

Mitä tietokonemetafora tarkoittaa?

Pian toisen maailmansodan jälkeen digitaaliset tietokoneet alkoivat tunkeutua suuren yleisön tietoisuuteen. Lehdistö ryhtyi kutsumaan niitä jättimäisiksi, elektronisiksi aivoiksi. Tarvittiin jokin metafora kuvaamaan maallikolle, mitä nämä uudet laitteet oikeastaan ovat. Samoihin aikoihin tutkijat puolestaan alkoivat kertoa toisilleen, että itse asiassa aivot ovat eräänlaisia tietokoneita. Tämä rinnastus käytännössä valoi perustan koko modernille kognitiotieteelle. Kuitenkin nykyään myös monet kognitiotieteilijät katsovat, että aika on ajanut metaforan ohi. Toiset taas pitävät rinnastusta lähes käsitteellisenä totuutena. Mitä oikeastaan tarkoittaa, että aivot ovat tietokone, ja mitä tämän rinnastuksen kiistämällä halutaan sanoa?

Psykologi Robert Epsteinin essee ”Aivot vailla sisältöä” (”The Empty Brain”) edustaa toisaalta hyvin ajanmukaista kognitiotieteen filosofiaa, jota käsitellään tarkemmin tämän numeron toisessa tekstissä¹, toisaalta siinä on kaikuja lähes sadan vuoden takaisesta behaviorismista. Behavioristit olivat kiinnostuneita tarkastelemaan, miten organismi mukautuu ympäristöönsä oppimalla liittämään sille edullisia reaktioita ympäristön ärsykeisiin. 1990-luvulta lähtien tällaiset lähestymistavat ovat tehneet paluuta kognitiotieteeseen ja mielenfilosofiaan, ja mielenkiinto on osin siirtynyt jälleen organismien toiminnallisiin valmiuksiin ja niiden tapoihin olla vuorovaikutuksessa ympäristönsä kanssa. Tosin siinä missä kovan linjan behavioristit eivät mielellään puhuneet mielestä alkuunkaan, tämän vuosituhannen kognitiotieteilijät ja mielenfilosofit ovat yhä kiinnostuneempia siitä, miten juuri mielen sisältöjä ja prosesseja sekä muita mielenfilosofiaan liittyviä teemoja kuten kokemusta, merkitystä, minuutta ja tietoisuutta voitaisiin ymmärtää paremmin kiinnittämällä huomio mielen, ruumiin ja ympäristön jatkuvuuteen pikemmin kuin niiden eroihin. Behaviorismin kultakauden ja tämän uudenlaisen kognitiotieteen filosofian välissä, eli noin 1950-luvulta vuosituhannen taitteeseen, alaa hallitsi *tietokonemetafora*, jonka mukaan aivot ovat eräänlainen tietokone ja mielen toiminta tietojenkäsittelyä.

Epsteinin tekstin sanoma on selkeä: aivot eivät toimi kuten tietokone, ja aivojen rinnastaminen tietokoneeseen on aikaansaanut väärinkäsityksiä aivojen sekä mielen luonteesta ja sitä myötä huonoa tiedettä. Olen tästä pitkälti samaa mieltä, vaikka en aivan kaikkiin hänen johtopäätöksiinsä yhdykään. Tarkoitukseni tässä on selvittää, mitä Epsteinin (ja hänen kaltaistensa ajattelijoiden) kanta tarkoittaa eikä niinkään perustella sen paikan-

sapitävyyttä. Tietokonemetaforaa voi nimittäin tulkita monella tavalla: tutkimusta ohjaavana heuristisena rinnastuksena tai sitten kirjaimellisena, lähinnä tietokoneen ja laskennan määritelmiä koskevana filosofisena kantana². Nämä ovat hyvin erityyppisiä näkemyksiä, joiden pätevyys ei mitenkään itsestään selvästi riipu toisistaan. Lisäksi Epstein väittää, että tietokonemetafora on vain yksi mielen metafora muiden joukossa. Aiemmin mieltä on verrattu vaikka mihin kulloisenkin ajankohtaisen teknologian mukaan: hydrauliseen laitteistoon, kellokoneeseen, puhelinvaihteeseen ja nyt meidän aikamme tietokoneeseen. Käsittääkseni kuitenkin tietokonemetaforassa on piirteitä, jotka, tulkinnasta riippuen, erottavat sen muista mainituista konemetaforista ja jotka tekevät siitä filosofisesti mielenkiintoisen riippumatta sen lopullisesta pätevydestä.

IEpsteinin mukaan tietokonemetafora on aiheuttanut väärinkäsityksiä aivojen ja mielen luonteesta, koska aivot eivät yksinkertaisesti toimi samalla tavalla kuin digitaaliset tietokoneet. Tietokonemetaforan puolustajat taas yleensä katsovat, että aivoja (tai kenties ihmisiä) voidaan tiettyssä mielessä pitää kirjaimellisesti tietokoneina ja mielen prosesseja tietojenkäsittelynä riippumatta siitä, missä määrin ne muistuttavat varsinaisia digitaalisia tietokoneita ja niiden toimintaa. Tämä on tietenkin täysin totta. Laskettavuuden teoriassa nimittäin tietokoneiden, tai laskentaa suorittavien automaattien, määrittellään koostuvan kolmesta asiasta: muistista, joukosta konetiloja ja tilasiirtymäfunktiosta. Jälkimmäistä voi ajatella eräänlaisena mekanismina, joka määrää miten koneet siirtyvät tilasta toiseen riippuen siitä, mitä ne lukevat muististaan. Konetilat vastaavat koneen jonkinlaisia fyysisiä tiloja, mutta laskennan kannalta niillä ei oleteta olevan muita

ominaisuuksia kuin miten ne ohjaavat laskentaa eli informaation lukemista ja kirjoittamista. Tarkalleen ottaen muististakaan ei tarvitse olettaa mitään erityistä, vaan muistilaitteettomatkin järjestelmät täyttävät tietokoneen määritelmän, mikäli tietokoneella voidaan tarkoittaa mitä tahansa äärellistä tilasiirtymien muodostamaa automaattia, joka toteuttaa ainakin jonkinlaisen algoritmin.

Tässä mielessä tiiliskiveäkin voidaan pitää eräänlaisena tietokoneena. Tiiliskiven voidaan nimittäin katsoa toteuttavan automaatin, joka koostuu yhdestä tilasta ja tyhjästä tilasiirtymäfunktiosta. Tällainen systeemi laskee esimerkiksi päätösongelman tyhjälle kielelle tai toteuttaa algoritmin, joka tulostaa tyhjän merkkijonon. Jos tiiliskiven kylkeen piirretään luku ja luetaan se sitten siitä, voidaan systeemin katsoa laskevan identiteettifunktion $f(x) = x$. Suomeksi sanottuna tällaiset viritykset eivät siis tietenkään tee mitään, mutta näin kuvattuna tiiliskiven voidaan teknisesti ottaen katsoa toteuttavan hyvin määritellyn algoritmin.

Tarkastelemalla tiiliskiven sisäisiä, oletettavasti varsin monimutkaisia, alkeishiukkasten muodostamia rakenteita, on ehkä mahdollista löytää vastaavuuksia monimutkaisempienkin algoritmien kanssa. Jääköön tämä ajatuskulku nyt kuitenkin tarkemmin kehittelemättä. Asian ydin on, että tiiliskivet eivät tee mitään erityisen mielenkiintoista, mutta jos tiiliskivi tietystä mielessä on tietokone tai toteuttaa jonkun algoritmin, niin aivan varmasti sitten aivotkin. Lisäksi aivojen voi täydellä varmuudella tulkita tekevän paljon monimutkaisempia ja mielenkiintoisempia laskentoja. Keskeinen tietokonemetaforan kriitikeistä onkin, että jos se tulkitaan kirjaimellisesti ja jonkinlaisena käsitteellisenä väitteenä aivoista sekä samalla tietokoneen käsite määritellään tarpeeksi laveasti, niin väite, että aivot ovat eräänlaisia tietokoneita, on varmasti tosi, muttei välttämättä kovin mielenkiintoinen.³

Usein tietokonemetaforan katsotaan viittaavan *klassiseen komputationalismiin*. Epsteinin kritiikki näyttää kohdistuvan ensisijaisesti juuri tietokonemetaforan tähän tulkintaan. Teorian filosofisesti keskeisin anti on sen kyky tarjota naturalistinen selitys, miten rationaalisuus tai muu systemaattinen käsitteellinen ajattelu on ylipäätään mahdollista.⁴ Mikäli ihminen on fysiikan lakien alaisuudessa toimiva kausaalinen järjestelmä, miten on mahdollista, että samaan aikaan toimintamme näyttäisi olevan (ainakin suurin piirtein) myös jonkinlaisille rationaalisuuden tai mielekkyyden kriteereille perustuvaa? Miten fysiikalisen mekanismin on ylipäätään mahdollista noudattaa systemaattisesti jonkinlaisia käsitteellisiä päättelyperiaatteita? On ehkä helppo ymmärtää, miten yksinkertainenkin organismi voi reagoida valikoivasti ja mielekkäästi ympäristönsä välittömiin ärsyksiin, mutta on huomattavasti hankalampaa ymmärtää, miten mekaaninen koje voisi tehdä harkintaan perustuvia päätöksiä.

Merkittävä kognitiotiedettä ja tietokoneteknologiaa edeltävä askel asian ymmärtämisessä oli, kun 1800- ja 1900-lukujen taitteessa keksittiin, miten logiikka voidaan muuttaa formaaliksi kalkyyliksi, jonka avulla

loogisia päätelmiä voidaan tehdä soveltamalla muodollisia päättelysääntöjä. Näiden sääntöjen käyttö edellyttää vain symbolirimpusja muuntelevien sääntöjen mekaanista soveltamista välittämättä siitä, mitä nuo symbolit tai säännöt tarkoittavat. Osoittautui, että kaikki (ensimmäisen kertaluvun) logiikan päätelmät voidaan tehdä tällä tavoin mekaanisesti.

Aivojen ja tietokoneiden jonkinlainen toiminnallinen yhteys on sikäli ilmeinen, että tietokoneella voidaan korvata ”aivotyötä” eli tehdä päätelmiä ja laskelmia, hallinnoida ja etsiä tietoja, ja niin edelleen. Tietokoneteknologia kehitettiin tällaisia tarkoituksia varten aluksi automatisoimaan monimutkaisia laskutoimituksia, mutta koska tietokoneet kykenevät minkä tahansa formaalin kalkyylin suorittamiseen, oli ilmeistä, että tällainen koje voidaan valjastaa myös loogiseen päättelyyn. Näin järjestyksessä voidaan siis ainakin tässä mielessä mekanisoida, joten mielekäs naturalistinen selitys mielen toiminnasta saadaan ottamalla mielen ja tietokoneiden välinen vastaavuus vakavasti. Komputationalinen mielenteoria on pitkälti tästä huomiosta kumpuava hypoteesi, jonka mukaan mielen toiminta kirjaimellisesti koostuu eräänlaisista loogisista tai vastaavista säännöistä, joita noudattaen aivot systemaattisesti manipuloivat symbolisia mentaalisia representaatioita ja niistä muodostettuja rakenteita.

Tällaisia representaatioita ovat esimerkiksi sanoja vastaavat kielelliset käsitteet ja symbolirakenteita taas niistä muodostetut lauseet. Taustalla tässä on siis myös tietokonemetaforasta periaatteessa riippumaton versio *representationalismista*, jonka mukaan ajatukset tai muut mielen sisällöt ovat siinä mielessä symbolisia, että ne viittaavat kohteisiinsa samaan tapaan kuin sanojen tai lauseiden ajatellaan jotenkin viittaavan merkityksiinsä. Näiden symbolien katsotaan olevan tavalla tai toisella tallennettuna aivoihin ajatusten tai muun kognitiivisen sisällön perusyksiköinä. Epsteinin esseen otsikko ”Aivot vailla sisältöä” ei siis tarkoita, että aivot ovat fyysisesti tyhjää täynnä, vaan että ne eivät ole mielensisältöjen säiliöitä. Hänen komputationalisminvastaisuutensa kumpuaa pohjimmiltaan antirepresentationalismista. Vaikka representationalismi ei edellytä komputationalismia, riippuu jälkimmäinen kuitenkin edellisestä.⁵

Joka tapauksessa on syytä huomata, että toisin kuin Epstein väittää, komputationalismi ei sentään perustu yksinkertaiseen virhepäätelmään, jossa aivojen tai ihmisten ajatellaan olevan tietokoneita pelkästään, koska sekä ihmiset että tietokoneet voivat toimia älykkäästi. Teoria on pikemminkin ”Miten on mahdollista?” -kysymykseen vastaava empiirinen hypoteesi, joka kumpusi formaalin järjestyksen mekanisoinnin tutkimuksesta sekä tietyistä kielen, mielen ja järjen luonnetta koskevista filosofisista käsityksistä.

Komputationalismiin liittyy kaksi mielenkiintoista ajatusta: 1) *Metafora*, jonka mukaan aivot ovat eräänlainen tietokone ja mieli tietokoneohjelma. Tämä on varsin mielekäs ratkaisu mieli–ruumis-ongelmaan. Jos ohjelmien ja koneiden suhde on ongelmaton, niin

mielen ja aivojen (tai ruumiin) suhde on samalla tavalla ongelmaton. 2) *Metodologinen periaate*, että mielen toimintaa ja rakennetta voi tutkia sellaisenaan tarvitsematta tutkia aivoja, vastaavasti kuin ei tarvitse tutkia, miten tietokoneet toimivat ymmärtääksemme, miten tietyt tietokoneohjelmat toimivat. Mielen toiminnan katsotaan olevan omalaksinen suhteessa aivoihin vaikkakin aivo-toiminnasta tiukasti riippuvainen. Tämä käsitys kulkee nimellä *funktionalismi*. Tietenkään koneen ja sen toiminnan erottelu ei liity pelkästään tietokoneisiin, mutta koska fyysisesti hyvin erilaiset tietokoneet voivat suorittaa täsmälleen samoja ohjelmia, ja toisaalta hyvin erilaisia ohjelmia voidaan ajaa fyysisesti identtisissä koneissa, niin laitteiston ja ohjelmien rakennetta ja toimintaperiaatteita ei ole mielekäästä samastaa. Mikäli tämä pätee mieleen ja aivojen suhteeseen, voidaan niitä pitää erillisinä ilman substanssidualismia. Vastaavanlainen fyysisen koneen ja sen toiminnan erottelu ei ole mielekäs, tai ainakaan mitenkään ilmeinen, vaikkapa höyrykoneen tapauksessa, ja sikäli tietokone-metafora on filosofisesti erityislaatuinen verrattuna sitä edeltäviin mielen konemetaforiin.⁶

Toinen tietokone-metaforan erityislaatuinen piirre on tullutkin jo mainittua: tietokoneiden määritelmä on varsin abstrakti eikä edellytä tietynlaisia fyysisiä toteutusta, mistä syystä aivojen voidaan katsoa olevan kirjaimellisesti jonkinlaisia tietokoneita. Mutta mitä aivoista oikeastaan täytyy olettaa, että vertaus tietokoneisiin olisi jotenkin mielenkiintoinen? Että aivot toteuttavat Turing-täydellisen pitkäaikaisesta muistista koostuvan formaalin järjestelmän, jossa data ja sen käsittelyohjeet (eli ohjelmat) on tallennettu toisistaan erillisinä kombinatorisina tietorakenteina, ja että pitkäkestoisesta muistista nämä tiedot voidaan lukea työmuistiin, missä niitä voidaan käsitellä muistiin koodattujen rekursiivisten sääntöjen avulla (komputationalisti kutsuisi tätä ajatteluksi)? Näin voi luonnehtia nykyisten ohjelmoitavien digitaalisten tietokoneiden keskeisiä ominaisuuksia, ja suurin piirtein jotain tällaista myös tietokone-metaforaa kirjaimellisesti tulkinneet komputationalistit ovat aivoista tai mielen toiminnasta väittäneet. Kuitenkin tietomme aivoista viittaa lähinnä siihen, ettei tämä pidä suurin piirteinkään paikkaansa.

Tässä yhteydessä symbolirakenteiden rekursiivinen käsittely on erityisen tärkeä käsite, joka voidaan määrittellä monella tapaa. Tietojenkäsittelytieteessä sillä tyyppillisesti tarkoitetaan, että ohjelma voi kutsua toisia ohjelmia tai säännöt voivat viitata toisiin sääntöihin tai tietorakenteet toisiin tietorakenteisiin. Esimerkiksi pääohjelma voi koostua monista aliohjelmissä. Tarpeen mukaan pääohjelman suorittaminen keskeytetään, kun jokin sen tuloste käsitellään aliohjelmalla, ja sitä jatketaan kun aliohjelma on tehnyt tehtävänsä. Näin siis monimutkainen tehtävä voidaan pilkkoa osiin yksinkertaisemmiksi tietojenkäsittelytehtäviksi, ja yksinkertaisista ohjeista voidaan rekursiivisesti koostaa monimutkaisia hierarkisia rakenteita eli ohjelmia. Komputationalismin yhteydessä rekursio liittyy vastaavasti järjestykseen ja muuhun ongelmanratkontaan mutta myös käsitteisiin ja

ajatuksiin: käsitteitä säännönmukaisesti, kuten loogisesti, yhdistelemällä voidaan muodostaa monimutkaisia käsitteitä, joiden tulkinta taas perustuu niiden purkamiseen rakennetekijöihinsä.

Perustavimmanlaatuiset rakennetekijät ovat jonkinlaisia yksinkertaisia käsitteitä, jotka ymmärretään sellaisenaan ja joista voidaan rekursion avulla koostaa mielivaltaisen suuri joukko monimutkaisia ajatuksia. Komputationalismissa tällaisen mekanismin on ajateltu olevan oleellisen tärkeä selittämään kykyämme sekä ymmärtää että muodostaa periaatteessa mielivaltaisen monimutkaisia käsitteitä ja ajatuksia.⁷

Tässä näkyy komputationalismin vahva yhteys formaaliin logiikkaan ja analyttiseen kielifilosofiaan ja toisaalta taas ero behaviorismiin: rekursiivinen prosessointi nimittäin edellyttää sekä työmuistia että laskennan rakenteesta huolehtivaa kognitiivista kontrollijärjestelmää, eli verrattain monimutkaista organismin sisäistä kognitiivista prosessointia. Mikään ei kuitenkaan viittaa siihen, että aivomme prosessoisivat ajatuksia kuten tietokone muistiin tallennettuja bittijonoja. Jos siis tietokone-metaforan katsotaan tarkoittavan käsitteellisen teesin sijasta sitä, että tietokoneteknologia toimii aivotoininnan mallina, on väite kirjaimellisesti epätosi ja tutkimusheuristiikkana harhaanjohtava.

Itse asiassa edes varsinaiset tietokoneet eivät aivan tarkalleen toteuta yllä mainittua ehtoja, koska niiden muistin äärellisyydestä johtuen ne eivät ole Turing-täydellisiä. Toisin sanoen ne eivät voi laskea kaikkia periaatteessa laskettavia funktioita, koska yksinkertaisesti ne eivät voi tallentaa työmuistiin mielivaltaisen pitkiä syötteitä tai laskennan välituloksia. Tässä mielessä ne ovat teknisesti ottaen äärellisiä automaatteja. Tosin tämä huomio voidaan sivuuttaa, mikäli sallitaan epärealistinen oletus, että fyysisten tietokoneiden muisti voisi olla ääretön. Tällaiset sivuseikat lähinnä osoittavat vain, että kun asiasta ruvetaan keskustelemaan määritelmien ja periaatteiden tasolla, asiat muuttuvat hankaliksi ja yhteys tosimaailman ilmiöihin, kuten siihen miten tietokone-metafora voi auttaa tai haitata ymmärrystämme aivoista ja mielestä, voi hämärtyä.

II

Olipa Epsteinin pessimismi miten perusteltua tahansa, lienee myönnettävä, että tietokone-metaforan ja komputationalistisen teorian puitteissa on tehty paljon hyvääkin tutkimusta. Lisäksi on tietenkä selvää, että jos kerran kykenemme vaikkapa piirtämään muistimme varassa kuvan joskus aiemmin näkemästämme setelistä, niin jotainhan aivoihimme on tallentunut, tai jotain siellä ainakin on muuttunut sen seurauksena, että olemme nähneet näitä seteleitä jossain vaiheessa elämäämme⁸. Kuten Epstein koettaa selittää, tästä ei kuitenkaan välttämättä seuraa, että setelistä olisi tallennettu aivoihin jokin representaatio siinä mielessä, että setelin kuva tai siinä olevien visuaalisten osien rakenne tai ohjeet setelin piirtämiseen tai mitään vastaavaa olisi tallennettu jonnekin päin aivoja. Tämä on tietenkä mahdollista, mutta esi-

”Yksittäistä neuronია tai kyt- kentää tarkastelemalla ei voida mitenkään päätellä, mitä se, tai koko verkko ylipäätään, tekee.”

merkiksi keinotekoisissa neuroverkoissa, joita on käytetty laajalti kognitiivisten kykyjen mallintamiseen, muisti-
jäljet ovat neuronien välisten kytkentöjen vahvuuksien
muutoksia. Nämä muutokset mahdollistavat sen, että
verkko voi tulevaisuudessa reagoida tuttuun (tai tutun
kaltaiseen) syötteeseen jollain tarkoituksenmukaisella ta-
valla. Tämä ei edellytä, että verkossa itsessään olisi koo-
dattuna mitään syötteeseen liittyviä ohjeita tai represen-
taatioita.

Neuroverkkoja käytetään usein koneoppimisessa,
ja ne voidaan esimerkiksi opettaa jäsentämään lauseita
ilman, että ne sisältävät representaatioita kieliopista tai
lauseenjäsenistä⁹. Tämä on sikäli mielenkiintoista, että
juuri lauseenjäsentämistä on pidetty paraatiesimerkkinä
toimesta, joka edellyttää symbolirakenteiden rekursiivista
käsittelyä. Virkkeethän voivat koostua useista lauseista,
jotka taas koostuvat lauseenjäsenistä, joista voidaan kie-
lioppisääntöjen avulla tuottaa periaatteessa mielivaltaisen
monimutkaisia mutta ymmärrettäviä ilmauksia.¹⁰ Tosin
viimeistään käsitteitä venyttämällä voidaan katsoa, että
kaikki neuroverkot sisältävät ainakin jotain representaa-
tioita. Tämä riippuu myös verkosta, mutta usein mie-
lekkäämpää on ajatella, että verkot lähinnä koodaavat
hajautetusti syötteissä esiintyvien piirteiden tilastollisia
yhteyksiä ja oppivat liittämään tarkoituksenmukaisen
vasteen kuhunkin syötteeseen. Tätä *assosiatiivisuus* pää-
piirteisään tarkoittaa nykyisissä kognitiotieteissä. Siis
esimerkiksi jos verkko opetetaan tunnistamaan kasvo-
kuvia, se oppii reagoimaan valikoivasti tiettyihin, ehkä
hankalasti määriteltäviin, kasvojen piirteisiin ja niiden
yhdistelmiin. Näin verkko lopulta kykenee reagoimaan

valikoivasti tiettyihin ihmisiin esimerkiksi liittämällä
oikean nimen kuhunkin kasvokuvaan tai vaikka luokit-
telemaan henkilöiden tunnetiloja ilmeiden perusteella.¹¹
Tämä perustuu verkkojen löytämiin tilastollisiin korre-
laatioihin syötteissä esiintyvien piirteiden välillä, mutta
piirteiden representaatioita ei tarvitse tallettaa verkkoon.
Tämä liittyy mainittuun hajautettuun koodaukseen.

Tietty neuronin voi olla herkkä tietyille syöteava-
ruuden piirteille, mutta jokainen neuronin ja kytkentä
osallistuu kaiken syötteen käsittelyyn. Kaikki verkon
neuronit ovat yksittäin tarkasteltuna yleensä samanlaisia,
ja miten ne osallistuvat verkon laskentaan riippuu syste-
emin kokonaisuudesta. Yksittäistä neuronin tai kyt-
kentää tarkastelemalla ei voida mitenkään päätellä, mitä
se, tai koko verkko ylipäätään, tekee. Mikäli neuroverkot
sisältävät representaatioita tai sääntöjä, ne eivät ainakaan
ole koodattuna verkosta erikseen luettavina symbolisina
tietorakenteina vaan kytkentöjen muodostamana holis-
tisena toiminnallisena rakenteena. Verkoissa siis ei ole sa-
manlaista prosessoinnin, datan ja tallennettujen ohjeiden
välistä erottelua kuin mikä luonnehtii ohjelmoitavia digi-
taalisia tietokoneita.

Laskentaahan se on neurolaskentakin, mutta peri-
aatteessa lienee mahdollista laatia neuroverkon ohjaama
systeemi, joka kykenee jäljentämään vaikka näkemänsä
setelin ilman, että verkkoon on varsinaisesti tallennettu
mitään sieltä suoraan luettavissa olevaa muistijälkeä se-
telistä samaan tapaan kuin verkot voidaan opettaa jäsen-
tämään lauseita ilman, että ne sisältävät representaatioita
kieliopista. Tosin on olemassa viitteitä, että jokin tietty
muistikuva voidaan ehkä palauttaa mieleen stimuloimalla

oikein valittua pientä aivosolujen ryhmää, ja että poistamalla tuo ryhmä voidaan estää kyseisen muistikuvan palautuminen¹². Yllä sanotun perusteella tästä ei kuitenkaan seuraa, että tuo muistikuva olisi millään tavalla tallennettu kyseisiin soluihin. Muistijälkeä voi ajatella kognitiivisen prosessin *lopputuloksena*, jonka tuottamiseen kyseiset neuronit osallistuvat. Ylipäätään ihmisen muistia voi usein olla hedelmällisempää tarkastella toiminnallisena kykyä eikä faktojen, ohjeiden ja menneiden tapahtumien säilymisenä. Tämä on verrattain selvää taitojen kuten pyörällä ajamisen taustalla olevan proseduraalisen muistin tapauksessa, mutta muistin konstruktiiivisesta luonteesta johtuen sama vaikuttaisi pätevä myös episodiseen eli tapahtumamuistiin. Eli muistikuvia ei haeta jostain säilöstä vaan ne ainakin osittain luodaan tiettyissä tilanteissa tiettyjä tarpeita varten. Tällainen kykyjen painottaminen valmiiksi tallennetun tiedon kustannuksella on usein antirepresentationalististen mielenteorioiden keskeinen lähtökohta.

Tapahtumamuistin tapauksessa näiden näkökulmien ero tulee esiin esimerkiksi suhtautumisessa muistikuvamme muuttumiseen ajan kuluessa. Säilömetaforan mukaan tämä on tallentuneen tiedon vääristymistä, joka viestii muistin heikkoudesta. Toisaalta, jos mieleen palauttaminen on eräänlainen muuhun toimintaan kytkeytyvä kyky – esimerkiksi, että muistelu on osittain tilannesidonnaista päättelyä – on luontevaa ajatella, että mielikuvamme muokkautuvat tarkoituksenmukaisesti sen mukaan, miten niitä käytämme, samoin kuin muutkin kognitiiviset kykymme.¹³

Antirepresentationalistisen näkökulman eroista komputationalismiin Epstein tarjoaa esimerkkinä kahta erilaista selitystä siitä, miten pesäpallon pelaaja kykenee ottamaan pallosta kopin. Komputationalistinen teoria edellyttää, että pelaaja muodostaa jonkinlaisen käsityksen lyönnin voimakkuudesta ja kulmasta, ja laskee tämän perusteella pallon todennäköisimmän lentoradan. Tätä mallia hän käyttää ohjatakseen raajojansa siten, että pallo lopulta löytyy pelaajan räpylästä. Antirepresentationalistinen perspektiivi tarjoaa yksinkertaisemman, havainnon ja toiminnan dynaamiseen vuorovaikutukseen, perustuvan selonteon: Pelaajan tarvitsee vain liikkua siten, että pallon havaittu sijainti näyttää pysyvän vakiokulmassa suhteessa lähtöpaikkaansa ja havaittuun ympäristöön. Ohjaamalla liikettään suoraan havainnossa olevien invarianssien perusteella hän lopulta päätyy samaan paikkaan pallon kanssa. Epsteinin mukaan tämä menetelmä on hämmästyttävän yksinkertainen eikä sisällä representatioita, laskentaa tai algoritmeja.

Toisaalta tässä koetellaan algoritmin käsitteellisiä rajoja. Esimerkiksi tietojenkäsittelytieteilijä Jeffrey Shallit ojensi melko poleemisissa blogikirjoituksessaan Epsteinin siitä, ettei tämä ole ymmärtänyt pelaajan itse asiassa toteuttavan eräänlaista algoritmia yllä olevassa kuvauksessa¹⁴. Ei käy kiistäminen, että palloa kiinni ottaessaan ihminen tai hänen aivonsa tekevät jotain sellaista, jota voidaan kuvata algoritmeilla tai algoritmina. Eri asia on, miten tämä tosiasia sinänsä auttaa ymmärtämään kyseistä

toimintaa. Borssikeiton reseptiä voi pitää eräänlaisena algoritmina ja borssin keittämistä algoritmin toteuttamisena, mutta kokkaamisen komputationaalinen teoria ei varmaankaan lisäisi ymmärrystämme ruoanlaitosta ja siihen liittyvistä kognitiivisista prosesseista kovin merkittävästi. Jos algoritmin katsotaan tarkoittavan mitä tahansa aikaansaamismenetelmää, katkeaa yhteys algoritmin matemaattiseen määritelmään. Vielä oleellisempaa on, että jos käyttäytyminen lasketaan algoritmin suorittamiseksi, jää epäselväksi, miten tämä enää liittyy siihen, että aivot ovat tietokoneita. Aivot saattavat käsitellä tietoa tai olla käsittelemättä, mutta usein ongelmien ratkaiseminen tai sääntöjen seuraaminen ei tapahdu ainakaan pelkästään aivojen suorittamana laskentana vaan organismin ja ympäristön välisessä vuorovaikutuksessa, missä osa ratkaisusta ja prosessoinnista tapahtuu konkreettisenä toimintana ja osin ympäristöön hajautettuna. Aktiivisen havainnoinnin ja autonomisen robotiikan tutkimuksesta tiedetään, että esimerkiksi monia hahmونتunnistukseen ja toiminnanohjaukseen liittyviä tietojenkäsittelyongelmia voidaan helpottaa huomattavasti, kun ongelmaa suorittava järjestelmä voi itse aktiivisesti etsiä informaatiota ja manipuloida ympäristöään.¹⁵

Itse asiassa joissain tapauksissa yksinkertaiset robotit voivat ratkaista ongelmia, joita niiden ”kognitiivinen järjestelmä” ei periaatteessakaan voi ratkaista¹⁶. Tässä ei ole mitään mystistä. Ongelman formaali määritelmä muuttuu, kun huomioidaan robotin kyvyt ja sen toiminta omassa ympäristössään. Tietenkin näiden laitteiden ”aivot” ovat yleensä tietokoneita ja tällaiset tutkimukset tehdään tarkoituksella niin, että robotin toimintaa voidaan tarkastella algoritmin suorittamisena. Asian ydin on kuitenkin siinä, että tällöin algoritmi ei sijaitse robotin ”päässä” ja algoritmin suorittamisen kannalta keskeiset muuttujat on osin hajautettu ympäristöön ja robotin ”ruumiiseen”. Näin siis vastaavasti siitä, että organismin vuorovaikutusta ympäristönsä kanssa voidaan tarkastella algoritmina, ei seuraa, että tuo algoritmi lasketaan organismin aivoissa. Yhtä lailla höyrymoottorin keskipakosäätimen varsien kulmankin voidaan katsoa representoivan moottorin kierrosnopeutta ja säätimen kontrolloivan höyryputken venttiiliä tämän informaation perusteella. Kuitenkin sen sanominen, että tällainen järjestelmä tiettyssä mielessä on (analoginen) tietokone, joka suorittaa eräänlaista algoritmia, on vain ymmärrystämme haittaava tapa puhua keskipakosäätimen ja höyrykoneen toiminnasta.¹⁷ Epstein pyrkii sanomaan, että tietokonepuhe aivoista on vastaavasti harhaanjohtavaa. Tosin väitteen ”aivot eivät ole tietokone” voidaan katsoa myös olevan epätosi, mutta lähinnä sellaisessa mielessä, joka kertoo ehkä enemmänkin jotain tietokoneen käsitteestä kuin aivoista. Shallit tuo tämän esiin seuraavassa lainauksessa:

[Komputationalismin taustalla oleva] todellinen päättely menee jotakuinkin näin: ihmiset voivat käsitellä tietoa (tiedämme tämän, koska ihmiset kykenevät esimerkiksi ymmärtämään ja kertomaan kokonaislukuja). Ihmiset kyke-

nevät tallentamaan tietoa (tiedämme tämän, koska kykenen muistamaan sosiaaliturvatunnukseni ja syntymäpäiväni). Systemit, jotka sekä tallentavat että käsittelevät tietoa ovat tietokoneita.

On syytä huomata, että tässä puhutaan ihmisistä eikä aivoista. Ihmiset voivat syödä hampurilaisia, mutta seuraako tästä, että aivot voivat syödä hampurilaisia¹⁸? Tietenkään ihminen ei voi suorittaa yhteenlaskuja ilman aivoja, muttei tosin tehdä mitään muutakaan. Kun laskemme sormillamme, onko laskeminen (pelkästään) informaation prosessointia aivoissa? Entä kun tämä tapahtuu paperilla? Kun ihminen oppii liittämään tiettyihin lukuihin tietyn toimituksen (kuten vaikka, että $2+3=5$), edellyttävätkö tämä, että hänen aivoihinsa on tallentunut lukujen representaatioita ja ohje miten yhteenlaskuja suoritetaan?

On selvää, että ihminen voi käsitellä informaatiota, ja kykymme suorittaa yhteenlaskuja on tästä hyvä esimerkki. Yksinkertaisissa laskutoimituksissa tämä ei kuitenkaan edellytä lukujen tai yhteenlaskualgoritmin tallentamista aivoihin, vaan tämä voidaan myös tehdä assosiativisesti oppimalla oikea vaste tiettyihin tehtäviin kuten neuroverkkojen tapauksessa. Monimutkaisissa tapauksissa laskut yleensä tehdään paperilla pilkkomalla laskutehtävä yksinkertaisiin komponentteihin. Tällöin ihminen todella soveltaa allekkainlaskualgoritmia ja hänen toimintaansa voi selittää ja tarkastella algoritmin toteuttamisen perspektiivistä, siis informaation prosessoinnin käsittein. Myönnytykseni eivät jää tähän: Kykenemme myös sisäistämään tällaisia algoritmeja ilman paperin tai sormien käyttöä. Kun suoritamme niitä mielessämme, algoritmin toteuttavat prosessit mitä ilmeisimmin tapahtuvat aivoissamme. Tällöin aivomme siis itse asiassa ikään kuin simuloivat eräänlaista tietokonetta, ja koska tietokoneen simulaatiolla ja implementaatiolla ei oikeastaan ole mitään eroa, tällöin aivomme siis toimivat kirjaimellisesti laskutehtäviä tekevänä tietokoneena!

Kykymme toteuttaa algoritmeja mielessämme on kuitenkin erittäin rajallinen, ja jo kolminumeroisten lukujen kertolasku ilman paperia ja kynää on useimmille melko haastavaa. Joka tapauksessa rajallisuudestaan huolimatta tämä kyky näyttäisi osoittavan, että ihmisen kognitio sisältää työmuistin ja kontrollijärjestelmiä, joiden avulla voimme tietyissä tilanteissa käsitellä informaatiota ja toimia kuten tietokone. Toisaalta taas rajallisuudestaan johtuen tämä kyky ei kuitenkaan riitä selittämään useimpia tärkeitä kognitiivisia kykyjämme, kuten arkista järjenkäyttöä ja päätöksentekoa, eikä tällainen tietoisesti kontrolloitu symbolinen kalkyyli luultavasti myöskään muistuta niitä. Tästä syystä tämä tosiasia ei myöskään tarjoa kovin hyvää yleistä mallia mielen tai aivojen toiminnasta, vaikka algoritmien suorittaminen selvästi on kyky, jonka aivomme mahdollistavat.¹⁹

On syytä silti muistaa, että kun laskemme allekkain, otamme kiinni palloa, keitämme borssia tai koitamme jäljentää seteliä, aivomme eivät tee mitään maagista. Kaikki kognitiotieteilijät uskovat, että nämä toimet edellyttävät ai-

votoimintaa ja että aivot toimivat tavalla tai toisella mekaanisesti. Computationalismin kannalta keskeinen kysymys kuuluu, onko tämä aivotoiminta jonkinlaista tietojenkäsittelyä vai ei. Jotkin ei-klassiset komputationaaliset lähestymistavat, kuten neurolaskenta, voivat hyvinkin osoittautua oikeanlaisiksi kuvauksiksi mielen ja aivojen toiminnasta. Komputationaalinen neurotiede, joka analysoi aivojen toimintaa algoritmisin menetelmin, vaikuttaa olevan terve ja hedelmällinen tutkimusohjelma. Nämä teoriat ovat myös immuuneja Epsteinin kritiikille, koska ne eivät edellytä perinteistä representationalismia tai vahvaa yhteyttä digitaalisten tietokoneiden ja aivojen välillä.

Toisaalta komputationaalisessa neurotieteessä käytetyt neurolaskentamallit muistuttavat huomattavasti enemmän esimerkiksi signaalinkäsittelyä kuin perinteisiä laskennan malleja. Edellä olen koettanut selittää, että neurolaskennassa nimenomaan menetetään yhteys aivojen ja tietotekniikan sekä tietokoneiden taustalla olevan laskennan mallin välillä. Tämä voi säästää komputationalismin, mutta tällöin jää kovin epäselväksi, mitä työtä varsinainen *tietokonemetafora* enää oikeastaan tekee. Mitä tahansa fysikaalista järjestelmää voidaan joka tapauksessa mallintaa laskennallisesti. Esimerkiksi ihmisen immuunijärjestelmä on hyvin monimutkainen systeemi, jota voidaan kuvata algoritmisesti ja jonka formaaleja ominaisuuksia on käytetty tekoälysovelluksissa.²⁰ Tästä huolimatta emme ehkä sanoisi, että immuunijärjestelmä on tietokone, koska immuunijärjestelmän tehtävä ei ole käsitellä tietoa vaan suojella organismeja patogeeneiltä. Vastaavasti aivojen tehtävä ei niinkään ole käsitellä tietoa kuin mahdollistaa tarkoituksenmukainen käyttäytyminen monimutkaisessa ympäristössä. Jos joka tapauksessa tiedämme, ettei tämä tapahdu tietokoneen toimintaa muistuttavilla tavoilla, miksi verrata aivoja tietokoneeseen? Tutkijalle tosin voi olla käytännössä varsin hyödyllistä kuvata aivotoimintaa tietojenkäsittelytieteiden termein ja signaalinkäsittelyn menetelmin. Tällaisen metodologian kannalta on kuitenkin melko epäoleellista, ovatko aivot kirjaimellisesti tietokoneita vaiko eivät. Tietokonemetafora voidaan tällöin ehkä säilyttää nimenomaan metaforana, mutta sen pätevyys tässä yhteydessä koskee lähinnä tietokoneen käsitteen joustavuutta ja tapaamme tehdä tiedettä, ei niinkään aivojen luonnetta. Näin metafora vaikuttaisi olevan parhaimmillaankin melko irrelevantti ja pahimmillaan asioita sotkeva.

III

Tehdäänpä pieni väliyhteenveto edellä sanotusta:

1. Käsittelevätkö ihmiset tietoa? Kyllä.
2. Käsittelevätkö ihmiset ainakin joskus symbolirakenteita? Kyllä.
3. Onko tällainen tietojenkäsittely ainakin joissain tapauksissa algoritmista? Kyllä.
4. Ovatko aivot tietokoneita ainakin jossain mielessä? Kyllä.
5. Auttavatko kohtien 1.–4. opetukset meitä ymmärtämään jotain oleellista mielen ja aivojen toiminnasta? Ehkä, mutta enemmänkin ehkä ei.

”Se, että aivojen ja mielen toimintaa ei kannata yleisellä tasolla tarkastella tietokoneiden kanssa samanlaisena, ei tarkoita, etteivätkö ihmiset voisi systemaattisesti prosessoida symbolista informaatiota sääntöjen avulla.”

Se, että aivot ovat jossain mielessä tietokoneita, ei tarkoita, että mielen ja aivojen toimintaa olisi kovin hedelmällistä mieltää tietojenkäsittelytieteestä, formaalista logiikasta ja tietotekniikasta perittyjen käsitteiden avulla. Erityisesti klassinen komputationalistinen mielenteoria, jonka mukaan psykologiset teoriat tulisi muodostaa jonkinlaisen loogisen kalkyylin mukaan, on enemmän ihmismielen toiminnan ymmärtämistä haittaava kuin hyödyttävä ajatus. Vielä pahemmin harhaanjohtavaa on aivojen vertaaminen tietokoneeseen, mikäli tämän tulkitaan tarkoittavan, että konkreettisten tietokoneiden toimintaperiaatteet ovat siinä määrin samankaltaisia aivotoinnin kanssa, että vertaaminen tietokoneeseen auttaa meitä muodostamaan hyviä teorioita aivotoinnasta. Ydinongelma tässä on, että ”tietokoneen” käsitteellä on kaksoismerkitys, joka helposti unohtuu: yhdessä mielessä tietokoneella tarkoitetaan tietynlaista konkreettista teknologiaa ja toisessa mielessä taas tietynlaista hyvin abstraktia matemaattisten mallien joukkoa.

Silti kohdat 2. ja 3. on tärkeää pitää mielessä. Se, että aivojen ja mielen toimintaa ei kannata yleisellä tasolla tarkastella tietokoneiden ja tietynlaisten tietokoneohjelmien kanssa samanlaisena, ei tarkoita, etteivätkö ihmiset voisi systemaattisesti prosessoida symbolista informaatiota sääntöjen avulla. Tämä kyky on myös arvaustenkin erittäin tärkeä ja vieläpä kasvavissa määrin, kun ympäristömme haasteet kouluissa ja työpaikoilla alkavat yhä enemmän muistuttaa tietojenkäsittelytehtäviä. Kyseisen kyvyn luonteen ja rajojen tutkiminen on tärkeää monestakin syystä, ja sen taustalla olevien mekanismien ymmärtäminen on kognitiotieteen kannalta erittäin mie-

lenkiintoista. Kuitenkin tämän tyyppinen algoritminen toiminta on melko pieni osa ihmisen käyttäytymistä, ja mielen luonteen selvittämiseksi tärkeämpää on tarkastella, miten aivot ja mieli yleensä mahdollistavat jokapäiväiset askareemme.

Vaikka aivojen ja tietokoneiden samastaminen on ongelmallista, aivojen ja mielen suhdetta voi silti olla valaisevaa ajatella analogisena tietokoneiden ja tietokoneohjelmien suhteen kanssa. Ensinnäkin tästä saadaan varsin järkevä oloinen ratkaisu mieli–ruumis-ongelmaan: metafora auttaa meitä ajattelemaan mieltä prosessina eikä oliona. Toiseksi tämä auttaa ymmärtämään, miten mieli ei myöskään ole sama asia kuin aivot, vaikka sekä sen olemassaolo että toiminta riippuu aivoista. Mielen prosesseja on syytä ajatella abstraktimmalla tasolla kuin aivoprosesseina, ja tällöin on usein hyödyllistä puhua representaatioista ja ehkä myös informaationkäsittelystä, kunhan tutkija on tarkkana, etteivät metaforat johda häntä harhaan.

Yllä mainitun kaltainen tietokonemetaforan tulkinta ei kuitenkaan ole täysin ongelmaton. Jos mielen toiminta on eräänlaista laskentaa ja kaikki periaatteessa mahdolliset laskennat voidaan suorittaa tietokoneilla, tästähän näyttäisi seuraavan, että tietokoneella voi olla mieli samalla tavalla kuin ihmisellä. Kuitenkin ihmismielen olemusta tuntuisi luonnehtivan jonkinlainen merkityksellisyys, eli ajatukset ja muut mielenliikkeet ovat sisällöllisiä organismille itselleen. On epäselvää, onko tietokoneen representaatioilla merkitystä tietokoneelle itselleen ja voiko niillä periaatteessakaan olla. Erityisesti voiko paikalleen kahlitulla monimutkaisella laskuko-

”Useimmat sekä kognitiivisen psykologian tutkimat että meitä arjessa kiinnostavat mielen ilmiöt ovat luonteeltaan toiminnallisia, maailmaan suuntautuvia ja toteutuvat vuorovaikutuksessa ympäristön kanssa.”

neella olla *inhimillisellä tavalla* sisällöllisiä mielentiloja ja -prosesseja?

En tiedä vastausta tähän kysymykseen, muttei sillä taida lopulta olla niin väliäkään. Asiasta oli jokunen vuosikymmen sitten paljonkin keskustelua, ja lopputulos oli nähdäkseni lähinnä sen tajuaminen, että kysymyksenasettelussa on ongelma. Riippumatta siitä, onko mieli eräänlainen tietojenkäsittelyprosessi vaiko ei, ihmiset eivät ole omaan sisäiseen todellisuuteensa sulkeutuneita ja paikalleen sidottuja koneita vaan aktiivisia, avoimessa maailmassa toimivia autonomisia organismeja. Mielen tietokone-metafora ohjaa liian helposti ajattelemaan mieltä samalla tavalla puhtaasti aivoissa tapahtuvana prosessina kuin algoritmien laskenta on puhtaasti tietokoneessa tapahtuva prosessi. Kuitenkin useimmat sekä kognitiivisen psykologian tutkimat että meitä arjessa kiinnostavat mielen ilmiöt ovat luonteeltaan toiminnallisia, maailmaan suuntautuvia ja toteutuvat vuorovaikutuksessa ympäristön kanssa.

Ei ehkä ole erityisen mielekästä miettiä, miten mielentilat tai representaatiojärjestelmät sinänsä voisivat olla merkityksellisiä. Meille merkityksellistä on ensisijaisesti meitä ympäröivät asiat ja niiden kanssa toimiminen. Mitä mielentilat ja mielen prosessit lienevätkään, mitä luultavimmin niiden sisältö, ja niihin liittyvä merkityksellisyys, palautuu mielen, ruumiin ja ympäristön toiminnalliseen yhteyteen, eli organismin ja ympäristön vuorovaikutukseen. Sen lisäksi, että toiminnan ohella havaitseminen on lähtökohtaisesti aktiivista, on myös ajattelu pohjimmiltaan maailmaan suuntautuvaa toi-

mintaa. Toki voimme ajatella myös puhtaan muodollisia asioita, esimerkiksi kehitellä itse jonkinlaisia merkki-pelejä, joita pelaamme ajatuksissamme, mutta tämä vastanee suurin piirtein juuri sitä, mitä tietokoneet tekevät; tosin sillä erotuksella, että edes tuollaisina hetkinä emme ole täysin irti käytännöllisestä inhimillisestä todellisuudesta. Lähtökohtaisesti ajattelu syntyy toiminnasta eikä toisinpäin, mistä syystä mielensisältöjä kannattaa ensisijaisesti ajatella organismin eikä representaatiojärjestelmän ominaisuutena.

Edellisen kaltaiset korulauseet ovat toki yhtä vähän tosia tai epätosia kuin aivojen vertaaminen tietokoneeseen, mutta toivottavasti hieman valaisevampia. Idea tällaisissa maalailuissa on tiivistää joitakin lähtökohtia ymmärtää mieltä siten, että osaisimme esittää oikeita kysymyksiä oikeilla tavoilla ja hahmottaa edes alustavasti, minkä tyyppisiä vastauksia ylipäätään haluamme. Käsityksemme muuttuvat ja täsmentyvät tutkimuksen edetessä, ja tästä syystä tietokone-metaforakaan ei ole johtanut tutkijoita täysin harhaan. Reilut viisi vuosikymmentä komputationalistista kognitiotiedettä ja mielenfilosofiaa ovat muuttaneet huomattavasti sitä, miten edellisen kappaleen kaltaiset argumentit ylipäätään ymmärretään. Lisäksi kaiken edellä sanotun jälkeenkin informaationkäsittelyyn perustuvalla lähestymistavalla on paikkansa kognitiotieteissä. Aivoja ja mieltä tutkittaessa puhe informaationprosessoinnista on kaikkein hedelmällisintä, kun tutkitaan jotain muuta kuin päättelyä, päätöksentekoa, uskomusten muodostamista ja vastaavia kokonaisen ihmisen toimintaa kuvaavia kognitiivisia prosesseja,

jotka sijoittuvat niin sanotusti persoonatasolle. Esimerkiksi kun tarkastellaan näköaistimuksen muodostumista, on informaationprosessoinnista puhuminen usein hyödyllistä. Yleensä tällöin tosin on selvää, että kyse on hermostollisten mekanismien abstraktioista ja käytännöllisestä terminologisesta valinnasta ja on usein samantekevää, onko tarkasteltu mekanismi jossain mielessä algoritmin implementaatio tai tietokone tai sellaisen osa vaiko ei. Tällöin kyse ei enää oikeastaan edes ole mielen tietokonemetaforasta vaan

yksinkertaisesti tehokkaasta tavasta kuvata aivojen tai mielen toiminnallista rakennetta.

Tietokonemetaforaan liittyy läheisesti myös kysymys ihmismielen kaltaisen tekoälyn mahdollisuudesta. Mistään edellä sanotusta ei kuitenkaan seuraa, että keinotekoiset mielet olisivat mahdollittomia. Yllä sanottu vain tarkoittaa, että keinotekoisia mieliä, tai ainakin edes etäisesti eläimellisen (mukaan lukien inhimillisen) kaltaisia sellaisia, kannattaa etsiskellä enemmän robotiikan ja keinoelämän kuin klassisen tekoälyn suunnalta.

Viitteet & Kirjallisuus

- 1 Ks. Haanilan, Salmisen & Telakiven artikkeli tässä numerossa.
- 2 Jälkimmäisessä tapauksessa kyse ei tietenkään enää ole varsinaisesti metaforasta, mutta tällainen hieman hämäävä nimitys on jäänyt käyttöön. Joskus puhutaan myös ”tietokoneanalogiasta”.
- 3 Oma tietokonemetaforaan liittyvä keskustelunsa koskee kysymystä, voidaanko minkä tahansa tarpeeksi monimutkaisen järjestelmän tulkita toteuttavan minkä tahansa algoritmin, ja onko asialla väliä. Ks. esim. Gualtiero Piccinini, Computational Modeling vs. Computational Explanation. In *Everything a Computer, and Does It Matter to the Philosophy of Mind?* *Australian Journal of Philosophy*. Vol. 85, No. 1, 2007, 93–115.
- 4 Ks. seuraava alaviite ja Allen Newell, Physical Symbol Systems. *Cognitive Science*. Vol. 4, 1980, 135–183.
- 5 Ks. luku 1 teoksesta Jerry A. Fodor, *The Language of Thought*. Harvard University Press, Cambridge, MA, 1975. On syytä huomata, että tietojenkäsittelytieteissä laskenta määritellään sikäli abstraktisti, ettei prosessoitavien symbolien edellytetä olevan millään tavalla sisällöllisiä. Kuitenkin komputationaalinen mielenteoria rakentaa yhteyden mielen ja formaalin laskennan välille juuri soveltamalla representationalismin sellaista muotoa, joka väittää, että laskennassa käytetyt symbolit voivat olla sisällöllisiä jollain tapaa edustamalla (eli representoimalla) käsitteitä tai käsitteiden viittauskohteita. Klassisesta komputationalismista on toki myös muotoiluja, joiden mukaan mentaaliset representaatiot tai prosessit eivät ole ainakaan sillä tavalla sisällöllisiä, että ne vastaisivat luonnollisen kielen käsitteitä, ks. esim. Steven Stich, *From Folk Psychology to Cognitive Science. The Case Against Belief*. MIT Press, Cambridge, MA 1983.
- 6 On ehkä syytä huomata, että vaikka komputationalismia ja funktionalismia pidetään usein saman teorian eri puolina, ne ovat oikeastaan toisistaan riippumattomia. Funktionalistinen ratkaisu mieli–ruumis-ongelmaan voi siis selvitä, vaikka komputationalismi kuopattaisiin. Funktionalismin voi tiivistää väitteen, että mieli on prosessi eikä olio, ja komputationalismi pyrkii selvittämään, minkälainen prosessi se on. Ks. Gualtiero Piccinini, Functionalism, Computationalism, and Mental States. *Studies in the History and Philosophy of Science*. Vol. 35, No. 4, 2004, 811–833.
- 7 Ks. Jerry A. Fodor & Zenon W. Pylyshyn, Connectionism and Cognitive Architecture. A Critical Analysis. *Cognition*. Vol. 28, No. 1–2, 1988, 3–71 ja Jerry A. Fodor & Ernest Lepore, The Red Herring and the Pet Fish. Why Concepts Still Can't Be Prototypes. *Cognition*. Vol. 58, No. 2, 1996, 253–270.
- 8 Ks. Epsteinin essee tässä numerossa.
- 9 Hyvä neurolaskennan yleisteos, jossa näitäkin asioita käsitellään, on esim. William Bechtel & Adele Abrahamsen, *Connectionism and the Mind. Parallel Processing, Dynamics, and Evolution in Networks*. 2nd ed. Blackwell, Malden, MA 2002.
- 10 Tästä syystä ajattelun systemaattisuuden selittämisen ohella erityisesti chomskylainen kielitiede oli tärkeä vaikuttava tekijä komputationalismin taustalla. Noam Chomsky nimittäin osoitti, että mikäli luonnollista kieltä tarkastellaan periaatteessa äärettömänä kokoelmana lauseita, jotka puolestaan ovat säännönmukaisia symbolirakenteita, voivat nämä rakenteet olla niin monimutkaisia, että niiden jäsentäminen edellyttää rekursiivisia sääntöjä ja sitä myötä työmuistia ja kontrolloitua kognitiivista prosessointia. Tämä käsitys kielestä voidaan kyllä mielekkäästi myös kiistää, mutta ks. Noam Chomsky, *Syntactic Structures*. Mouton, Haag 1957 ja Noam Chomsky, *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA 1965.
- 11 Garrison W. Cottrell & Janet Metcalfe, EMPATH, Face, Emotion, and Gender Recognition Using Holons. Teoksessa *Advances in Neural Information Processing Systems*, vol. 3. Toim. Richard P. Lippmann, John E. Moody & David S. Touretzky. Morgan Kaufmann, San Francisco 1991, 564–571.
- 12 Xu Liu, Steve Ramirez, Petti T. Pang, Corey B. Puryear, Arvind Govindarajan, Karl Deisseroth & Susumu Tonegawa, Optogenetic Simulation of a Hippocampal Engram Activates Fear Memory Recall. *Nature*. Vol. 484, No. 7394, 2012, 381–385.
- 13 Ks. esim. luku 2 teoksessa William Bechtel, *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Psychology Press, New York 2006.
- 14 Jeffrey Shallit, Yes, Your Brain Certainly Is a Computer. *Recursivity*-blogi 19.5.2016. Verkossa: recursed.blogspot.fi/2016/05/yes-your-brain-certainly-is-computer.html. Kiitoksia Sami Torssoselle kyseisen tekstin saattamisesta tietooni ja tässä yhteydessä kiitokset myös Olli Herraselle. Tämä kirjoitus syntyi pitkälti näitä kahta Epsteinin ja Shallitin tekstiä koskevasta ajatustenvaihdostamme.
- 15 Dario Floreano, Toshifumi Kato, Davide Marocco & Eric Sauser, Coevolution of Active Vision and Feature Selection. *Biological Cybernetics*. Vol. 90, No. 3, 2004, 218–228.
- 16 Ks. edellinen viite.
- 17 Ks. Tim van Gelder, What Might Cognition Be, If Not Computation? *The Journal of Philosophy*. Vol. 92, No. 7, 1995, 345–381.
- 18 Tästä väkeväästä antirepresentationalistisen huomion muotoilusta kiitän Aatu Koskensiltaa.
- 19 Hyvä katsaus arkisen intuitiivisen toiminnan ja kontrolloidun järkeilyn psykologisista eroista löytyy teoksesta Daniel Kahneman, *Thinking, Fast and Slow*. Farrar, Straus & Giroux, New York 2011. Lyhyt ja ajanmukainen yhteenvedo symbolisen laskennan haasteista arkijärjen mallintamisessa on esim. Ernest Davis & Gary Marcus, *Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence. Communications of the ACM*. Vol. 58, No. 9, 2015, 92–103. Kognitiivisten kykyjen yleisestä ei-rationaalisesta luonteesta erinomainen katsausartikkeli on Daniel Kahneman & Gary Klein, Conditions for Intuitive Expertise. A Failure to Disagree. *American Psychologist*. Vol. 64, No. 6, 2009, 515–526.
- 20 Ks. luku 5 teoksessa Dario Floreano & Claudio Mattiussi, *Bio-Inspired Artificial Intelligence. Theories, Methods, and Technologies*. MIT Press, Cambridge, MA 2008.