – sem gerir vísindamanninum kleift að hafa áhrif á heilastarfsemina, og (h, i) mælitækjum sem lýsa efnislegri stöðu og eiginleikum mannsheilans. Til að mæla innri stöðu heilans verðum við að útbúa nanóróbóta eða nanógriplur með rafskaut sem geta skriðið inn í dýpstu afkima heilans. Mikilvægt er að þær geti bæði ert og mælt starfsemi taugafrumna, á hárfínni rýmdar- og tímaupplausn. Ef rafskautin gætu einungis mælt starfsemi heilans yrðum við að finna aðrar aðferðir til að hafa áhrif á starfsemi hans, til dæmis að sprauta efnaupplausn inn á vel valda staði. Framkvæmd í þessari uppsetningu felst í að hafa sambærileg áhrif á vélvitundarrásina og mannsheilann (d) og bera saman (1) munnlega frásögn þeirra og (2) efnislega stöðu og starfsemi þeirra. Mismunur í munnlegri frásögn gefur til kynna mismun á meðvitaðri upplifun; samræmi í efnislegri stöðu og virkni hjálpar til við að þróa tilgátur um hvaða þættir skipta máli og við hvaða aðstæður meðvitund sprettur fram.

Við verðum að vera viss um að nanóróbótarnir skemmi ekki heila sjálfbodaliðans, til að halda okkur innan siðlegra marka. Sú krafa þarf þó ekki að hamla okkur of mikið ef við getum valið stærð og staðsetningu þeirra svæða þar sem taugaboð eru tímabundið stöðvuð eða þeim breytt, án varanlegra áhrifa.

Tilraunasniðið gerir okkur kleift að rannsaka raunverulegan

mannshuga, sem getur sagt okkur hvað hann upplifir hverju sinni, og bera það saman við viðbrögð vélvitundarrásarinnar á vísindalegan hátt. Þar sem við getum gert hvað sem er við vélvitundarrásina, án siðferðislegra vandkvæða, getum við notað kerfið til að afmarka, á mjög afgerandi hátt, uppbyggingu og skipulag hugsandi kerfis sem er bæði nauðsynlegt og nægjanlegt til að meðvituð upplifun verði til.

Hvort sem víxlverkunarkenningin eða einhver önnur nútímakenning um meðvitund er rétt er víst að því lengur sem rannsóknirnar með þessu tilraunasniði stæðu yfir því ljósara yrði það okkur hvaða þættir eru nauðsynlegir og nægjanlegir fyrir meðvitaða upplifun.

Til að færa okkur enn nær áreiðanlegri vitneskju um meðvitund getum við bætt við einni útgáfu í viðbót. Hún gengur út á að færa upplifanir úr einu kerfi í annað með utanaðkomandi breytingum á starfsemi þess. Í kvikmyndinni *Brainstorm* kynnti Douglas Trumbull okkur fyrir tæki sem gat afritað upplifanir og fært þær milli huga. Svipaða hugmynd væri hægt að nota í tilraunakerfinu okkar. Við gætum til dæmis fært ferli úr vélvitundarrásinni í mannsheila með því að lesa fyrst stöðu rásarinnar á einhverju tímabili þegar það upplifir eitthvað, til dæmis þegar það segist vera syfjað, og síðan sett heila sjálfbodaliðans í sambærilega stöðu, með hjálp nanógriplna, út frá kenningum okkar um hverjar séu nauðsynlegar og nægjanlegar forsendur fyrir meðvitaðri upplifun. Ef maðurinn segist verða syfjaður eru það rök fyrir því að vélvitundarrásin hafi verið að segja satt. Hægt væri að framkvæma sömu aðgerð í hina áttina, það er úr mannsheila í rás, og líka milli tveggja mannsheila. Jákvæðar niðurstöður úr slíkum tilraunum gæfu okkur enn haldbærari rök fyrir því að okkur hefði tekist að framkalla meðvitaðar upplifanir í vélvitundarrás.

Að útskýra alheiminn

Yrðu niðurstöður fengnar með þessu tilraunasniði einhvern tíma *sönnun* á því að vél hefði meðvitaða upplifun? Stutta

svarið er „nei“. Niðurstöðurnar verða aldrei 100% öruggar, því eins og við höfum þegar rætt þá er eina leiðin til að vera viss um að eitthvert fyrirbæri sé meðvitað *að vera það fyrirbæri*. Við verðum því að láta okkur nægja líkindi. Tilraunaaðstaðan sem hefur verið lýst myndi hins vegar gera okkur kleift að draga mjög áreiðanlegar ályktanir, með háum líkindum.

Þeim sem eru að vonast eftir útskýringu á því hvers vegna meðvituð upplifun er til þykir þetta kannski vonbrigði: Fjárfesting í rannsóknartækjum upp á tugi milljóna og áratuga rannsóknir en engin skýring á því *af hverju við höfum meðvitund*. Er það þess virði?

Vestræn vísindi hafa reynst öflugasta aðferðin til að útskýra hvernig heimurinn er settur saman og hvernig hann virkar. Þau hafa lýst orsakasamhengi í alheiminum sem forverar okkar gátu vart látið sig dreyma um að vita – nokkurn tíma. Kraftarnir sem halda saman alheiminum eru þeir sömu og halda gervihnetti á lofti. Hvernig? Jú, kraftar sem halda atómum saman gera mismunandi efni möguleg; úr efnunum verða til hlutir; aðdráttarafl milli hluta togar gervihnött að plánetu um leið og hann leitar út í geiminn á fljúgandi ferð. Hann getur ekki ákveðið hvort hann á að detta eða þjóta út í geiminn og hringsólar því í kringum Jörðina. Hvernig? Við getum lýst því mjög nákvæmlega með formúlu. En af hverju? Af hverju er raunveruleikinn þannig en ekki öðruvísi?

Það er alltaf hægt að finna stað í orsakasamhenginu þar sem það er óútskýrt, þar sem við getum ekki útskýrt hlutina vísindalega. Orsakakeðjan slitnar og við hættum að geta rakið hana áfram, að minnsta kosti í bili, og verðum að viðurkenna að við getum ekki svarað spurningunni um hvað valdi hverju. Ekki þarf til dæmis að spyrja oft „af hverju?“ áður en komið er að spurningu sem enginn getur svarað með vísindalegri vissu. Af hverju er aðdráttarafl í alheiminum? Af hverju er alheimurinn til? Af hverju er hann svona stór? Af hverju er til kraftur sem dregur atóm hvert að öðru?

Við nánari athugun kemur í ljós að vestræn vísindi hafa alltaf útskýrt heiminn með röð af „hvernig“-skýringum en í raun aldrei skýrt „af hverju“. Það er því óraunsætt að ætla að

við getum skýrt *af hverju* meðvituð upplifun er til *áður en* við skýrum *hvernig hún verður til*. Líkast til munum við skýra hvernig meðvitund sprettur fram löngu áður en við skýrum af hverju hún er til. Reyndar verðum við að vera undir það búin að við munum *aldrei* geta útskýrt af hverju hún er til, ekkert frekar en við getum skýrt það um alheiminn, aðdráttaraflíð eða efni.

Ef við sættum okkur ekki við þá kenningu að ekkert okkar hafi meðvitaðar upplifanir – sem afneitar okkar eigin upplifun af því að vera til – þá verðum við að skýra hvað það er sem gefur kjötvélinni upplifanir en neitar öðru efni um þær, meira að segja greindustu vélmennum. Hvað gæti gefið gervigreindu gervimennunum okkar vélvitund með mannlegri upplifun? Er það eitthvað sem er auðvelt að búa til inni á rannsóknarsetri, þegar við komumst að því hvað það er, eða verður erfitt að útbúa það jafnvel þegar við skiljum nákvæmlega hvernig það virkar? Þetta eru mikilvægar spurningar – og svörin enn mikilvægari. Tökum dæmi. Auðvelt er að skilja og skilgreina 0 gráður á Kelvin-kvarða en erfitt að búa til þær aðstæður á rannsóknarsetri (engum hefur tekist það, þótt nokkrir hafi komist mjög nálægt því); dæmi um eitthvað sem var erfitt að komast að og skilgreina en auðvelt að útbúa eru skammtafræðileg fyrirbæri (á síðasta áratug var fyrsta skammtafræðilega tölvan búin til³⁰). Svörin við þessum spurningum munu ákvarða hvort við getum búið til gervi-meðvitund inni á gervigreindarsetri einhvern tíma næstu 50 árin eða hvort slíkt heldur áfram að vera heimspekilegar vangaveltur næstu 500 árin.

Þakkir

Ég vil þakka Katrínu Elvarsdóttur, Hrafni Þorra Þórisssyni, Þóri S. Guðbergssyni, Þorsteini Þorsteinssyni, Þórdísi Helgadóttur og Steinari Erni Atlasyni fyrir góðar athugasemdir við

³⁰ Sem vísbending um hversu auðvelt það reyndist gerðu vísindamennirnir sér lítið fyrir og bjuggu til fyrstu skammtafræðilegu tölvuna í kaffibolla („Computing: Putting a Quantum Computer to Work in a Cup of Coffee“. *Science*, 275:5298, 17. janúar 1997, 307–309).

handritið að þessari grein. Verkefni sem hér er lýst eru að hluta til styrkt af Rannís og Marie Curie European Reintegration styrk 6. rannsóknaráætlunar Evrópusambandsins. Ritun þessarar greinar var að hluta til styrkt af Zentrum für interdisziplinäre Forschung, Universität Bielefeld, Þýskalandi, og af Háskólanum í Reykjavík.

Heimildir

- Bergenheim, M., H. Johansson, B. Granlund, J. Pedersen (1996). Experimental evidence for a sensory synchronization of sensory information to conscious experience. Í S.R. Hameroff, A.W. Kaszniak og A.C. Scott (ritstj.), *Toward a Science of Consciousness: The First Tucson Discussions and Debates*, 303–310. Cambridge, MA: MIT Press.
- Campbell, M.S. (1996). „An Enjoyable Game“: How HAL Plays Chess. Í D.G. Stork (ritstj.), *HAL's Legacy: 2001's Computer as Dream and Reality*. Cambridge, MA: MIT Press.
- Chalmers, D.J. (1994). A Computational Foundation for the Study of Cognition. Philosophy-Neuroscience-Psychology Technical Report 94-03, Washington University.
- Chalmers, D.J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Emmeche, C., S. Köppe, F. Stjernfelt (2000). Levels, Emergence and Three Versions of Downward Causation. Í P.B. Andersen, C. Emmeche, N.O. Finnemann og P.V. Christiansen (ritstj.), *Downward Causation. Minds, Bodies and Matter*, 13–34. Aarhus: Aarhus University Press.
- Gonzalez, E., M. Broens, P. Haselager (2004). Consciousness and Agency: The Importance of Self-Organized Action. *Networks*, **3–4**:103–113.
- Hackett, J., H. Kaas (2004). Auditory Cortex in Primates: Functional Subdivisions and Processing Streams. Í Michael S. Gazzaniga (ritstj.), *The Cognitive Neurosciences III: Third Edition*, 215–232. Cambridge, MA: MIT Press.
- Hameroff, S.R., A.W. Kaszniak, A.C. Scott (ritstj.), *Towards a Science of Consciousness: The First Tucson Discussions and Debates*, 301–310. Cambridge, MA: MIT Press.
- Klein, S.A. (2002). Libet's Temporal Anomalies: A Reassessment of the Data. *Consciousness and Cognition*, **11**:198–214.
- Libet, B., E.W. Wright Jr., B. Feinstein, D.K. Pearl (1979). Subjective

- referral of the timing for a conscious sensory experience. *Brain*, **102**:193–224.
- Metzinger, T. (2005). Précis: Being No One. *Psyche*, An Interdisciplinary Journal of Research on Consciousness, **11**(5):1–35.
- Muscoletta, N., P.P. Nyak, B. Pell, B.C. Williams (1998). Remote Agent: To Boldly Go Where No AI System Has Gone Before. *Artificial Intelligence*, **103**(1–2):5–47.
- Simon, H.A. (1969). *The Sciences of the Artificial*. Cambridge, Massachusetts: MIT Press.
- Swanson, L.W. (2001). Interactive Brain Maps and Atlases. Í M. A. Arbib og J. S. Grethe (ritstj.), *Computing the Brain*, 167–177. San Diego: Academic Press.