

ROBOTTIEN TIETOISUUS, MORAALI JA VAPAA TAHTO?

Theravāda-buddhalainen näkökulma

Eeva-Mari Virtanen
Sosiaalietiikan pro gradu -tutkielma
Syyskuu 2016

Tiedekunta/Osasto – Fakultet/Sektion Teologinen tiedekunta		Laitos – Institution Systemaattinen teologia	
Tekijä – Författare Eeva-Mari Virtanen			
Työn nimi – Arbetets titel Robottien tietoisuus, moraali ja vapaa tahto? Theravāda-buddhalainen näkökulma			
Oppiaine – Läroämne Sosiaalietiiikka			
Työn laji – Arbetets art Pro gradu	Aika – Datum 1.9.2016	Sivumäärä – Sidoantal 92	
Tiivistelmä – Referat <p>Tämän tutkimuksen tarkoitus on selvittää miten tietoisuus on ymmärretty theravāda-buddhalaisuudessa, millaisia taustaoletuksia tähän tietoisuuskäsitykseen liittyy erityisesti suhteessa moraaliin ja vapaaseen tahtoon ja voivatko tähän tapaan ymmärretyt tietoisuus, moraali ja vapaa tahto ilmetä kehollisessa keinoälyssä. Tutkimusmetodini on systemaattinen analyysi. Theravāda-buddhalaisen oppia on tutkittu, mutta aiheesta voivatko tietoisuus, moraali ja vapaa tahto, siten kun ne ymmärretään tässä opissa, ilmetä kehollisessa keinoälyssä, on edeltävää pohdintaa vain vähän. Hahmotin kokonaiskuvaava aiheestani tutustumalla theravāda-buddhalaisen opin perusteisiin yksilön tietoisuudesta, moraalista ja tahdonvapaudesta, tietoisuuden teorioihin ja keinotekoisesta moraalin kehitykseen viimeaikaisen tutkimuksen valossa. Kirjallisuuden lisäksi olen hyödyntänyt buddhalaisuudesta, neurotieteestä, tietoisuudesta ja tahdonvapaudesta kirjoitettuja tutkimusartikkeleita.</p> <p>Ensimmäisessä pääluvussa vastaan kysymyksiin, mitkä ovat tietoisuuden, yksilön, ilmenemisen ehdot theravāda-buddhalaisen opin mukaan ja mikä on tietoisuuden ja moraalisen toimijuuden ja vapaan tahdon suhde? Buddhalaisessa opissa itseys kielletään ja yksilön ymmärretään olevan viidestä osasta koostuva entiteetti: (1) muoto, (2) ajatukset ja tunteet, (3) käsitteen muodostus, (4) halut ja toiveet ja (5) tietoisuus. Tietoisuus on avain, jonka avulla yksilö voi täydellistyä Buddhan oppien mukaan moraalissa ja viisaudessa. Vapaan tahdon tulkitaan olevan vapaata toimintaa, jota mielihalu ei haittaa.</p> <p>Toisessa pääluvussa perehdyn siihen, miten tietoisuuden käsitteä tulkitaan kognitiivisen neurotieteen ja theravāda-buddhalaisen perinteen näkökulmasta. Tietoisuuden olemuksesta on vaihtelevia tulkintoja. Kognitiivisen neurotieteen näkökulmasta tietoisuus ymmärretään aivojen tuottamaksi ilmiöksi. Toistaiseksi ei ole pystytty selittämään miten tämä tapahtuu. Theravāda-buddhalaisen opin mukaan tietoisuuden virta taas tulkitaan muodosta (aivoista) riippumattomaksi toiminnalliseksi entiteetiksi. Tietoisuuden ymmärretään ankkuroituvan neljään muuhun yksilön muodostavaan osatekijään. Tietoisuus myös jatkaa olemassaoloaan tiedostamattomana elämänvirtana vielä yksilön lakattua olemasta.</p> <p>Kolmannessa pääluvussa kysyn, voivatko tietoisuus, moraali ja vapaa tahto ilmetä kehollisessa keinoälyssä viimeaikainen tutkimuksen valossa. Tietoisuus voidaan ymmärtää kahdella tapaa, mekanismeina tai sisältönä. Mekanismit voidaan rakentaa myös keholliseen keinoälyyn, mutta tietoisuuden sisällön implementointi on käsitteen tuntemattomuuden vuoksi haaste. Kehollisen keinoälyn kyky moraaliseen toimintaan on sidoksissa moraalisen toimijan määrittelyyn. Kehollinen keinoäly voi olla toiminnallisesti moraalinen ja se voi kyetä eettiseen päättelyyn. Eettisen päätöksenteon on yhä haaste. Ihmisen vapaa tahto ja vapaa tahto osoittautuivat kahdeksi erilaiseksi käsitteeksi. Vapaan tahdon implementointi keholliseen keinoälyyn on viimeaikainen tutkimuksen valossa verrattain vaikeaa.</p> <p>Theravāda-buddhalaisen opin mukaan ymmärretyt tietoisuuden ilmeneminen kehollisessa keinoälyssä edellyttää soveltuvaa olomotoa. Mikäli tietoisuus ei ilmene, myös theravāda-buddhalaisen opin mukaan ymmärretty moraali ja tahdonvapaus ovat keholliselle keinoälylle mahdottomuus.</p>			
Avainsanat – Nyckelord buddhalaisuus, moraali, robotiikka, tahdonvapaus, tekoäly, tietoisuus			
Säilytyspaikka – Förvaringställe Helsingin yliopiston kirjasto, Keskustakampuksen kirjasto, Teologia			
Muuta tietoa			

Sisällysluettelo:

Johdanto	1
Tutkimuskysymys ja metodi.....	2
Lähteet ja keskeinen kirjallisuus.....	5
Tutkimuksen rakenne	7
1. Theravāda-buddhalaisuus; itseys, moraali ja vapaa tahto.....	9
1.1 Tietoinen itseys theravāda-buddhalaisessa ajatusmaailmassa	10
1.2 Tapahtumien kausaalisuus ja ehdonvarainen seuraamus.....	14
1.3 Kärsimyksen lopettaminen; moraali ja myötätunto.....	17
1.4 Vapaa tahto buddhalaisessa ajattelussa.....	24
1.4.1 Vapaa tahto, determinismi ja kombatilismi	25
1.4.2 Buddhan opetus vapaasta moraalisesta toiminnasta.....	29
2. Tietoisuus	31
2.1 Teoriat tietoisuudesta.....	33
2.2 Näkökulma tieteelliseen materialismiin.....	39
2.3 Tietoisuus theravāda-buddhalaisessa filosofiassa	44
3. Robottien tietoisuus, moraali ja vapaa tahto?	50
3.1 Älykkäät koneet.....	51
3.2 Robotin tietoisuus	52
3.2.1 Tietoinen kone.....	52
3.2.2 Tunteet ja aistit tietoisuuden mekanismeina	57
3.3 Keinotekoinen moraali.....	59
3.3.1 Moraalisen toiminnan kategoriat ja ohjelmoitu moraali.....	59
3.3.2 Keinotekoinen moraalitoimija.....	65
3.4 Kehollisen keinoälyn vapaa tahto.....	71
3.4.1 Järki, autonomia ja vapaa tahto	71
3.4.2 Täydellinen vapaus	77
Johtopäätökset	80
Lähde- ja kirjallisuusluettelo.....	87
Lähteet.....	87
Kirjallisuus.....	88
Internet-lähteet	91

Johdanto

Tämän tutkimuksen innoittajana on ollut James Hughes'n artikkeli "Compassionate AI and Selfless Robots" kirjassa *Robot Ethics, the Ethical and Social Implications of Robotics*, MIT Press 2012. Hughes kertoo buddhalaisuuden olleen vahvasti dialogissa länsimaisen kognitiotieteen kanssa 2000-luvun ensimmäisen vuosikymmenen aikana tietoisuuden ja itsen luonteeseen liittyen. Hughesin mukaan buddhalainen ajatusmaailma on verrattain avoin koneessa kasvavalle tietoisuudelle. Artikkelissaan Hughes muun muassa esittää Dalai Laman todenneen:

Jos tietokoneen fyysinen olomuoto saavuttaa valmiuden toimia olomuotona tietoisuudelle, tietoisuuden virta ehkä todella saattaisi asettua koneeseen..¹

Robotiikka, keinoäly ja sen kehittäminen ovatkin esilläolevia teemoja. Robotiikan kehitys on muun muassa yksi osa Euroopan komission tavoitetta kehittää digitaalisia yhteismarkkinoita. EU:n kahdeksannessa puiteohjelmassa Horisontti 2020, 2014 – 2020, rahoitetaan yli sataa keinoälyä tutkivaa ja kehittävää hanketta.² Alla lainaus euRobotics AISBL'n laatimasta strategisesta tutkimusagendasta, joka on osa kahdeksannen puiteohjelman julkisen ja yksityisen sektorin yhteistyöhanketta robotiikan saralla.³

Robottiteknologia tulee hallitsevaksi tulevilla vuosikymmenellä. Se tulee vaikuttamaan joka osa-alueella niin kotona kuin työelämässä. Robotiikalla on potentiaali muuttaa elämää ja työtapoja, kasvattaa tuottavuutta ja turvallisuutta, tarjota parempia palveluja ja luoda uusia työpaikkoja. Sen vaikutus tulee kasvamaan ajan myötä, kuten myös vuorovaikutus ihmisten ja robottien välillä.⁴

Hughesin artikkeli sai minut kiinnostumaan buddhalaisen opin itseys- ja tietoisuuskäsitteestä ja siitä, voidaanko sanoa buddhalaisen ajatusmaailman olevan verrattain avoin robotissa ilmenevälle tietoisuudelle, kuten Hughes artikkelissaan kirjoittaa. Jäin pohtimaan, voiko robotti olla tietoinen moraalitoimija.

¹ "If the physical basis of the computer acquires the potential to serve as a basis for a continuum of consciousness ... a stream of consciousness might actually enter into a computer" Hughes 2012, 69, käänös oma.

² EC, strategy; EC, DSM; EC, robotics.

³ EC, PPP; euRobotics AISBL

⁴ "Robotics Technology will become dominant in the coming decade. It will influence every aspect of work and home. Robotics has the potential to transform lives and work practices, raise efficiency and safety levels, provide enhanced levels of service and create jobs. Its impact will grow over time as will the interaction between robots and people." EC, Robotics 2020, 3, käänös oma.

Tutkimuskysymys ja metodi

Tämän tutkimuksen tarkoitus on selvittää miten tietoisuus on ymmärretty theravāda-buddhalaisuudessa, millaisia taustaoletuksia tähän tietoisuuskäsitykseen liittyy erityisesti suhteessa moraaliin ja vapaaseen tahtoon ja voivatko tähän tapaan ymmärretyt tietoisuus, moraali ja vapaa tahto ilmetä kehollisessa keinoälyssä. Tutkimusmetodini on systemaattinen analyysi.

Tutkimuskysymykseni tarkentuivat tutkimusprosessin aikana. Keskeiset kysymykseni ovat: Mitkä ovat tietoisuuden, yksilön, ilmenemisen ehdot theravāda-buddhalaisen opin mukaan ja mikä on tietoisuuden ja moraalisen toimijuuden ja vapaan tahdon suhde? Miten tietoisuuden käsittä tulkitaan niin tieteessä kuin theravāda-buddhalaisen perinteen näkökulmasta? Voivatko tietoisuus, moraali ja vapaa tahto ilmetä kehollisessa keinoälyssä viimeaikainen tutkimuksen valossa?

Buddhalaista ajatusmaailmaa lähestyessä huomaa vääjäämättä, että buddhalaisuus on laaja määre ja pitää sisällään useita koulukuntia. Rajatakseni tutkimustani mielekkäästi, tein valinnan tutkia buddhalaista oppia pāli-kaanonin⁵ näkökulmasta. Pāli-kaanon on viimeaikaisen tutkimuksen valossa varhaisin kirjoitettu teksti, jossa ensin suullisessa muodossa säilynyt Buddhan opetus koottiin kirjalliseen muotoon noin 100 eKr. Kaanonin oppiin kaikkein tiiviimmin yhdistetty koulukunta on theravāda-buddhalaisuus.⁶ Viitatessani siis työssäni ”buddhalaisuuteen”, keskustelen theravāda-buddhalaisuudesta, ellen tekstissä ole toisin indikoinut.

Liittyen buddhalaiseen termistöön, olen valinnut esittää käsitteistä sekä sanskritin- että pālinkielisen vastineen. Esitän ensin sanskritinkielisen termin, jonka jälkeen pālinkielisen vastineen, mikäli termeissä on eroa. Sanskrit oli buddhalaisen ajatusmaailman syntyaikoina n. 500 eKr. bramaanioppineiden ja rituaalien kieli. Pāli jälleen oli yhdistelmä muinaisessa Intiassa puhutun kielen murteita. Oletetaan, että buddhalaisuuden syntyessä Buddha teki tietoisin valinnan opettaa kansankielellä.⁷

Tutkimuskysymyksiini palaten, selvitän mitkä ovat tietoisuuden, yksilön, ilmenemisen ehdot theravāda-buddhalaisen opin mukaan. Selvitän, miten yksilön tietoisuuteen limittyvä käsite itseystä ymmärretään theravāda-buddhalaisessa

⁵ *Pāli*, kaanonin kielen nimi, juontaa termistä pāli-bhasā, *buddhalaisten tekstien kieli*. Gethin 1998, 41.

⁶ Gethin 1998, 41.

⁷ Gethin 1998, 41–42.

opissa. Tutkin, miten tietoisuus nähdään theravāda-buddhalaisessa opissa suhteessa moraaliseen toimintaan ja vapaan tahdon ilmenemiseen. Pohdin taustalla, voidaanko tietoisuuden väittää olevan edellytys moraalille toimijuudelle. Toisessa tutkimuskysymyksessäni kysyn, miten tietoisuuden käsitettä on analysoitu länsimaisen tieteen kentällä kognitiivisen neurotieteen valossa ja miten tietoisuus vastaavasti ymmärretään pāli-kaanonin theravāda-buddhalaisesta perspektiivistä.

Ymmärrys ihmisen aivojen rakenteesta ja toiminnasta alkoi kasvaa 1800-luvulla. Kognitiiviseen psykologiaan pohjautuva tieteenalana, kognitiivinen neurotiede, tutkii, miten aivot tuottavat yksilössä ilmenevät ajatukset, tunteet ja ilmiöt. Kognitiivisen neurotieteen vallitseva paradigma on, että aivot mahdollistavat henkiset tapahtumat.⁸ Tämä näkemys sointuu yhteen eliminatiivisena materialismina tunnetun filosofian suuntauksen näkemyksen kanssa. Eliminatiivinen materialismi hahmottaa tunteet, sikäli, kun ne ovat olemassa, aivojen tilaksi. Erityisesti filosofit Paul ja Patricia Churchland ovat argumentoineet voimakkaasti eliminatiivisen materialismin näkökulman puolesta.⁹ Keskustellessani länsimaisen tieteen näkökulmasta viittaan siis ennen kaikkea kognitiivisen neurotieteen ja eliminatiivisen materialismin teoriaan aivoista henkisten ilmiöiden takana. Vastausta kysymykseen, miten aivot mahdollistavat henkiset ilmiöt, ei toistaiseksi ole löydetty. Tietoisuuden analysointi on siis haastava tehtävä tietoisuuden käsitteen karatessa tarkalta tieteelliseltä määritelmältä. Willard Van Orman Quine on todennut tietoisuutta koskien: ”Tiedämme millaista on olla tietoinen, muttemme osaa esittää sitä tyydyttävien tieteellisin termein.”¹⁰ Keskustellessani tutkimuksessani tietoisuudesta käsitteenä, olen tehnyt ratkaisun määritellä sen ”immateriaaliseksi läsnäolon tunteeksi”¹¹. Keskustelen tietoisuuden käsitteestä ensisijaisesti fenomenalisena, aistihavaintoon ja kokemukseen perustuvana¹².

Viimeisenä tutkimuskysymyksenäni kysyn, voivatko tietoisuus, moraalit ja vapaa tahto ilmetä kehollisessa keinoälyssä viimeaikainen tutkimuksen valossa. Tässä vaiheessa on hyvä pohtia myös, mitä on keinoäly. Keskustellessani keinoälystä, rajaan tässä tutkimuksessa keinoälyn samojen suuntaviivojen mukaan,

⁸ Ward 2015, 3, 15.

⁹ Stanford Encyclopedia of Philosophy.

¹⁰ “We know what it is like to be conscious, but not how to put it in satisfactory scientific terms.” Honderich 2014, 7, käännös oma.

¹¹ Haikonen 2007, 2.

¹² Honderich 2014, 46.

kuin Steve Torrance artikkelissaan ”Ethics and Consciousness in Artificial Agents”. Rajaan käsittelemäni keinoälyn ulkopuolelle virtuaaliset keinoälyt, hybridit, joissa elävää organismia on paranneltu mekaanisesti ja mahdolliset näennäisbiologiset keinotekoiset entiteetit. Näennäisbiologisilla keinotekoisilla entiteeteillä – Torrancea lainaten – viitataan keinotekoisiiin entiteetteihin, joilla on niin paljon, biologisia tai näennäisbiologisia, ominaisuuksia, ettei niitä voi enää kutsua koneiksi. Keskustelenkin keinoälystä viitaten keholliseen keinoälyyn, joka on fyysisen rakenteensa perusteella selkeästi kone, ei organismi.¹³

Viitataan kehollisella keinoälyllä tutkimuksessani myös käsitteeseen robotti, tšekinkielisestä sanasta *robota*¹⁴ juontavaan termiin. Käsitteen esitteli Karel Čapekin näytelmässään R.U.R. Rossum’s Universal Robots¹⁵. Nojaudun tutkimuksessani myös seuraavaan P.W. Singerin määritelmään: Robotti on

[t]utkimuksessa ”aisti – ajattele – toimi” – oppimalliin perustuva kone. Ne ovat ihmisen rakentamia kolmella pääkomponentilla varustettuja laitteita. Pääkomponentit ovat ”anturit”, jotka havainnoivat ympäristöä ja sen muutoksia, ”prosessorit” tai ”keinoäly”, joka reagoi muutoksiin ja ”efektorit” joiden avulla robotti pystyy toimimaan ympäristössään, havainnosta prosessoimansa päätöksen perusteella, samalla aiheuttaen muutosta ympäristössään.¹⁶

Käyttäessäni tekstissäni käsitettä kone, mikä juontaa lähdetekstin termistöön (machine), viitataan siis ihmisen rakentamaan ”aisti – ajattele – toimi” – oppimallille perustuvaan koneeseen – robottiin.

Entä mitä on moraalinen toiminta? Wendell Wallach ja Colin Allen toteavat kirjassaan *Moral Machines* hyvän moraalisen toimijan pystyvän sääntelemään käytöstään sen valossa, millaista harmia tai hyvää kyseinen käytös voi aiheuttaa, tai mitkä velvollisuudet ottaa huomioon tai jättää huomiotta¹⁷. Tähän nojaten, keskustelenkin moraalista toimimisesta käsitteenä moraaliseen tietoon pohjautuvana tekemisenä. Mutta onko tämän toiminnan oltava tahdottua? Myös vapaa tahto ilmiönä ja käsitteenä synnyttää ihmetystä, ja sen luonteesta on vaihtelevia näkemyksiä. Kuten tietoisuuskin, vapaa tahto vaikuttaa ennen kaikkea

¹³ Torrance 2008, 496.

¹⁴ Robota-sananj juuret ovat proto-slaavilaisessa sanassa orbota, jonka juuret jatkuvat puolestaan venäjän rabotaan, työ, mikä taas juontaa indo-eurooppalaisesta lähteestä orbh, lapsi, orjatyttö, edelleen sukua armenian sanalle orb, hindin arbha, lapsenlapsi sekä kreikan sanalle orphanos, orpo, mikä on sukua myös saksan sanan Arbeit, työ, kanssa, fr.oldict.com 2016.

¹⁵ Whitby 2012, 235.

¹⁶ “...machines that are built upon what research call the ”sense-think-act” paradigm. That is, they are man-made devices with three key components: “sensors” that monitor the environment and detect changes in it, “processors” or “artificial intelligence” that decides how to respond, and “effectors” that act on the environment in a manner that reflects the decisions, creating some sort of change in the world around a robot.” O’Meara 2012, 164, käänös oma.

¹⁷ Wallach & Allen 2009, 16.

kokemukselliselta ilmiöltä. Immanuel Kant on kirjoittanut vapaasta tahdosta siten, että vapaan tahdon on oltava vapaa ulkopuolisten syiden vaikutuksesta. Vapaan tahdon on myös kyettävä itsemääräämisoikeuteen tai autonomiaan.¹⁸ Tarkoitukseni ei ole keskustellessani tässä tutkimuksessa tietoisuudesta, moraalista ja tahdonvapaudesta kyseenalaistaa käsityksiä tietoisuuden, moraalisen toiminnan tai vapaan tahdon moninaisista olemuksista. Paremminkin, pohdintani kietoutuu kysymykseen, onko tietoisuus, siten kun se koetaan, ja tiedostettu, fenomenaalinen vapaa tahto tarpeellista moraalille toiminnalle. Edellä esittämäni määrittelyt haasteellisille käsitteille tietoisuus, moraalit ja vapaa tahto eivät tavoita näiden käsitteiden kaikkia vivahteita. Toivon silti, että ne tarjoavat lukijalle tarpeellisen ajatuksen, mitä tarkoitan käsitteistä keskustellessani.

Lähteet ja keskeinen kirjallisuus

Theravāda-buddhalaisen opin ymmärrystä tietoisuudesta ja sen suhdetta moraalisiin ja tahdonvapauteen on tutkittu, mutta tähän nivoutuvasta tutkimukseni aiheesta – voivatko tietoisuus, moraalit ja vapaa tahto, siten kun ne ymmärretään theravāda-buddhalaisessa opissa, ilmetä kehollisessa keinoälyssä – ei aiheeni valinnan inspiraationa toimineen Hughes'in artikkelin "Compassionate AI and Selfless Robots" lisäksi ole kirjoitettu edeltävää tutkimusta. Olenkin hahmottanut kokonaiskuvaa aiheestani useiden kirjoittajien teoksiin tutustumalla. Minulle oli tärkeää ammentaa perusteet theravāda-buddhalaisesta opista yksilön tietoisuudesta, moraalista ja tahdonvapaudesta, tietoisuuden teorioista ja keinotekoisesta moraalien kehityksestä.

Tein ratkaisun opiskella theravāda-buddhalaisuuden perusteita yksilön tietoisuudesta, moraalista ja tahdonvapaudesta. Tässä minua auttoi Rupert Gethin kirjassaan *The Foundations of Buddhism*. Pāli-kaanonin ja theravāda-buddhalaisuuden itseyskäsitykseen sekä tietoisuuden roolin ymmärrykseen minua on johdattanut mitä mainioimmin Steven Collins kirjassaan *Selfless persons. Imagery and thought in Theravāda Buddhism*. Buddhalaisen etiikan ja psykologian tutkimisessa tukeuduin vahvimmin Damien Keown'n teokseen *The Nature of Buddhist Ethics*. Buddhalaisesta vapaan tahdon käsitteestä on kirjoitettu varsin vähän. Parhaiten tietoa tältä saralta ovat tarjonneet Asaf Federman artikkelissa "What Kind of Free Will did the Buddha Teach?", B. Alan Wallace artikkelissa "A

¹⁸ O'Neill 1989, 52–53.

Buddhist View of Free Will: Beyond Determinism and Indeterminism” ja Nicholas F. Gier ja Paul Kjellberg artikkelissa “Buddhism and the Freedom of Will: Pali and Mahayanist Responses.

Tutkin tietoisuuden teorioita kognitiivisen neurotieteen näkökulmasta lähteenäni Ted Honderichin teos *Actual Consciousness*. Tässä Honderich pohtii filosofian ja tieteen näkökulmasta vakiintuneimpia teorioita tietoisuuden olemuksesta. Kiinnostavan haasteen tieteen paradigmalle tietoisuudesta aivojen tuottamana löysin B. Alan Wallacen kirjasta *Contemplative Science: Where Buddhism and Neuroscience Converge*. Kirja on myös sillanrakentaja pohdintaani theravāda-buddhalaisen opin tietouskäsityksestä.

Olen tutkinut ajatuksia keinotekoisien tietoisuuden, moraalien ja tahdonvapauden implementoinnista keholliseen keinoälyyn Wendell Wallachin ja Collin Allenin yhteistyössä kirjoittamaan *Moral Machines: Teaching Robots Right from Wrong* –teoksen valossa. Kirja on kiinnostava yleisteos viimeaikaisen tutkimuksen näkökulmista robottien mahdolliseen moraaliseen toimintaan, joskin kirjoittajat tarjoavat verrattain suoraviivaisen, funktionalistisen näkökulman. Kirjoittajat myös lainaavat laajalti filosofien ja tutkijoiden käsityksiä keinoälyn luonteesta, mikä on puolestaan laajentanut lähdeaineistoani tarkistaessani alkuperäisiä lähteitä. *Moral Machines* perustuu toiminnalliseen teoriaan moraalista ja toimijuudesta. Etsin työhöni laajempaa näkökulmaa robottien mahdollisesta moraalista toimijuudesta David J. Gunkelin teoksesta *The Machine Question. Critical Perspectives on AI, Robots, and Ethics*. Saadakseni työhöni tekniikan näkökulman aktuaalisen kehollisen keinoälyn koostamisesta, tutkin Pentti Haikosen pohdintoja keinoälyn tietoisuuden luonteesta ja tämän toteuttamisesta teoksesta *Robot Brains: Circuits and Systems for Conscious Machines*.

Tahdonvapautta käsittelen Immanuel Kantin ajatuksiin nojautuen. Näitä olen ammentanut Onora O’Neillin teoksesta *Constructions of Reason, Explorations of Kant’s Practical Philosophy*. Pohdin tahdonvapautta myös kirkkoisä Augustinuksen ja René Descartesin aatteiden valossa Christopher Gilbertin artikkelista ”Grades of Freedom”. Olen hyödyntänyt lisäksi buddhalaisuudesta, neurotieteestä sekä tietoisuudesta ja vapaasta tahdosta kirjoitettuja tutkimusartikkeleita. Viittaan muun muassa seuraaviin kirjoittajiin: Paul Churchland, James Gips, Christof Koch, Giulio Tononi, John Searle, John Sullins sekä Reijo Työrinoja.

Tutkimuksen innoittaja James Hughes on kirjoittanut roboetiikkaan liittyen teoksen *Citizen Cyborg: Why Democratic Societies Must Respond to the Redesigned Human of the Future*. Tällä hetkellä Hughes työskentelee The Institute for Ethics and Emerging Technologies (IEET):ssa Cyborg Buddha -projektissa¹⁹ ja kirjoittaa kirjaa *Cyborg Buddha: Using Neurotechnology to Become Better People*.

Kuten totesin, James Hughesin artikkelin ja tutkimuksen lisäksi tekoälyn tietoisuutta, moraalia tai tahdonvapautta ei ole lähestytty buddhalaisen filosofian näkökulmasta. Aihetta sivuava tutkimus liittyy pääasiassa buddhalaisen filosofian ja kognitiotieteen vuoropuheluun. Tutkimus on keskittynyt aivojen tutkimiseen, neuroverkkoihin ja älykkääseen roboetiikkaan. Tutkimusta on tehty myös tietoisuuden luonteesta niin buddhalaisen filosofian kuin vastaavasti kognitiivisen neurotieteen valossa. Robotiikan ja keinoälyn tutkimus etenee yhä kasvavissa määrin, mikä näkyy myös arkipäivän uutisotsikoissa ja myös jo hieman arkielämässä.

Tutkimuksen rakenne

Aloitan tutkimukseni ensimmäisen pääluvun perehtymällä theravāda-buddhalaiseen oppiin itseyden luonteesta vastaten samalla tutkimuskysymyksen, mitkä ovat tietoisuuden, yksilön, ilmenemisen ehdot theravāda-buddhalaisen opin mukaan ja mikä on tietoisuuden ja moraalisen toimijuuden ja vapaan tahdon suhde? Pohdin yksilön ilmenemiseen liittyen buddhalaisuuden keskeistä oppia ei-itsestä (anātman/pāli: anattā). Theravāda-buddhalaisessa opissa asiat ja tapahtumat koetaan kausaalisiksi. Keskeisellä sijalla ovat käsitteet ja ehdonvarainen seuraamus, Neljä Jaloa Totuutta sekä Jalo Kahdeksanosainen Polku. Keskustelen, miten tietoisuus linkittyy theravāda-buddhalaisen opin mukaan näihin käsitteisiin ja niiden mukana hyvään moraaliseen toimintaan ja viisauteen. Näen mielekkääksi tässä yhteydessä käsitellä vapaan tahdon käsitettä theravāda-buddhalaisessa ajattelussa viimeaikaisen tutkimuksen valossa ja pohtia edelleen tietoisuuden toiminnan roolia myös suhteessa theravāda-buddhalaiseen käsitykseen vapaasta tahdosta.

Tutkielman toisessa pääluvussa keskityn pohtimaan tietoisuuden käsitettä ja miten tätä on tulkittu. Vastaan tutkimuskysymykseeni, miten tietoisuuden käsittämistä tulkitaan niin tieteessä kuin theravāda-buddhalaisen perinteen näkökulmasta.

¹⁹ Cyborg Buddha.

Aloitin länsimaisen tieteen näkemyksestä esitellen Ted Honderichin ajatusten mukaiset viisi johtavaa ajatusta tietoisuudesta sekä funktionalismin käsitteen, joka Honderichin mukaan on tällä hetkellä läpikotaisin ja hallitsevin käsitys tietoisuudesta. Esittelen käsitteen ”Hard Problem” – miten tietoisuus ilmenee, miten aivot tuottavat tietoisuuden – ja John Searlen näkemyksen tästä ongelmasta. Jatkan esittelemällä B. Alan Wallacen näkökulman tietoisuuden luonteesta. Tämä vaikuttaa eittämättä haastavan tieteellisen paradigman tietoisuudesta aivojen tuottamana.

Siirryn tämän jälkeen pohtimaan tietoisuuden käsitteen ymmärrystä theravāda-buddhalaisessa filosofiassa. Käyn läpi theravāda-buddhalaista tietoisuuden käsitteistöä. Keskeisiä käsitteitä ovat rakennetietoisuus, (abhisamkhāra- viññāṇa), tietoisuus elämän ylläpitäjänä, ravintona (āhāra), tietoisuuden laskeutumisen vertauskuva Gandharva (pāli: Gandhabba)-henki, ihmispersoonallisuuksien ketju (attabhāva) sekä bhavaṅga-mieli, tapahtumien taustalla virtaava tiedostamaton elämänjatkumo.

Kolmannessa pääluvussa tutkin keinoälyn kehitystä viimeaikaisen tutkimuskirjallisuuden valossa ja vastaan tutkimuskysymykseeni voivatko tietoisuus, moraalit ja vapaa tahto ilmetä kehollisessa keinoälyssä viimeaikainen tutkimuksen valossa. Pohdin ensin keinoälyä ja rakennetun robotin älykkyyttä. Tämän jälkeen keskustelen tietoisuuden ilmenemisestä robotissa ja tunteiden ja aistien asemasta tietoisuuden ilmenemisessä. Etenen näiden pohdintojen jälkeen keskustelemaan keinotekoisesta moraalista. Lähestyn moraalista toimijuutta, moraalien kategorioita ja sitä, miten viimeaikainen tutkimus näkee moraalien rakentamisen mahdollisuudet. Pohdin, millainen olisi hyvä keinotekoinen moraalitoimija. Tästä etenen edelleen kysymykseen vapaasta tahdosta. Keskustelen vapaasta tahdon käsitteestä nojautuen kirkkoisä Augustinuksen, Rene Descartesin ja Immanuel Kantin näkemyksiin. Pohdin kysymystä, mitä lopulta on täydellinen vapaus. Onko se – kuten Augustinus toteaa – ”enkelien ja siunattujen vapaus taivaassa”?

Lopuksi kokoon pääluvuissa käsitellyt aiheet tutkimusraporttini johtopäätöksiin tarjoten vastauksen tutkimuskysymyksiini. Pyrin tarjoamaan tiiviin katsauksen aihepiireistä, joista moni jo itsessään mielenkiintoisuutensa ja laajuutensa perusteella ansaitsisi oman pro gradu-tutkielmansa.

1. Theravāda-buddhalaisuus; itseys, moraalit ja vapaa tahto

Buddhalaisuutta tutkiessa on ensin ymmärrettävä kulttuurillinen tausta, josta tämä ajattelu lähti aikanaan kehittymään. Steve Collinsin mukaan keskeisiä aiempia taustatekijöitä olivat nykyisen Intian ”sanskritisatio” tai ”braminisatio”. Tämä tarkoitti arjalaisheimojen kielen, kulttuurin ja käytäntöjen ja bramiini-pappien ryhmän leviämistä koko nykyisen Intian niemimaan alueelle n. 1500–1000 eKr. Tätä sanskritinkielisen kulttuurin leviämistä kuvaavat uskonnolliset tekstit, Vedat, Brāhmaṇat sekä Upaniṣadit. Collinsin mukaan buddhalaisuus syntyi ennen kaikkea vastareaktiona nykyisen Intian alueella vallalla tuolloin olleelle bramiinuskolle. Buddhalaisuus lainasi bramiinuskolta kuitenkin käsitteet, uudelleensyntymisestä (saṃsāra), tekojen vaikutuksesta kohtaloon (karma/pāli: kamma) ja vapautumisesta (mokṣa).²⁰

Buddhalaisuus sai alkunsa noin 500 vuotta ennen länsimaista ajanlaskun alkua. Mies, joka tunnetaan Buddhana, syntyi Siddhārtha Gautama (pāli: Siddhattha Gotama) -nimisenä aristokraattiseen perheeseen Kapilavastu -nimiseen kaupunkiin nykyisen Intian–Nepalin rajalla²¹. Buddhalaisen ajattelun keskeinen tavoite on, Buddhan esimerkkiä seuraten, valaistuminen. Tätä kuvataan traditiossa sanalla bodhi, herääminen. Valaistunut henkilö ymmärtää kärsimyksen olemuksen, sen syyt, lakkaamisen ja tavan päätyä kärsimyksen lakkaamiseen. Buddhalaisuus määritteli uskontona ja tämä määrittely on Rupert Gethiniä lainaten mielekäs siten, että hartaudenharjoitukset ja rituaalit koetaan buddhalaisuudessa yleisesti perusteltuina ja olennaisina valaistuksen saavuttamiseksi. Buddhalaisuutta ei kuitenkaan määritellä uskontunnustuksen kautta, eikä Buddhalaisuus muistuta uskontoa siten, että siihen kuuluisivat pelastava/-t jumala/-t.²²

Buddhalaisuus voidaan nähdä järjestelmänä, jossa harjoitellaan käytöstä, meditaatiota ja viisautta. Buddhan viitoittamaa polkua seurataan, jotta saavutetaan kärsimyksen lakkaaminen. Jos buddhalaisuus määriteltäisiin tällaisena

²⁰ Collins 1982, 29–30

²¹ Siddhārtha Gautaman elinvuodet ajoitetaan useimmissa julkaisuissa vuosiin 566 – 486 eKr. Tämä juontaa lähteistä, joiden mukaan kuningas Aśoka, Mauryan-valtakunnan kolmas hallitsija ja buddhalaisuuden suosija kruunattiin 218 vuotta Buddhan kuoleman jälkeen (268 eKr.). Ajoituksesta on kuitenkin vaihtelevia versioita ja vuosiluvut lähteissä tulisi nähdä suuntaa antavina arvioina. Nykyinen tutkimus asettaa Buddhan kuoleman lähemmäs vuotta 400, kuin 500, eKr. Keown 2004; Gethin 1998, 14.

²² Gethin 1998, 14–15, 65.

järjestelmänä, se voitaisiin nähdä Gethinin sanoja lainaten käytännöllisenä (practical) tapana elää todellisuuden kanssa. Buddhalaisuus on ensisijaisesti vastaus eettiseen ongelmaan: miten elää parasta mahdollista elämää^{23, 24}.

Tarkastelen tässä tutkielmani ensimmäisessä pääluvussa lähemmin miten theravāda-buddhalaisessa opissa ymmärretään tietoisuuden, yksilön, ilmenemisen ehdot. Haluan myös löytää vastauksen kysymykseeni, mikä on tietoisuuden ja moraalisen toimijuuden ja vapaan tahdon suhde theravāda-buddhalaisen opin valossa.

1.1 Tietoinen itseys theravāda-buddhalaisessa ajatusmaailmassa

Esittelen tässä luvussa theravāda-buddhalaisen opin tietoisien yksilön, ilmenemisestä. Keskeisen tärkeä käsite on oppi ei-itsestä (anātman/pāli: anattā). Kuten jo kirjoitin, Steve Collins kertoo buddhalaisen ajattelun pohjautuvan bramiinitradiotille. Bramiinitradiotiosta periytyvät myös buddhalaisuuden keskeiset opit elämän kierrosta, eteenpäin soljumisesta ja uudelleensyntymisestä, tekojen vaikutuksesta kohtaloon ja vapautumisesta uudelleensyntymän kierrosta. Buddhalainen traditio tuntee bramiinitradiotion käsitteen vapautuminen sanalla nirvāṇa (pāli: nibbāna). Collinsin ajattelua lainaten, tiedostettu eläminen ymmärretään theravāda-buddhalaisessa traditiossa yhtenä osana useiden eri elämien sarjasta. Nämä elämät voivat tapahtua useissa muodoissa ja maailmoissa²⁵. Uudelleensyntymän ja seuraavan onnellisuuden tai kärsimyksen määrittävänä tekijänä ovat teot ja niiden moraalinen arvo. Tavoitteena on vapautuminen, nirvāṇa. pääsy pois uudelleensyntymän kierrosta. Damien Keown'ia siteeraten, tämä on ”buddhalaisuuden *summum bonum*... ...Samsāra on ongelma, johon nirvāṇa on ratkaisu^{26, 27}.

²³ Keown 1992, 1.

²⁴ Gethin 1998, 64.

²⁵ Näitä ovat theravāda -koulukunnan mukaan mielihalun maailmat (Helvetti, eläinten maailma, henkien maailma, demonien maailma, ihmisten maailma ja kuusi jumalallista (divine) maailmaa), jalostuneen olomuodon maailmat (neljä eriasteista meditaation tasoa, joille kullekin ominaista eri määrä taivaita) ja hahmottomat maailmat (äärettömän avaruuden taso, äärettömän tietoisuuden taso, tyhjyyden taso, ei astimisen eikä aistimattomuuden (neither perception nor non-perception) taso) Collins 1982, 217.

²⁶ ”The *summum bonum* of Buddhism... ...Samsāra is thus the problem to which nirvāṇa is the solution.” Keown 2004, nirvāṇa, käänös oma. Kursiivi kuten alkuperäisessä tekstissä.

²⁷ Collins 1982, 29.

Kuten jo mainitsin, buddhalaisen opin keskeinen kulmakivi on oppi ei-itsestä (anātman/pāli: anattā).²⁸ Steven Collins on antanut kirjassaan *Selfless Persons* puheenvuoron kolmelle buddhalaiselle ajattelijalle tämän opin esittelyssä ja koen tutkimustani pohjustaakseni parhaaksi toimia samoin. Annan Collinsia lainaten äänen Walpola Rāhulalle teoksesta *What the Buddha Taught*:

Buddhalaisuus erottuu ainutkertaisena ajattelun historiassa kieltäessään tällaisen Sielun, Itsen tai *Ātmanin* olemassaolon. Buddhan opetukseen viitaten, idea itsestä on kuvitteellinen, väärä uskomus, jolla ei ole vastinkappaletta todellisuudessa, ja se tuottaa haitallisia ajatuksia ”minä” ja ”minun”...²⁹

G.P. Malalasekera on kuvannut käsitettä seuraavin sanoin:

Yhtä kaikki, tämä oppi Ei-Sielusta, tai *Anattā*, on buddhalaisuuden peruskallio ja kaikki muut Buddhan opetukset ovat kiinteästi yhdistyneet sen kanssa . . . Nyt, mikä on tämä ”Sielu”, jonka olemassaolon Buddha kiistää? Lyhyesti sanottuna, sielu on pysyvä, erillinen, jatkuvasti olemassa oleva ja häviämätön entiteetti, jonka yleisesti uskotaan löytyvän ihmisestä.³⁰

Mahāthera Nyanatiloka kuvaa oppia:

Buddha opettaa, että ilmauksemme ego, itse, sielu, persoonallisuus, jne. ovat vain tavanmukaisia termejä, jotka eivät viittaa mihinkään oikeaan itsenäiseen entiteettiin. Ja hän opettaa, että on olemassa vain psyko-fyysinen hetki hetkeltä muuttuva olemassaolon prosessi. . .³¹

Kuten ylläolevista lainauksista voi ymmärtää, buddhalaisesta opillisesta näkökulmasta itse ei ole ”pysyvä, ikuinen ja absoluuttinen entiteetti, joka on muuttumaton substanssi muuttuvan, ilmiöitä sisällään pitävän, maailman takana³²”.

Buddhan opetukseen perustuen toimija, ihminen, koostuu osatekijöistä (skandha/pāli: khandha). Tietoisuus käsitteenä ja ilmiönä on yksi näistä³³. Ajattelutapaa, jossa itseys nähdään osasista koostuvana, kuvaa esimerkiksi buddhalainen teksti Milindapañha, Milindan kysymykset.

Tarina kertoo muinaista Bactariaa hallitsevan kreikkalaisen kuningas Milindan ja buddhalaisen munkki Nāgasenan tapaamisesta. Kuninkaan saapuessa

²⁸ Collins 1982, 3–4.

²⁹ “Buddhism stands unique in the history of human thought in denying the existence of such a Soul, Self, or *Ātman*. According to the teachings of the Buddha, the idea of self is an imaginary, false belief which has no corresponding reality, and it produces harmful thoughts of ‘me’ and ‘mine’...” Collins 1982, 4, käännös oma. Kursiivit kuten alkuperäisessä tekstissä.

³⁰ “Yet this doctrine of No-soul or *Ānatta* is the bedrock of Buddhism and all other Teachings of the Buddha are intimately connected with it... Now what is this ‘Soul’ the existence of which the Buddha denies? Briefly stated, the soul is the abiding, separate, constantly existing and indestructible entity which is generally believed to be found in man.” Collins 1982, 4, käännös oma. Kursiivit ja pisteet kuten alkuperäisessä tekstissä.

³¹ “The Buddha teaches that what we call ego, self, soul, personality, etc., are merely conventional terms not referring to any real independent entity. And he teaches that there is only to be found this psychophysical process of existence changing from moment to moment.” Collins 1982, 5, käännös oma.

³² Collins 1982, 4.

³³ Gethin 1998, 135.

tapaamaan munkkia Nāgasena esittelee itsensä, mutta toteaa sitten nimensä olevan vain nimilappu, oikeastaan henkilöä Nāgasena ei ole. Kuningas hämmästyttää ja toteaa munkin puhuvan tyhjänpäiväisyyksiä. Nāgasena vastaa kuninkaalle kysymällä, miten tämä saapui munkkia tapaamaan. Kuningas vastaa ”vaunuilla”, jolloin Nāgasena esittää kysymyksen: ”Mitkä ovat vaunut?”, ja jatkaa tiedustellen, ovatko vaunut akseli, pyörät tai kenties valjaat. Kuningas vastaa, etteivät vaunut ole mikään näistä, mutta toteaa olevan järkevää puhua vaunuista, koska tämä kuvaa asiaa, jossa muun muassa akseli, pyörät ja valjaat ovat ja toimivat yhdessä. Nāgasena toteaa asian olevan juuri näin. Samoin on aivan järkevää puhua Nāgasenasta, viitatessa olentoon, joka vain koostuu itse asiassa viidestä osatekijästä. Mutta ei ole olemassa itsenäistä itseä, joka olisi Nāgasena.³⁴

Buddhalaisessa opissa itsestä siis ymmärretään myös yksilön olevan osistaan koostuva entiteetti, ikään kuin kuningas Milindan vaunut. Ihmistä voi kutsua tavanomaisesti mielekkäällä nimilapulla, mutta nimilapun takana onkin osasia, jotka toimivat joka hetki yhdessä ja luovat kokonaisuuden, tässä tapauksessa ihmisen. Olemassaolo selittyy koettujen asioiden kausaalisen yhteytenä. Gethin kertoo, että buddhalainen perinteen parissa on analysoitu ihmisen olemusta useilla eri tavoilla³⁵. Näistä hänen mukaansa kuitenkin ehkä tärkein pohjautuu myös Milindapañha-tekstistä lainatussa esimerkissä mainittuun viiteen fyysiseen ja henkiseen osatekijään.³⁶ Alla olevaan taulukoon on koostettu nämä viisi osatekijää, siten kuin Gethin ne kuvaa kirjassaan *The Foundations of Buddhism*. Halujen ja toiveiden, merkitystä perittynä edellisestä olomuodosta on selvennetty erityisesti Collinsia mukailleen kirjasta *Selfless Persons*,

(1) muoto	<i>rūpa</i>	vaikutelma kehosta ja aisteista, ja ympäröivästä maailmasta
(2) ajatukset ja tunteet	<i>vedanā</i>	nousevat fyysisen maailman vaikutuksesta
(3) käsitteen muodostus	<i>saṃjñā (pāli: saññā)</i>	kyky luokitella asiat
(4) halut ja toiveet, henkiset muodostelmat, perityt voimat.	<i>saṃskāra (pāli: saṃkhāra)</i>	nämä nousevat 2&3 vaikutuksesta
(5) tietoisuus	<i>viññāna(pāli: viññana)</i>	kyky ajatella, havainnoida

Tauluko 1: Viisi osatekijää³⁷

³⁴ Gethin 1998, 139.

³⁵ Näistä esimerkkeinä kaksiosainen nimi-muoto, nāmarūpa, –malli, neliosainen nähdyt, kuullut, ajatellut ja tiedostetut (asiat), 12 piiriä sekä kuusiosainen aistipohjainen, āyatana, –malli. Āyatana, –mallissa voidaan nähdä myös eri aistien kohteiden kuusi eri luokkaa. Viimeisen muunnelman tunnetaan myös 18 elementtinä, jolloin edellä mainittuihin lisätään kuusi tietoisuuden luokkaa. Gethin 1998, 140–141; Collins 1982, 82.

³⁶ Gethin 1998, 135, 140–143.

³⁷ Gethin 1998, 135; Collins 1982, 202.

Kokemuksen voidaan tulkita muodostuvan viidestä henkisestä ja fyysisestä ilmiöistä: rūpa, vedanā, saṃjñā (pāli: saññā), saṃskāra (pāli: saṃkhāra), ja vijñāna (pāli: viññana). Nämä viisi ilmiötä kuvaavat perustavalla tavalla yksilön. Ihmisellä on vaikutelma kehosta ja aisteista, ja häntä ympäröi olemassa oleva maailma. Tätä kutsutaan muodoksi (rūpa). Ihmismielessä tapahtuu jatkuvasti henkisiä tapahtumia ja fyysisen maailman nostattamia ajatuksia ja tunteita (vedanā). Nämä aistikokemukset edelleen tunnistetaan, (saṃjñā / pāli: saññā), ja tätä kautta luokitellaan. Aistikokemukset herättävät ihmisessä tahtotilaa ilmaisevia tunteita, haluja ja toiveita (saṃskāra/ pāli: saṃkhāra). Lisäksi on viides osatekijä, perustava itsetietoisuus (vijñāna/ pāli: viññana). Ihminen on ajatteleva olentoja, joilla on havaintoja ja ajatuksia.³⁸

Opillisen totuuden mukaan itseä ei siis tule ymmärtää itsenäisenä, pysyvänä ja muuttumattomana. On kuitenkin huomattava, Collinsia lainaten, että opillisen totuuden itsestä rinnalla theravāda-buddhalaisuudessa tunnetaan niin kutsuttu ”tavanmukainen” (conventional) totuus itsestä. Theravāda-tradition mukaan ei-itseyden opin, anattā, oikea soveltaminen on mahdollista vain todelliselle virtuoosille, toisin sanoen, oppineelle munkille. Useimmat theravāda-buddhalaisuuden harjoittajat ymmärtävät itsen ”tavanomaisella” tavalla, toisin sanoen, he näkevät ”itsen”, joka jatkaa myös seuraavassa elämässä, ja pyrkivät hyvillä teoilla turvaamaan hyvän uudelleensyntymän. Valaistumattomalla henkilöllä on, ja kuuluu olla, tunne ”minästä” (asmi-māna). Kuvitelma ”minä olen”, on ikään kuin kahle – mutta ellei kahleita olisi, ei olisi mitään mistä päästä vapaaksi. Ilman kahleita valaistuminenkaan ei olisi mahdollista.³⁹

Gier ja Kjellberg toteavat, ettei pāli-kaanon sinällään kiellä kokemuksellista itseä, vain sen, että itse olisi jokin muuttumaton ja pysyvä. Gierin ja Kjellbergin mukaan pāli-kaanonin kuvaama itse tulee ymmärtää toiminnalliseksi. Osatekijät tulisi ymmärtää paremminkin toimintoina kuin entiteetteinä.⁴⁰ Ajatus yksilöstä voidaankin kuvata monitahoisena virtana niin henkisiä kuin kehollisiakin ilmiöitä ja ”itse” on nimilappu, joka annetaan näille henkisille ja fyysisille ilmiöille koska ne liittyvät toisiinsa ja seuraavat kausaalisesti toisiaan.⁴¹ Pohdin seuraavassa luvussa tarkemmin, mitä tämä kausaalisuus tarkoittaa. Ennen sitä, on kuitenkin

³⁸ Gethin 1998, 135–136.

³⁹ Collins 1982, 94, 150, 153.

⁴⁰ Gier&Kjellberg 2004, 288.

⁴¹ Gethin 1998, 138–139

hyvä tiivistää: theravāda-buddhalaisen opin mukaan itseä ei ole pysyvänä entiteettinä. Valaistumattomalla henkilöllä kuuluukin olla, tunne ”minästä”, asmi-māna, kuvitelma ”minä olen”. Yksilö koostuu osatekijöistä, mutta voidaan tulkita paremminkin toiminnoiksi kuin entiteeteiksi.

1.2 Tapahtumien kausaalisuus ja ehdonvarainen seuraamus

Buddhalaisessa ajattelumallissa yksilö ja hänen olemisensa jaetaan siis fyysisiin ja henkisiin osatekijöihin ja Gethinin mukaan buddhalaisen ajattelun tärkein päämäärä on pyrkiä osoittamaan, että fyysiset ja henkiset tapahtumat (dharma/pāli: dhamma)⁴² ovat toisistaan riippuvaisia. Asiat ja tapahtumat eivät ilmene umpimähkään, vaan kaiken tapahtuvan takana on kausaalinen yhteys. Tapahtumat ovat seurauksia toisista tapahtumista.⁴³ Yksilö koostuu kausaalisista toiminnoista. Samoin tapahtumat yksilön ympärillä tapahtuvat kausaalisesti. Asiat ovat liittyneitä toisiin asioihin.

Miten tämä kausaalisuus sitten ymmärretään? Tarina kertoo kulkija Śāriputran tutustumisesta Buddhan opetukseen. Munkki Aśvajit lausui hänelle säkeen:

Niistä dharmoista, jotka ilmenevät syystä, Tathagata on todennut syyn ja niin ikään lakkaamisen; näin kuuluu suuren askeetin opetus.⁴⁴

Tämän kuullessaan Śāriputra ymmärsi oitis dharman olemuksen. Gethinin mukaan säe kiteyttää Buddhan opetuksen ja paljastaa salaisuuden kärsimyksen lakkaamisen takana. Hän lisää: ”Jos vain ymmärtäisimme sen.”!⁴⁵ Gethin antaa ymmärtää, että Buddhan opetus on avoin tulkinnalle. Ehkä opetus tarkoittaa, että tietoisuuden avulla voidaan tunnistaa syy kärsimykselle – mielihalu. Tietoisuuden avulla yksilö voi myös kukistaa mielihalun ja seurata Buddhan viitoittamaa polkua valaistumiseen.

Kausaalisuuden periaatetta (idampratyaatā) on Gethinin mukaan kuvattu lauseella: ”Kun tämä on olemassa, tuo on olemassa; tämä ilmenee, tuo ilmenee; kun tämä ei ole olemassa, tuo ei ole olemassa, kun tämä lakkaa, tuo lakkaa.” Hän toteaa

⁴² Dharma voi viitata myös universaalien lain luonnolliseen järjestykseen tai buddhalaisten opetuksien kokonaisuuteen, mutta tässä yhteydessä se viittaa elementteihin, joista kokemusmaailma koostuu. Dharmat voivat olla ulkoisia tai sisäisiä henkisiä tapahtumia tai luonteenpiirteitä. Keown 2004, Dharma.

⁴³ Gethin 1998, 140–141.

⁴⁴ “Of those dharmas which arise from a cause, the Tathagata has stated the cause, and also the cessation; such is the teaching of the Great Ascetic.” Gethin 1998, 141, käännös oma.

⁴⁵ “if we could but understand it”, Gethin 1998, 141, käännös oma.

tärkeimmän tavan buddhalaisessa ajattelussa hahmottaa syytä ja seurausta olevan ehdonvaraisen seuraamuksen (pratītya-samutpāda/pāli: paṭicca-samuppāda) ketju ja sen 12 lenkkiä (nidāna) ⁴⁶

Esittelen ja määrittelen seuraavassa, Collinsin kirjan *Selfless Persons* koostetussa, taulukossa 2 ehdonvaraisen seuraamuksen 12 kohtaa. Selvyiden vuoksi olen tehnyt ratkaisun käyttää pālinkielisen termin rinnalla sekä englannin- että suomenkielistä käännöstä.

C	karma (1-2)	↑	(1)	<i>avijjā-paccayā</i>	ignorance	tietämättömyys	Mennyt	
			(2)	<i>saṃskāra</i> (pāli: <i>saṃkhāra</i>)	(mental) formations/ karmic activities	[karmaa (pāli: kamma) aiheuttavat] henkiset muodostelmat	elämä (1-2)	
D	uudelleen- syntyminen (3-7)	↓	(i)	(3)	<i>viññāna</i>	consciousness	tietoisuus	
			(4)	<i>nāma-rūpa</i>	name-and-form	nimi ja muoto		
			(5)	<i>saḷāyatana</i>	the six senses	kuusi aistia		
			(6)	<i>phasso</i>	sense-contact	aisti-kosketus	Nykyinen	
			(7)	<i>vedanā</i>	feeling	tunne	elämä	
			(8)	<i>taṇhā</i>	craving	mielihalu	(3-10)	
A	kamma (8-10)	↑	(ii)	(9)	<i>upādāna</i>	grasping (attachment)	takertuminen	
				(10)	<i>bhava</i>	becoming	(joksikin) tuleminen	
				(11)	<i>jāti</i>	birth	syntymä	
B	uudelleen- syntyminen 11-12	↓	(ii)	<i>jarā-maraṇam</i>			Tuleva	
				(12)	<i>soka-parideva- dukkha- domanass- ūpāyāsā sambhavanti</i>	old age and death, distress, grief, suffering, sorrow and unrest	vanhuus, kuolema, ahdistus, murhe, kärsimys, suru ja rauhattomuus	elämä (11-12)

Taulukko 2: Ehdonvarainen seuraamus⁴⁷

Collins toteaa, että ehdonvaraisen seuraamuksen 12 kohdan lista on nähty myös hämmentävä, koska syntymä on nähty esiintyvän siinä kaksi kertaa; kohdissa 3 ja 4 (tietoisuus ja nimi-ja-muoto) ja kohdassa 11 (syntymä). Theravāda-tradition piirissä tämä on ratkaistu siten, että ehdonvarainen seuraamuksen on nähty itse asiassa viittaavan kolmeen eri elämään. Lista on mahdollista tulkita niin ikään akselilla (i), kohdat 1–6 viitaten elämään mennyt–nykyinen, ja (ii), kohdat 7–12,

⁴⁶ “This existing, that exists; this arising, that arises; this not existing, that does not exist; this ceasing, that ceases.” Gethin 1998, 141, käännös oma.

⁴⁷ Collins 1982, 107, 203–204.

nykyisyys–tulevaisuus. Tässä tulkinnassa koetaan niin aistin kuin tunteenkin edustavan nykyistä hetkeä, ja nykyhetki on olemassa edeltävien tekojen (karma/päli: kamma) tuloksena. Jälleen, ihmisen mielihalu aisti- ja tunnekokemuksia kohtaan pyöryttää karmisen pyörän uudestaan liikkeelle. Collins kertoo ehdonvarainen seuraamusta tulkitun myös jakamalla se neljään kohtaan, yllä esitellyn taulukon mukaisesti kohtiin A, B, C ja D. Mielihalu on jatkuvuuden pääsyy. Osat 8–10, A, karman kehitys, ovat syitä menneisyydessä. Kohdat 11–12, B, ovat vaikutuksia nykyisessä elämässä, uudelleen syntymiseen johtava prosessi. Kohdat 1–2, C, edustavat syitä nykyisyydessä, syntyvää karmaa. Kohdat 3–7, D, edustavat tulevia vaikutuksia, uudelleensyntymää. Collins kuvaa ehdonvaraista seuraamusta kokonaisuudessaan symbolina sille, että elämän pyörä jatkaa kiertämistään riippumatta mistään itsenäisestä toimivasta itsestä.⁴⁸

Näin ehdonvarainen seuraus piirtää linjat sille, miten syy kaikelle tapahtuvalle on itse asiassa kausaalisuudessa. Yksi asia johtaa toiseen⁴⁹. Rähulan modernissa tulkintaa lainaten:

Kun on A, on B; kun A saa alkunsa, B saa alkunsa. Kun A ei ole, B ei ole; kun A lakkaa olemasta, B lakkaa olemasta.⁵⁰

Rähulan tulkinta on perin looginen. Hän käyttää esimerkissään muuttujia A ja B. Koska yllä oleva Collinsin kirjan pohjalta koostetussa taulukossa esiintyvät niin ikään muuttujat A, B, C ja D, nimeän Rähulan kuvaamat muuttujat uudestaan A_1 ja B_1 , sekä edelleen $C_1, D_1 \dots$, jossa A_1 = tietämättömyys, B_1 = henkiset muodostelmat, C_1 = tietoisuus, D_1 = nimi ja muoto, ja niin edelleen. Asia A_1 johtaa asiaan B_1 , joka edelleen johtaa asiaan C_1 , ja niin edelleen. Tämä voidaan myös esittää loogisesti muodossa jos, ja vain jos, A_1 , niin B_1 , jos, ja vain jos, B_1 , niin C_1 , ja niin edelleen. Syllogismin keinoin argumentti voidaan ketjuttaa seuraavasti: joss A_1 , niin B_1 , joss B_1 , niin C_1 ja edelleen, joss C_1 niin D_1 . Siis joss A_1 niin D_1 .

Kuvaan alla ehdonvaraista seuraamusta Max Blackin logiikan opeista ammentaen kaavan muodossa:⁵¹

$A_1 \leftrightarrow B_1$	tietämättömyys \leftrightarrow henkiset muodostelmat
$B_1 \leftrightarrow C_1$	henkiset muodostelmat \leftrightarrow tietoisuus
$C_1 \leftrightarrow D_1$	tietoisuus \leftrightarrow nimi ja muoto
$A_1 \leftrightarrow D_1$	tietämättömyys \leftrightarrow nimi ja muoto

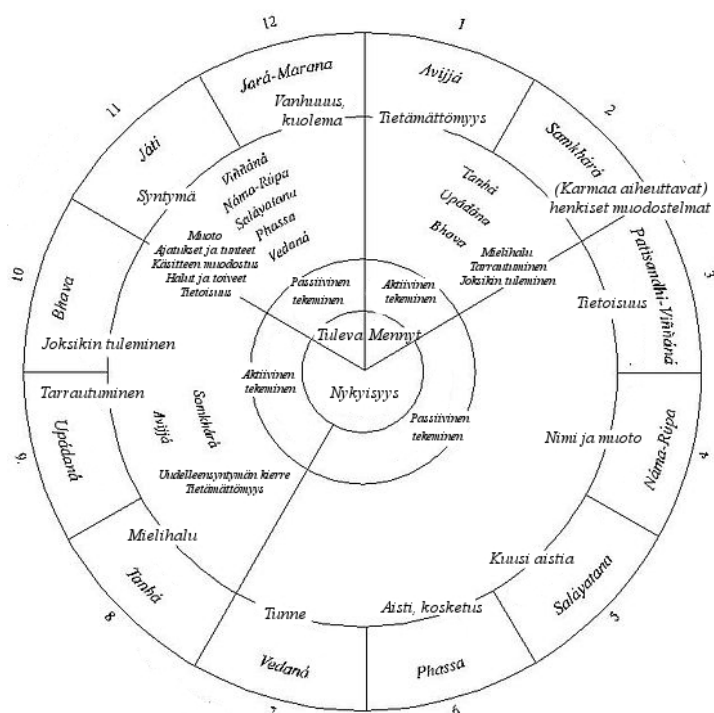
⁴⁸ Collins 1982, 106, 203–204.

⁴⁹ Gethin 1998, 142.

⁵⁰ Collins 1982, 107.

⁵¹ Collins 1982, 106; Black 1952, 85–86.

Tämä tarkoittaa siis sitä, että tietämättömyys on syy henkisten muodostelmien muodostumiselle (saṃskāra/pāli: saṃkhāra). Nämä henkiset muodostelmat johtavat edelleen tietoisuuden muodostumiseen. Ehdonvarainen seuraamus on samalla kuvaus elämän kierrosta ja eteenpäin soljumisesta (saṃsāra) –Keown kuvasi tätä oppia ”ongelmaksi”, johon valaistuminen ja vapautuminen (nirvāṇa) on ratkaisu⁵². Ennen kuin lähdän pohtimaan, miten tämä ratkaisu saavutetaan, kiinnitän huomiota vielä allaolevaan kuvaan 1. Kuvassa on elämän pyörä, joka havainnollistaa ehdonvaraista seuraamusta ja edellä selitettyä kausaalisuutta. Kuva on lainattu buddhalaisten julkaisujen seuran, Buddhist Publication Society:n, verkkosivuilta ”The Buddha and His Teachings”. Olen täydentänyt tiedot suomenkielisiin käännöksiin Steve Collinsin kirjasta *Selfless Persons*.



Kuva 1: Ehdonvarainen seuraamus (praṭītya-samutpāda/pāli: paṭicca-samuppāda)⁵³

1.3 Kärsimyksen lopettaminen; moraalit ja myötätunto

Totesin viime luvussa Damien Keown todenneen saṃsāran olevan ongelma ja nirvāṇan olevan tähän ratkaisu. Tässä luvussa käsittelemme kärsimyksen lopettamisen teemaa buddhalaisessa ajattelussa ja sitä, miten kärsimyksen lopettaminen – valaistuminen – kietoutuu yhteen tietoisuuden ja moraalin käsitteiden kanssa.

⁵² Keown 2004, nirvāṇa.

⁵³ The Buddha and his teachings, käännös oma. Collins 1982, 107, 203–204.

Keskeisen tärkeitä käsitteitä tässä ajatuskulussa ovat Neljä Jaloa Totuutta Buddhan opetuksessa kärsimyksestä ja sen lakkaamisesta sekä Jalo Kahdeksanosainen Polku, ohje valaistumiseen. Gethin, samoin kuin Keown, käsittelee kärsimyksen lopettamisen teemaa seuraavalla tavalla: Buddhalainen ajatusmaailma voidaan tiivistää kahteen perusajatukseseen, (1) Ihmisellä on elämässä perusongelma, kärsimys (duḥkha/pāli: dukkha) ja (2) Buddhan opetus on ratkaisu tähän ongelmaan.⁵⁴

Duḥkha (pāli: dukkha) käännetään usein tarkoittamaan kärsimystä ja kipua, mutta Collins haluaa määritelmässään korostaa, että tarkemmin käännettynä kyseinen termi tarkoittaa turhautuneisuutta ja epätydyttävää. Tällä ei ole tarkoitus kuvata elämän tilaa yleensä, vaan vastakohtaa valaistumiselle. Yksilö ollessaan valaistumaton kokee olonsa epätydyttäväksi. Kärsimystä, turhautuneisuutta tai epätydyttävää oloa ei tule tarkkailla henkilökohtaisesta näkökulmasta, vaan siihen sisältyy kaikkien olentojen kokemus yhtä aikaa.⁵⁵ Elämän kierto on siis toisin sanoen epätydyttävää ja Buddhan opetuksen avulla on mahdollista saavuttaa valaistuminen ja päästä pois tästä epätydyttävästä kehästä.

Kärsimyksen lakkaaminen kietoutuu siis läheisesti ajatukseen moraalisesta toiminnasta ja myös buddhalaisuuden keskeisenä osana olevasta myötätunnosta. Siten kuin itseä ei ole, ”itse” ja ”toinen” tulee nähdä samanarvoisina. Collins siteeraa kirjassaan Louis de La Vallée-Poussinin lausetta kirjasta *The Way to Nirvāṇa*:

Jos buddhalainen läpikäy harjoitukset, jotka johtavat nirvāṇaan – siis harjoitukset, joiden johdosta yksikään uusi olento ei synny hänen tilalleen – hän tekee sen vähentääkseen elävien, kärsivien olentojen määrää yhdellä.⁵⁶

Collinsin mukaan saṃsāran voi nähdä edustavan neljäntyyppisiä olentoja; yksilö, hänen menneet minänsä, hänen tulevat minänsä ja olennot, jotka ovat vain ”toisia”. Koska yksilön on mahdoton tietää, kuka tai mikä on suhteessa häneen, ja kuka tai mikä on toinen, niin buddhalainen filosofia neuvoo suhtautumaan kaikkiin olentoihin samoin, rakastavalla ystävällisyydellä, myötätunnolla, sympaattisella ilolla ja mielentyyneydellä.⁵⁷ Ajatuksen voi nähdä heijastuvan mielekkäästi John Rawlsin näkemykseen, jossa (hyvien) päätösten taustalla ovat oikeat olosuhteet.

⁵⁴ Gethin 1998, 59.

⁵⁵ Collins 1982, 191.

⁵⁶ “If a Buddhist undergoes the discipline which leads to nirvāṇa - that is, the discipline owing to which no new being is to be born in his stead – it is in order to diminish by one the number of of living and suffering beings.” Collins 1982, 193, käännös oma.

⁵⁷ Collins 1982, 193.

Oikeissa oloissa omien etujen ei tulisi tulla järkevän, ymmärryksellä tehdyn, ja vapauden, tasa-arvon ja palkkion huomioivan päätöksen tielle.⁵⁸

Buddhan opetus tiivistyy neljään jaloon totuuteen⁵⁹; kärsimyksen olemus, sen syyt, lakkaaminen ja tapa saada kärsimys loppumaan. Nämä on toisinaan ymmärretty myös ikään kuin diagnoosina, sairaus, sairauden syy, parannuskeino sekä lääke.⁶⁰

Kun ylevä oppilas on ymmärtänyt kärsimyksen, kärsimyksen alkuperän, kärsimyksen lakkaamisen ja polun, joka johtaa kärsimyksen lakkaamiseen... hän tässä ja nyt tekee lopun kärsimyksestä.⁶¹

Kysymys on, mikä on tapa saada kärsimys loppumaan. Majjhima Nikāya -tekstissä Buddhan kerrotaan todenneen: ”Ja mikä johtaa kärsimyksen lakkaamiseen? Jalo kahdeksanosainen polku...⁶²” Jalo kahdeksanosainen polku sisältää suuntaviivat kärsimyksen lakkaamiseen; oikeat näkemykset, päätökset, puhe, toiminta, elinkeino, pyrkimys, mielentila ja oikea meditaatio. Polku voidaan jakaa moraaliin, meditaatioon ja viisauteen, minkä on tarkoitus ilmentää ymmärryksen, toiminnan ja tunteiden välillä olevaa suhdetta. Sekä käytös, tunteet että toiminta vaikuttavat päämäärän saavuttamiseen, kärsimyksen lakkaamiseen.⁶³

Damien Keown kertoo jalon kahdeksanosaisen polun olevan lyhennetty versio matkasta valaistukseen ja esiintyvän ensin kertaa pāli-kaanonin yhdestä vanhimmista osuuksista, pitkien puheiden tekstissä (Dīgha Nikāya) moraalisten käytäntöjen kokoelmassa (Sīlakkhandhavagga) puheessa VI, Puhe Mahātille.

Moraalisten käytäntöjen alkupäässä esiintyy Keown'n Jalon Kahdeksanosaisen Polun edeltäjäksi tulkitsema 13-osainen matka valaistukseen⁶⁴ ja tämä on jaettu kolmeen kategoriaan, moraalisuus (sīla), meditaatio (samādhi) ja

⁵⁸ Päivänsalo 2007, 8.

⁵⁹ Rupert Gethin haluaa kuitenkin painottaa, että huolimatta termistä ”totuus”, lukijan ei tule tulkita Neljää Jaloa Totuutta ikään kuin buddhalaisena uskontunnustuksena. Haasteena tulkinnalle on sana ”totuus”, satya (pāli: sacca), joka voi merkitä myös ”todelliset asiat/realiteetit”. Näin ollen, neljä totuutta ovat ikään kuin elämän tosiasioita, jotka pätevät yhtä hyvin niin ’buddhalaiselle’ kuin ’ei-buddhalaiselle’. Buddhalaisuuteen ei kuulu seurata näitä ’totuuksia’ orjallisesti, vaan yksinkertaisesti pyrkiä näkemään asiat niin kuin ne ovat. Gethin 1998, 60.

⁶⁰ Gethin 1998, 59, 63.

⁶¹ “When a noble disciple has thus understood the suffering, the origin of suffering, cessation of suffering, and the way leading to cessation of suffering... he here and now makes an end of suffering.” Nanamoli & Bodhi 1995, 135, käännös oma.

⁶² “And what is the way leading to the cessation of suffering? It is just this Noble Eightfold Path...” Nanamoli & Bodhi 1995, 135, käännös oma.

⁶³ Gethin 1998, 81; Keown 1992, 35.

⁶⁴ Ensimmäinen sutta, Puhe brahmaanin verkosta (Brahmajāla Sūtra/pāli: Brahmajālasutta), sisältää kolme traktaattia moraalisesta käytöksestä, listoja asioita, joista tulee pidättäytyä. Moraalisten käytäntöjen kokoelman toisessa sutassa Uskonnollisen elämän hedelmät (Sāmaññaphalasutta) moraaliset normit ensimmäisestä sutasta yhdistyvät yleiseksi malliksi kohti pelastusta. Moraaliset hyveet johtavat neljän harjoituksen (dhyana/pāli, jhāna) kautta täydellistymiseen (Arahat).

viisaus (prajñā/pāli: paññā), mikä toistuu myös jalon kahdeksanosaisen polun kategorisoinnissa. Keown'n tulkinta on, ettei polun yksittäisiä osia ilmeisesti nähdä niin tärkeitä, kuin yleistä kolmea kategoriaa, joihin polku on jaettu. Kategoriat osoittavat, millä alueilla kehitystä tarvitaan, kun taas yksittäiset säännöt osoittavat tarkemmin miten toimia ja mistä pidättäytyä. Keown esittää myös näkemyksen, jonka mukaan tietä täydellistymiseen olisi mielekkäintä tutkailla kaksiosaisen mallin valossa, moraali (sīla) ja viisaus (prajñā/pāli: paññā). Meditaatio tulisi tulkita vain välineeksi edetä näissä hyveissä.⁶⁵ Seuraava jaloa kahdeksanosaista polkua kuvaava taulukko 3 on koottu yhdistäen taulukko Steve Collinsin kirjasta *Selfless Persons* ja Rupert Gethinin kirjasta *The Foundations of Buddhism*.

sammā	ditṭhi	(1) oikea näkemys	neljän jalon totuuden näkeminen	pañña – viisaus
	saṅkappa	(2) päätökset	<ul style="list-style-type: none"> • mielihalun kieltäminen, • ystävällisyys, • myötätunto. 	
	Vācā	(3) puhe	pidättäytyminen: <ul style="list-style-type: none"> • valheellisesta, • erimielisyyttä aiheuttavasta, loukkaavasta puheesta ja • joutilaasta jaarittelusta. 	sīla – moraalisuus
	kammanta	(4) toiminta	pidättäytyminen: <ul style="list-style-type: none"> • elävien olentojen satuttamisesta • sen ottamisesta mitä ei ole annettu • sukupuolisista rikkomuksista. 	
	Ājīva	(5) elinkeino	joka ei perustu väärään puheeseen ja toimintaan.	
	Vāyāma	(6) pyrkimys	<ul style="list-style-type: none"> • ilmenemättömien turmelevien (unwholesome) tilojen estäminen, • ilmenevien turmelevien tilojen hylkääminen, • ilmenemättömien turmeltumattomien (wholesome) tilojen herättäminen, • herätettyjen turmeltumattomien tilojen kehittäminen. 	samādhi – meditaatio
	Sati	(7) mielentila	<ul style="list-style-type: none"> • kehon, • tunteen, • mielen ja • dharman mietiskely 	
	Samādhi	(8) keskittyneisyys	• neljän <i>dhyānan</i> ⁶⁶ harjoitus	

Taulukko 3: Jalo kahdeksanosainen polku⁶⁷

⁶⁵ Keown 1992, 35–38.

⁶⁶ Dhyāna (Pāli: jhāna); mietiskelyssä tavoitettu transsitila. Näitä on opin mukaan kahdeksan, joista neljä ensimmäistä ovat osa muodon maailmaa (realm of form) ja toiset neljä osa hahmotonta maailmaa (formless realm). Neljä korkeinta transsitilaa 5–8 ymmärretään myös jakautuneeksi neljännessä, viimeisestä muodon maailman transsitilasta. Derris 2004, 226

⁶⁷ Gethin 1998, 81–82; Collins 1982, 90.

Edellisessä luvussa keskusteltu Keown'n näkemys täydellistymisestä viisauden ja moraalin avuin saa tukea moraalisten käytäntöjen kokoelman puheesta IV, Puheesta Soṇaṇḍaṇḍalle. Puhe Soṇaṇḍaṇḍalle jatkaa lopullisen hyvän ja moraalin sekä viisauden vastavuoroisuuden määrittelyä. Tässä bramaani Soṇaṇḍaṇḍa määrittelee Buddhan johdattamana todellisen brahmaanin piirteet ja tulee johtopäätökseen, että tällaisia on vain kaksi, hyve (sīla) ja ymmärrys (prajñā/pāli: paññā). Keown'n mukaan puheesta IV voidaankin päätellä, että moraali on keskeinen osa ihmisen täydellistymistä. Buddhalaisen ajatusmallin tunnustamat hyvät piirteet ovat lähtöisin kolmesta päähyveestä: Suvaitsevaisuus, hyväntahtoisuus ja ymmärrys. Harhaluulot, viha ja kateus puolestaan muodostavat vastakohtan täydelliselle mielelle.⁶⁸

Buddhalaisessa ajattelussa pahe (niin kuin hyvekin) jakautuu moraaliseen ja älylliseen. Moraalisista paheista pahimmat ovat ahneus (rāga, lobha) sekä viha (dosa) Älyllisistä paheista vakavin on harhaisuus (moha). Sekä moraaliset että älylliset paheet kietoutuvat mielihalun (taṇhā/pāli: tṛṣṇā), käsitteen ympärille, edustaen sen kahta puolta.⁶⁹ Padmasiri De Silva on todennut kärsimyksen luonteesta buddhalaisessa ajatusmaailmassa: ”Kärsimys juontaa joko harhakäsityksistä tai mielihaluista⁷⁰.” Edelleen, Keown kirjoittaa mielihalusta juontavan tarrautumisen olevan kärsimyksen päälähde. Hän toteaa myös, että vaikka älylliset ja moraaliset paheet voi teoriassa erottaa toisistaan, näin ei voi tehdä mielen sisällä. Keown'n selittää asian seuraavasti: järki ja tunne ovat keskinäisessä vuorovaikutuksessa ja tietämättömyys⁷¹ ja mielihalu limittyvät. Keown toivookin voivansa osoittaa niin moraalisen kuin älyllisenkin täydellistymien olevan keskeisiä buddhalaisessa ihanteessa, kykenevyyden moraaliseen mielipiteeseen ollessa olennainen osa ihmisluontoa.⁷² Mielihalu, buddhalaisuuden keskeinen ongelma valaistumisen kannalta, on tulosta tunteiden ja käsityskyvyn yhtäaikaisesta väärästä toiminnasta. Mielihalun vallassa olevan henkilön koetaan olevan sekä harhainen oikeasta oikeaa laitaa koskien sekä henkisesti kiinni väärinkäsityksessä. Epämoraalinen käytös ei näin siis juonnu pelkästään tietämättömyydestä tai

⁶⁸ Keown 1992, 38–39.

⁶⁹ Keown 1992, 64–65.

⁷⁰ De Silva 2005, 46.

⁷¹ Tietämättömyys (avijjā) sisältää sekä älyllisen puutteellisuuden että ennakkoluuloisuuden, ja näin sisältää motivaation puutteen buddhalaisen ajatusmaailman mukaiseen hyveelliseen toimintaan.

⁷² Keown 1992, 65–66.

tunnepuolen epätasapainosta, vaan asioiden väärinkäsityksestä yhdistettynä tunteisiin, jotka ovat nousseet tämän väärinkäsityksen johdosta. Keown toteaa buddhalaisuuden ollessa kyseessä yleensä oletettavan ongelman olevan tiedon puute. Tämä oletus kuitenkin aliarvioi buddhalaista näkemystä tunteiden voimasta hallita ja manipuloida järkeä. Tunnepuolen ja käsityksen virhetoiminta ilmenee mielihaluna ja tietämättömyytenä, jotka buddhalaisessa ajattelussa mielletään paheiksi.⁷³

Keown toteaa pelastuksen perimmäinen ongelman buddhalaisesta näkökulmasta voitavan kuvata, miten puhdistaa mieli kaikesta pahuudesta. Gethiniä lainaten, buddhalainen ajatusmalli näkee mielen voivan tahraantua useasta tekijästä; ahneus, viha ja harhaluulot. Näille on kuitenkin vastatekijät, eittarrautuvuus (alobha), ystävällisyys (adveṣa/pāli: adosa tai maitrī/pāli: mettā) ja viisautus (prajñā/pāli: paññā). Buddhalaisuuden opin yksi keskeisistä tavoitteista on kehittyä näissä hyvissä ominaisuuksissa.⁷⁴ Tiivistäen edellä keskusteltua, Buddhan opetus on siis ratkaisu elämässä ilmenevään ongelmaan, kärsimykseen. Kärsimyksen päälähte ovat mielihalu ja tietämättömyys. Ongelma on mahdollista ratkaista seuraamalla Jaloa Kahdeksanosaista Polkua, täydellistyä sekä moraalissa, että viisaudessa. On yhä huomionarvoista, että koska – Collinsin mukaan – Buddhan opetus ei tunnusta pysyvää itseä, yksilön ei voi erottaa itseään toisesta. Kuten jo mainitsin, buddhalainen filosofia neuvoo suhtautumaan kaikkiin olentoihin samoin, rakastavalla ystävällisyydellä, myötätunnolla, sympaattisella ilolla ja mielentyyneydellä. Keskustelen seuraavassa myötätunnon käsitteen roolista theravāda-buddhalaisessa opissa.

Keown pohtii, sisältääkö buddhalaisuus käsitettä toisten huomioon ottamisesta ja oliko Buddhan eettinen täydellistymisessä moraalisen huolen tuntemista. Moraalisella huolella Keown'n tarkoittaa ei-itseen liittyvää huolta toisten hyvinvoinnista ja tuntemuksella ei-tiedollista tilaa, vastakohtana tiettyjen sääntöjen tiedolliselle hyväksymiselle. Keown esittää, että Buddha oli omaksunut tällaisen moraalisen huolen, mikä oli nähtävissä hänen sympatiassaan (anukampā) kaikkea elävää kohtaan. Tämä tiedostavuus ei ollut tulosta valaistumisesta vaan edelsi sitä. Buddhan itse on kerrottu käyttäneen termiä sympatia ja sen vastineita useasti kuvatessaan motivaatiotaan valaistua. Keown kertoo Buddhan niin ikään rohkaisseensa munkkeja opettamaan toisia, jotta myös nämä voisivat saavuttaa

⁷³ Keown 1992, 67, 69–71.

⁷⁴ Gethin 1998, 80; Keown 1992, 57.

valaistumisen. Huoli lähimmäisestä ja *sympatia* kaikkia eläviä olentoja kohtaan olivat Keown'n mukaan keskeinen piirre Buddhan luonnossa. *Sympatia* käsitetään tässä ei-rationaalisenä tuntemuksena, joka edeltää moraalisen päämäärien syntymistä, eikä järkipäisenä motiivina moraalisuudelle. Jos moraalisen toiminnan motiivina on henkilökohtaisen arvon nousu, tekoa ei ole saanut aikaan *sympatia*. Moraalisuus ei näin ole väline vaan tarkoitus itsessään ja niin ikään osa valaistumista.⁷⁵ Edellä oleva korostaa aikomuksen merkitystä theravāda-buddhalaisessa ajattelussa. Aikomuksen merkitys toiminnassa osana jatkuvuutta (ja sen loppumista) onkin keskeisen tärkeä. Keskustelen aikomuksen merkityksestä syvemmin seuraavassa pääluvussa käsitellessäni theravāda-buddhalaista ymmärrystä tietoisuuden luonteesta.

Collinsin mukaan theravāda-lähteet kuvaavat Buddhan myötätuntoa samoin siten, että eläessään vielä edellisessä elämässään nimellä Sumedha⁷⁶, hän päätti tulla erityiseksi Buddhaksi (Supreme Buddha) auttaakseen kaikkia olentoja valaistumaan. Buddhan syntymä on ”yhden ihmisen” syntymä, joka tuo mukanaan

Monien hyvän, monien onnellisuuden ja joka on syntynyt myötätunnosta tähän maailmaan, ihmisten ja jumalien hyvinvoinnin, hyvän ja onnellisuuden puolesta.⁷⁷

Collins kertoo Buddhan välittömästi valaistumisensa jälkeen epäröineen, tulisiko hänen opettaa ihmisiä, mutta päätös opettaa buddhalaista oppia syntyi myötätunnosta kaikkia olentoja kohtaan. Hän opasti seuraajiaan opettamaan monien hyväksi, monien onnellisuudeksi, myötätunnosta.⁷⁸

Myötätunto käsitteenä nivoutuu buddhalaisuudessa etenkin Buddhan opettamaan Neljään Jaloon Totuuteen, joiden tarkoituksena on kärsimyksen lopettaminen. Toive parantaa kärsimys yhdistyy Gethiniä lainaten *sympatian* (*anukampā*) tunteeseen, sekä myötätuntoon omaa tai toisten kärsimystä kohtaan. *Sympatian* tunne kaikkia kärsiviä kohtaan ei motivoi yksin Buddhan toimintaa vaan kaikkia buddhalaisuuden oppien mukaan eläviä.⁷⁹

⁷⁵ Keown 1992, 72–75

⁷⁶ Sumedhan kerrotaan olleen Buddhan nimi hänen eläessään Bodhisattvana, valaistusta kohti kulkevana yksilönä, Buddha Dīpaṅkaran aikana. Kerrotaan Dīpaṅkaran ennustaneen Sumedhan saavuttavan valaistumisen Siddhartha Gautamana. Keown 2004.

⁷⁷ “The good for man, the happiness of the many, (who is born) out of compassion for the world, for the welfare, good and happiness of gods and men.” Collins 1982, 194, käännös oma.

⁷⁸ Collins 1982, 194.

⁷⁹ Gethin 1998, 64.

1.4 Vapaa tahto buddhalaisessa ajattelussa

Haluan tässä luvussa selvittää, mikä asema yksilön vapaalla tahdolla oli theravāda-buddhalaisessa opissa ja miten vapaa tahto nähtiin osana oikein toimimista. Aiheesta kirjoittaneiden tutkijoiden käsitykset vaikuttavat toisistaan eriäviltä. Kysymykset käsitteiden, determinismin, kompatibilismin ja fatalismin, määrittelystä ovat merkittävässä osassa tulkitessa kysymystä tahdon vapaudesta theravāda-buddhalaisessa ajattelussa. Ensin käsittelen kuitenkin erästä kiinnostavaa näkökohtaa. Länsimainen käsitys tahdon vapaudesta ja theravāda-buddhalainen vastaava käsitys eivät ole tutkijoiden mukaan toisiinsa suoraan verrattavissa ja näenkin mielekkääksi keskustella tästä näkemyksestä.

Nicholas F. Gier ja Paul Kjellberg kirjoittavat artikkelissaan ”Buddhism and the Freedom of Will: Pali and Mahayanist responses” ettei kysymys vapaasta tahdosta osoittautunut merkittäväksi dilemmaksi Buddhan opetuksissa. Heidän mukaansa menneet ajattelijat nykyisen Intian alueella olivat keskeisesti kiinnostuneempia vapaan henkisyuden (spirituality) kuin vapaan tahdon kysymyksestä. Buddhalaiselle ajattelulle, toisin kuin modernistiselle näkemykselle, ei ollut ominaista jakaa kokemusmaailmaa ulkoiseen ja sisäiseen, mikä tuottaa ristiriidan sisäisen vapauden ja ulkoisen kausaalisuuden välille, ja juuri tämän Gier ja Kjellberg tarjoavat erääksi syyksi vapaan tahdon käsitteen ongelman ohittamiselle. Vapaan tahdon käsitteestä sellaisena, kuin viimeaikainen tutkimus sen tuntee, keskustellaankin ensimmäistä kertaa kirkkoisä Augustinuksen teksteissä.⁸⁰ Augustinusta on myös kutsuttu vapaan tahdon, *voluntas libera*, tai tarkemmin, tahdon vapaan päätöksen (free decision of the will), *liberum arbitrium voluntatis*, käsitteen keksijäksi.⁸¹ Nicholas F. Gier ja Paul Kjellberg rohkenevat kutsua tätä asetelmaa jopa ironiseksi. Kannattihan Augustinus heidän mukaansa lopulta näkemystä, jossa valta ihmisen tahdon yli oli lopulta Jumalalla⁸², mikä väitteenä poistaa yksilöltä tahdon vapauden.

⁸⁰ Gier & Kjellberg 2004, 278–280.

⁸¹ Saarinen 2011, 19.

⁸² Augustinuksen näkemys on kuitenkin se, että vaikka valta olikin Jumalalla, tämä ei ollut ristiriidassa ihmisen vapaan tahdon kanssa. Todellinen vapaus on tehdä, kuten Jumala haluaa. Risto Saarisen mukaan, Augustinuksen näkemys oli vapauden olevan vapautta toimia tai olla toimimatta. Ihmisen kykyä toimia oikein häiritsevät Augustinuksen kirjoituksissa kaksi epätäydellistä tahtoa. Juurtuneet tavat estävät kohti oikeaa pyrkimisen. Jumalan armoon heittäytyminen parantaa ihmisen tahdon, jolloin ihminen vapautuu haitallisista, häiritsevistä mielihaluista. Augustinuksen Tunnustuksista 9, 1,1; ”Ja tässä oli tulos: en enää tahtonut tehdä, mitä tahtoinkin ja aloin tahtoa tehdä, mitä Sinä tahtoit. Mutta vapaa valintani oli olemassa kaikki ne vuodet ja mistä synkästä ja salaisesta piilopaikasta se kutsuttiin esiin yhdessä hetkessä.”

Myös Asaf Federman toteaa vapaan tahdon dilemman olevan keskiaikaiseen teologiaan juontava eurooppalaisen ajattelun tuotos. Myöhemmin myös René Descartes artikuloi kirkkoisä Augustinuksen osoittaman suunnan mukaisen johtopäätöksen vapaata tahtoa koskien: järki on Jumalan lahja ihmiselle, ja mitä enemmän järki vaikuttaa toimintaan, sitä vapaampaa toiminta on. Kristinuskon Jumalan rooli on näin ollut merkittävä tekijä tahdonvapauden käsitteen muodostumisessa. Tästä syystä esimerkiksi Federman väittää, että tämä läntinen näkökulma ei ole lainkaan sovellettavissa buddhalaisuuteen ajatusmalliin, tämän kieltäessä yksittäisen hallitsevan jumaluuden olemassaolon. Federman esittääkin modernin vapaan tahdon käsitteen ongelman olevankin seuraava: samalla, kun uskotaan luonnonlakien kausaalisuuteen, uskotaan toimijoiden, jotka kuitenkin ovat osa tätä luonnon järjestystä, pystyvän nousemaan tämän kausaalisuuden yläpuolelle vapaan tahtonsa avulla. Muun muassa kartesiolainen ajattelu ja kombatilismi ovat pyrkineet ratkaisemaan, miten nämä kaksi ajatusta voidaan sovittaa yhteen.⁸³ Tutkijoiden mukaan theravāda-buddhalaista käsitystä ei vapaasta tahdosta ei siis voi käsitellä samoin taustaoletuksin kuin länsimainen käsitys vapaasta tahdosta ymmärretään. Pohdin seuraavassa luvussa, miten vapaa tahto sitten voitaisiin ymmärtää. Näkemyksiä ovat pohtineet Asaf Federman, B. Allan Wallace sekä Nicholas Gier yhdessä Paul Kjellbergin kanssa. Seuraavissa alaluvuissa pohdin heidän näkemyksiensä valossa vapaan tahdon luonnetta Buddhan opetuksessa. Keskustelen determinismistä ja kompatibilismista ja lopulta, mitä Buddha opetti vapaasta moraalisesta toiminnasta.

1.4.1 Vapaa tahto, determinismi ja kombatilismi

Asaf Federman argumentoi, että jotta voitaisiin mielekkäästi määritellä miten buddhalainen ajatusmaailma suhtautuu vapaan tahdon käsitteeseen, täytyy ensin selvittää, mitä vapaan tahdon käsitteellä tarkoitetaan. Pitää siis määritellä, onko vapaa tahto jotain kaikista olosuhteista, syistä ja seurauksista riippumatonta vaiko vastakohta fatalismille. Vapaan tahdon käsite voidaan Federmanin mukaan määritellä kahdella tavalla; vapaa tahto on sielulle ominainen, kehoa täysin hallitseva, voima (VT1) tai vapaa tahto on toimijan kyky hallita toimintaa tahdon

⁸³”And this was the result: Now I did not will to do what I willed, and began to will to do what thou didst will (*nolle quod volebam et velle quod volebas*). But there was my free decision during all those years and from what deep and secret retreat was it called forth in a single moment (*evocatum est in momento liberum arbitrium meum*). Saarinen 2011, 19–22, käänös oma.

⁸³ Federman 2010, 1, 3; Gilbert 2005, 210–214.

mukaisesti, silloin kun tälle ei ole estettä (VT2).⁸⁴ VT1-määritelmä pitää sisällään piirteitä, jotka ilmenivät niin René Descartesin ajattelussa kuin muinaisen Intian Upaniṣad⁸⁵-teksteissäkin. Sielu on olemisen ydin ja sisäinen hallitsija. Se hallitsee kehoa ollen siitä silti erillinen. Buddha kuitenkin kielsi sielun olemassaolon ja samalla varsin määrätietoisesti tällaisen vapaan tahdon näkemyksen, jossa *Ātman*, itse tai sielu, on sisäinen hallitsija ja havaitsemisen keskus. Näin ollen, Buddha kielsi opetuksissaan VT1:n, mutta hyväksyi aatteen, VT2, jonka mukaan yksilöillä on vapaa tahto, kun he voivat hallita toimiaan tahtonsa mukaan, kun tälle ei ole esteitä.⁸⁶

Kirjoittajat Asaf Federmanin ja B. Allan Wallace vaikuttavat olevan tutkimusartikkeleissaan erimielisiä keskeisen käsitteen, determinismin, luonteesta Buddhan opetuksessa. Tämä edellyttää käsitteen analysointia ennen varsinaisia johtopäätöksiä. Federmanin mukaan buddhalaiset tekstit tukevat ajatusta, ettei Buddha kieltänyt determinismia ilmiönä⁸⁷, kun taas Wallace näkee Buddhan kieltäneen determinismin sen tukiessa ajatusta, ettei ihminen ole vastuussa teoistaan⁸⁸. Vaikuttaakin siltä, että se minkä Wallace määrittelee determinismiksi, on Federmanin määritelmän mukaan fatalismia.

Federman tarjoaa determinismille seuraavan määritelmän: determinismi on ”väite, jonka mukaan menneisyys määrittelee ainutkertaisen tulevaisuuden⁸⁹”. Toisin sanoen, jokaista tekoa seuraa yksi tietty ennalta määrätty, järkkymätön tapahtuma tulevaisuudessa. Fatalismin Federman jälleen määrittelee eettiseksi asennoitumiseksi, jonka mukaan valinta on merkityksetön. Toiminnalla ei ole merkitystä, ja siksi on turhaa edes yrittää. Federman perustelee ajatustaan, että Buddha hyväksyi determinismin, buddhalaisen opetuksen kausaalisuutta korostavilla opeilla, näistä hän mainitsee esimerkkinä ehdonvaraisen seuraamuksen teorian. Federman kuitenkin korostaa, ettei ehdonvarainen seuraamus voi olla ehdoton kaikissa tilanteissa. Tuntemusten (sensations) ei esimerkiksi välttämättä tarvitse herättää mielihalua (craving). Lopuksi hän toteaa näkemyksensä, jonka mukaan theravāda-tekstien nojautuminen determinismiin on merkki laajemmasta

⁸⁴ Federman 2010, 3.

⁸⁵ Vedat, Brāhmaṇat ja Upaniṣadit: osa korpuksesta, joka Steve Collinsin mukaan kertoo sanskritinkielisen kulttuurin leviämisestä toisella vuosituhannella eKr. nykyisen Intian alueella. Collins 1982, 30.

⁸⁶ Federman 2010, 3–5, 11.

⁸⁷ Federman 2010, 11.

⁸⁸ Wallace 2011, 219.

⁸⁹ ”[t]he past determines a unique future”. Federman 2010, 11, käännös oma.

tavasta, jolla buddhalaisessa ajattelussa nähdään todellisuus deterministisen kausaalisuuden säätelemänä; ajattelu ei ole lainkaan niin jyrkkää, että yhdestä syystä syntyy vain yksi ainoa seuraus. Kausaalisuuden teoria ei niin ikään sodi lainkaan karman (pāli: kamma) teorian kanssa. Vaikka syyn ja seurauksen suhde ei ole tiukan deterministinen, tämä ei Federmanin mukaan kyseenalaista determinismia lainkaan. Lopputulos on ennalta määritelty, mutta ei vain yhdestä syystä riippuen, vaan tähän vaikuttaa lukuisten eri syiden yhdistelmä. Federman lainaa kuitenkin David Kalupahanan argumenttia, että tämä ei silti suinkaan tarkoita, että Buddhan käsitys kausaalisuudesta olisi ollut enemmän ehdollinen (conditional) kuin deterministinen, vaan tekojen vaikutus (karma/pāli: kamma) on vain yksi tuleviin tapahtumiin vaikuttavien tekijöiden lomassa. Buddhalainen ajattelu ei näin kiellä kausaalisuuden eittämättä determinismistä teoriaa, vaan fatalismin, jonka mukaan teot ovat merkityksettömiä. Wallace on samoilla linjoilla todetessaan Buddhan hylänneen väärinä kaikki teoriat, jotka millään tavoin horjuttivat ihmisen tajua moraalaisesta vastuusta tai viekoittelivat pois hyveen harjoituksesta. Vaikka determinismin käsite voi aiheuttaa hämmennystä, sekä Federman että Wallace toteavat ykskantaan Buddhan myös kieltäneen tämän vastakohtan, indeterminismin, sattumanvaraisuuden (addhiccāsamupanna-vāda). Maailma ei ole buddhalaisessa ajattelussa sattumanvarainen.⁹⁰

Gier ja Kjellberg tulkitsevat buddhalaisen ajattelun ehdollista determinismia ehdollisuuden (conditionality) käsitteellä. He tulkitsevat myös ehdonvaraisen seuraamuksen ehdolliseksi samoilla perusteilla, kuin Federman tulkitsee sen löyhästi kausaaliseksi ja deterministiseksi. Yksittäiset asiat ei sanele yhtä yksittäistä seurausta, vaan kaikki edeltävät asiat vaikuttavat seuraamuksiin. Ehdonvaraisen seuraamuksen ketjusta esimerkin ottaaksemme, jos yksi kahdestatoista lenkistä jää puuttumaan, uudelleensyntyminen ei tapahdu. Ehtojen on siis toteuduttava täysimääräisesti, jotta tietty tapahtuma toteutuu. Gier ja Kjellberg lainaavat Edward Conzen esimerkkiä; luoti on aiheuttanut miehen kuoleman. Mutta tapahtuiko näin? Eihän miehen kuolemaa aiheuttanut vain luoti. Sen voidaan pohtia aiheuttaneen myös ase, ampujan halu ampua mies, aseiden valmistaja... Pohdinnoissa voidaan edetä niin pitkälle, että on mahdollista todeta, ettei ampuja ei olisi nähnyt tähdätä ilman auringon valoa. Auringon täytyy näin olla syy myös.⁹¹

⁹⁰ Wallace 2011, 219, 224; Federman 2010, 11–13.

⁹¹ Gier & Kjellberg 2004, 285–286.

Kompatibilismin aate yhdistettynä buddhalaiseen näkemykseen vapaasta tahdosta on jakanut siteeraamieni tutkijoiden ajatuksia. Kompatibilismissa on kyse siitä, sopiiko ajatus determinismistä yhteen ajatuksen vapaasta tahdosta kanssa. Asaf Federman näkee Buddhan hyväksyneen kompatibilistisen moraalisen valinnan sallivan vapaan tahdon. Federman korostaa toimijan olevan vapaa, VT2:n mukaisesti, hallitessaan toimintaansa tahtonsa mukaisesti vailla toimintaa haittaavia pakotteita, esteitä tai pakkomielteitä. Hän liittyy buddhalaisuuden näkemykset vahvasti Daniel Dennettin filosofiaan. Dennett tuo kompatibilistiseen vapaan tahdon teoriaan kognitiivisen näkökulman, jonka mukaan vapaa tahto on itsekontrollia, johon liittyy joukko erityislaatuisia tiedollisia kykyjä; toimija voi esittää, pohtia ja kuvitella mahdollisuuksia. Vapaa tahto on näin ”epätäydellinen kyky ohjata itseä kohti kuviteltua maalia”.⁹²

B. Alan Wallace on eri mieltä kompatibilismista eikä hän hyväksy Dennettin näkemyksiä ihmisen toimijana vapaudesta. Se, ettei Wallace allekirjoita kompatibilismin aatetta tulee ymmärrettäväksi, kun muistetaan, ettei Wallace myöskään ajatellut Buddhan hyväksyneen determinismiiä. Eritoten Wallace vastustaa Dennettin näkemystä ihmisen vapaudesta. Dennettin näkemyksiin kuuluu ajatus, että ihminen soluista koostuva mekanismi, jokaisen solun ollessa tiukasti fysiikan lakien mukaan toimiva mekanismi. Dennett ei Wallacen mukaan selitä, miten näistä ymmärrystä vailla olevista soluista pystyy syntymään ihminen vapaana toimijana. Wallace huomauttaakin, että Dennettin lailla ajattelevat ajattelijat vaikuttavat älyllisesti vakuuttaneen itsensä, että ihmiskunnan olemassaolo ja todellisuus voidaan selittää tieteen avulla, mutta samalla tuntevat pakkoa todentaa moraalinen vastuu, jota ilman ihmiskunnan sivilisaatiota ei mielekkäästi voi olla olemassa. Ergo, he liittyvät moraalin väkinäisesti atomien ja solujen deterministiseen olotilaan. Wallace toteaa: ”Tästä pulmasta seuraa vain huonoa tiedettä ja filosofiaa.”⁹³

Tutkijoista Federman hyväksyy negatiiviselta vapaudelta vaikuttavan vapauden määrittelyn VT2. Toimija voisi siis hallita tekemisiään, kun tälle ei ole estettä. Federman hyväksyy determinismin ja myös ajatuksen kompatibilismista. Gier ja Kjellberg pohtivat samansuuntaisesti, mutta tulkitsevat tapahtumat ehdollisiksi. Wallace taas on eri mieltä sekä determinismistä että kompatibilismista

⁹² Wallace 2011, 222, Federman 2010, 8–9, 15.

⁹³ “This conundrum makes for bad science and bad philosophy.” Wallace 2011, 222–223, käänös oma.

yhdistettynä vapaan tahdon ajatukseen. Miten vapaus siis voisi selittyä Wallacen näkemyksen mukaan? Selvitän tätä vapauden mahdollisuutta seuraavassa luvussa, joka käsittelee Buddhan opetusta vapaasta toiminnasta.

1.4.2 Buddhan opetus vapaasta moraalisesta toiminnasta

B. Alan Wallace lähtee pohtimaan vapaata tahtoa Nicholas Maxwellin tarjoaman määritelmän avulla. Tämän mukaan vapaus on mahdollisuus saavuttaa arvokkaita asioita olosuhteiden sallimissa rajoissa. Wallace pohtii, että buddhalaisen ajattelun näkökulmasta yksilöillä ei käytännöllisesti katsoen ole vapaata tahtoa saavuttaa arvokkaita asioita niin kauan kuin tietämättömyys ja siitä johtuvat henkiset koettelemukset rasittavat heidän mieltään. Hän jatkaa Buddhan kuitenkin todenneen, etteivät tällaiset taakat kuulu yksilön luontoon synnynnäisesti, vaan ne voi hälventää henkisten harjoitusten avulla. Buddhalainen tradition painotus on, että epätietoisuudesta luonnon oikeasta olemuksesta kumpuavat henkiset koettelemukset, mielihalu, vihamielisyys ja harhakuvitelmat, rajoittavat tiedostavien yksilöiden vapautta. Tätä vapautta ja moraalia kuitenkin on mahdollista kehittää henkisten harjoitusten avulla, keskittyneisyyden ja huomiokyvyn kehittämisellä. Myös Asaf Federman toteaa, että henkisten rajoitteiden poistaminen on buddhalaisen opetuksen ytimessä matkalla vapauteen. Tekojen seurausten tietoinen reflektointi on välttämätöntä, jotta yksilö voisi saavuttaa valaistumisen. Wallace toteaa Buddhan opettaneen: ”Mieli kulkee kaikkien ilmiöiden edellä, kaikki ilmiöt säteilevät mielestä ja muodostuvat mielestä.⁹⁴” Tahto liittyy yksilön mielen hänen käytökseensä. Moraalisia seurauksia aiheuttaviksi teoiksi määritellään tahdonalaisesti tehdyt teot.⁹⁵

Buddha kielsi opetuksessaan hallitsevan perimmäisen toimijan, sielun tai itseyden, olemassaolon, mutta ei kykyä tehdä moraalista valintaa eikä vastuuta. Federman toteaa Buddhan opettaneen, että vaikka yksilö onkin vain rykelmä toisiaan seuraavia prosesseja, tämä ei tarkoita, etteikö tekemisiään silti voisi valita. Buddhan mukaan ihmisellä olevan kyky tehdä aloite (*ārabha-dhātu*). Myös aikomusta tarkoittavat pälinkieliset termit (*sañcetanā*, *sankappa*, *cetanā*) antavat ymmärtää Buddhan olleen sitä mieltä, että yksilö on kykenevä tietoisesti suunnittelemaan tekemisiään. Buddhalaiselle tahdon vapauden käsitteelle on

⁹⁴ “All phenomena are preceded by the mind, issue forth from the mind, and consist of the mind.” Wallace 2011, 221, käännös oma.

⁹⁵ Wallace 2011, 221, 224–226, 231; Federman 2010, 9–10.

keskeisen tärkeää myös tietoisuus mielentilasta, haluista ja toiveista. Kun yksilö on tietoinen mielentilastaan, haluistaan ja toiveistaan, hän voi pohtia oikeaa toimintaa ja toimia näin, ellei tälle ole esteitä. Vapaus ymmärretään buddhalaisuudessa keskeisesti vapaudeksi mieltä haittaavista tekijöistä.⁹⁶

Wallacen mukaan moderni buddhalainen ajatus vapaasta tahdosta ei keskity kysymykseen, säätelevätkö menneet tapahtumat ja olosuhteet tahtoa vaan pohtii, missä määrin yksilöt ovat vapaita tekemään päätöksiä, jotka johtavat omaan ja toisten onnellisuuteen. Gier ja Kjellberg taas toteavat buddhalaisessa ajattelussa vapaan toimijan olevan sellainen, joka on kykenevä olemaan tietoinen toimiensa vaikutuksesta itseän ja toisiin. Moraalinen vapaus piilee toimijoiden kyvyssä muodostaa oikeita, heidän tarpeidensa ja olosuhteiden kanssa sopusoinnussa olevia haluja. Heidän mukaansa, viimeaikaisen vapaata tahtoa koskevan keskustelun valossa buddhalainen ajatusmaailma ei niinkään perustu vapaan tahdon käsitteelle, vaan vapaalle toiminnalle. Buddhan näkemyksen mukaan ihmisen syvin mielihalu oli halu määrittellä olemuksensa ”itsenä”, sieluna. Gier ja Kjellberg huomauttavat, että on ironista, että se, mikä länsimaisille filosofeille oli vapauden keskeinen lähtökohta, määrittää buddhalaisuudessa vapauden suurinta estettä.⁹⁷

Kuten jo kirjoitin, vapaus ymmärretään buddhalaisessa ajattelussa keskeisesti vapaudeksi mieltä haittaavista tekijöistä. Theravāda-buddhalainen oppi yksilöstä ehdonvaraisesti ilmenevänä osatekijöiden summana yhdistyy oppiin viisaudesta ja moraalisesta toiminnasta tapoina saavuttaa valaistuminen tietoisuuden avulla. Yksilöllä on mahdollisuus tietoisuuden avulla torjua mielihalu ja tarrautuminen.

⁹⁶ Federman 2010, 9–10.

⁹⁷ Wallace 2011, 226; Gier & Kjellberg 2004, 284–285.

2. Tietoisuus

Kysyn tutkimuskysymyksissäni miten tietoisuuden käsitteä tulkitaan niin tieteessä kuin theravāda-buddhalaisen perinteen näkökulmasta. Tässä tietoisuutta käsittelevässä luvussa tarkoitukseni on vastata näihin kysymyksiin. Aloitan pohjustaen tehtävääni seuraavalla lainauksella:

”Siitä ei ole milloinkaan kirjoitettu mitään lukemisen arvoista.” (Stuart Sutherland tietoisuudesta.)⁹⁸

Ted Honderich keskustelee kirjassaan *Actual Consciousness* tietoisuuden käsitteestä ja esittää useiden teorioiden sekä ajatuksia herättävien, mutta myös tietoisuuden ristiriitaista olemusta kuvaavien, määritelmien muassa myös ylläolevan, psykologi Stuart Sutherlandin, toteamuksen tietoisuutta koskien. Lausahdus kuvastaa tietoisuuden käsitteen yhäkin ilmeistä tuntemattomuutta. Pitää kuitenkin muistaa, että tämä on esitetty ennen kuin uuden vuosituhannen pohdinnat ovat nähneet päivänvalon. Itse Stuart on Honderichin mukaan kuvannut tietoisuuden käsitettä seuraavin sanoin: Tietoisuus on ”omata havaintoja, ajatuksia ja tunteita”. Tietoisuutta koskien on lukuisia muitakin määritelmiä, niin Honderichin mukaan maalaisjärkisiä, kuin filosofisiakin. Mikä tekee käsitteestä haasteellisen, on siitä vallitseva erimielisyys. Honderich toteaaakin, että lähes kaiken, mitä koskaan on todettu tietoisuudesta, joku on kumonnut. Hän esittää kuitenkin olevansa varma, että joitakin totuuksia voidaan saavuttaa.⁹⁹

Honderich erottelee termit olla havainnollisesti tietoinen (perceptually conscious, havaita asioita), tiedollisesti tietoinen (cognitively conscious, tietää ja ajatella asioita) ja affektiivisesti tietoinen (affectively conscious, haluta ja tuntea, aikoa asioita). Hän näkee tämän erottelun samansuuntaisena jokapäiväisen tietoisuuden kokemuksen, näkemisen, ajattelun ja haluamisen, kanssa. Hän näkee myös, että mainitut tietoisuuden muodot limittyvät. Niillä on jokin perustavanlaatuinen yhteinen piirre. Seuraavaksi Honderich pohtii, mikä sitten on tämä tietoisuuden perustavanlaatuinen piirre. Millaista jollekin on olla tietoinen? Hän näkee aloituskohdan perushavainnossa, että joku joko on tietoinen ollessaan hereillä ja vastaavasti ei ole tietoinen ollessaan esimerkiksi syvässä unessa. Tämä antaa hänestä ”maalaisjärkisen” (common sense) määritelmän tietoisuudelle, jota myös John Searle on tukenut:

⁹⁸ ”Nothing worth reading has ever been written about it.” Honderich 2014, 4, käänös oma.

⁹⁹ Honderich 2014, 4, 7.

”Tietoisuudella” tarkoitan niitä tietoisena olemisen tiloja, jotka tyypillisesti alkavat herätessämme aamulla unianäkemättömästä unesta ja jatkuvat läpi päivän, kunnes taas nukahdamme.¹⁰⁰

Honderich väittää, että meillä jokaisella on jonkinlainen ote omaan tietoisuuteemme. Hän kuitenkin kiistää, että tämä olisi tulosta minkäänasteisesta itsetutkiskelusta (introspection), mikäli itsetutkiskelulla tarkoitetaan sisäistä visiota tai aistia omasta itsestä. Honderich näkee mahdollisuuden virheisiin ja mahdolliseen itsetetokseen, jos itsetutkiskelun käsite ymmärretään mystisenä tietona itsestä ja omista mielenliikkeistä. Hän pohtii, ”voiko jokin, jota ajattelun osa (piece of thinking) kyseisenä aikana koskee, myös sisältää tämän saman ajattelun osan¹⁰¹”. Honderich ei tietenkään kiellä, etteikö ihminen voisi keskittyä tiettyyn menneeseen ajatukseen, tai tietoisuuden tilaan, kuten esimerkiksi pohtiessaan paljonko kello hetki sitten sittenkään mahtoi olla.¹⁰²

Honderich kirjoittaa tietoisuuden käsitteeseen liittyvän paljon vaikeatajuisuutta ja epämääräisyyttä. Tietoisuuteen liittyy yhä ”vaikea ongelma” (Hard Problem) siitä, miten aivot tuottavat tietoisuuden. Tietoisuutta koskee myös yhä niin kutsuttu ”selityksellinen juopa” (explanatory gap) fyysistä ilmiöiden ja tietoisuuden välillä. Toiset ajattelijat, kuten Colin McGinn, väittävät, ettei ihminen voi milloinkaan ymmärtää tietoisuutta, kun taas toiset, kuten David Papineau ja Barbara Montero, eivät näe käsitettä laisinkaan selittämättömänä.¹⁰³

B. Alan Wallace puolestaan piirtää länsimaisesta tieteestä kuvan, jossa kristinuskon luoma viitekehys ohjaa rautaisin ottein tieteen paradigmoja. Tämä juontaa hänen mukaansa kirkkoisä Augustinuksen näkemyksistä ihmisestä perisyntin vuoksi pahana ja kykenemättömänä pelastumaan, ja siitä, miten näitä tekijöitä painotettiin protestanttisessa reformaatiossa. Wallace toteaa tämän vaikuttaneen siten, että länsimaisessa yhteiskunnassa ei ollut enää tarvetta tutkiskella tietoisuutta pelastuakseen, vaan yksin usko Jumalaan saattoi pelastaa ihmisen.¹⁰⁴

Länsimainen tieteellinen vallankumous on Wallacen mukaan saanut alkunsa kristinuskon muokkaaman taustaoletuksen Jumalasta maailman luoja ja vallitseva.

¹⁰⁰ “By “consciousness” I mean those states of awareness that typically begin when we wake up in the morning from a dreamless sleep and continue throughout the day until we fall asleep again.” Honderich 2014, 4, käänös oma.

¹⁰¹ ”Can the something that a piece of thinking at a time is on or about at a time also include that same piece of thinking?” Honderich 2014, 5.

¹⁰² Honderich 2014, 1–5.

¹⁰³ Honderich 2014, 7–8.

¹⁰⁴ Wallace 2007, 53.

Fyysikot laativat hänen mukaansa siten tavoitteen ymmärtää objektiivista universumia ”Jumala- silmän” perspektiivistä. Universumille muotoiltiin lait matematiikan avulla, joka nähtiin Jumalan kieleksi. Koska tiede oli näin keksinyt objektiiviseen maailmaan ympärillään – Wallace toteaa – ja tämä maailma oli olemassa täysin tietoisuudesta riippumatta, psyyken lähtökohdat jätettiin yleisesti huomiotta Kopernikuksen ajoista aina modernin psykologian syntyyn saakka. Psykyke huomioitiin Wallacen mukaan vasta 300 vuotta tieteellisen vallankumouksen alkamisen jälkeen, kun Sigmund Freud kehitti psykoanalyttisen teoriansa 1800-luvun viimeisinä vuosina.¹⁰⁵

Lähestyn seuraavassa tietoisuuden käsitteen ymmärrystä aloittaen länsimaisen tieteen näkemyksestä esitellen Ted Honderichin ajatusten mukaiset viisi johtavaa ajatusta tietoisuudesta sekä funktionalismin käsitteen. Pohdin myös John Searlen näkemystä tietoisuuteen linkittyvästä yhä ratkaisemattomasta kysymyksestä, miten aivot tuottavat tietoisuuden. Etenen B. Alan Wallacen edelliset näkemykset haastavaan näkökulmaan tietoisuudesta. Tämän jälkeen siirryn tietoisuuden käsitteen ymmärrykseen theravāda-buddhalaisessa filosofiassa. Pohdin miksi tietoisuus koetaan theravā-buddhalaisessa ajattelussa tärkeäksi elementiksi. Lähestyn tässä luvussa keskeisiä käsitteitä abhisamṅkhāra- viññāṇa, tietoisuus ravintona, elämän ylläpitäjänä, Gandharva-henki, jatkuvuus ja attabhāva sekä Bhavaṅga-mieli.

2.1 Teoriat tietoisuudesta

Ted Honderich esittelee viisi johtavaa ajatusta tietoisuuden luonteesta. Näistä ensimmäinen on qualia-käsite (yks. quale, mon. qualia). Qualia-käsitteeseen liittyy kiinteästi ”joltain vaikuttavuus”. Qualia kuvaa sitä, miltä mikäkin kokemus vaikuttaa kenellekin, täysin subjektiivisesti kulloisenakin hetkenä. Esimerkiksi, miltä piimä maistuu, miltä sade näyttää. David Chalmers on kuvannut qualiaa yhtämittäiseksi tietoisuuden kanssa ja esimerkiksi John Searle on pohtinut käsitettä samansuuntaisesti. Honderich kuitenkin painottaa, että vaikka qualia/quale useimpien ajattelijoiden toimesta ymmärretään keskeisesti tietoisuuden kuvaavaksi ominaisuudeksi, qualia/quale ei kuitenkaan ole koko kokemisen prosessi, vaan kokemuksen ominaisuus. Tietoisuus ei yksinään koostu vain näistä laadullisista ominaisuuksista, vaan tietoisuudessa on nähtävissä myös erilaisiin esityksiin

¹⁰⁵ Wallace 2007, 12, 23, 33, Rennison 2001, 14.

kantaaottava osa (propositional attitudes), muun muassa uskoa, epäillä, kieltää asioita.¹⁰⁶

Tietoisuuden luonnetta on pohdittu myös kysymyksellä miltä jostakin tuntuu olla tämä jokin. Filosofia on pohtinut tietoisuuden selitystä tältä näkökannalta kysymyksen ”Miltä tuntuu olla lepakko?” avulla. Vastauksena tarjotaan, että olenolle olla tietoinen on yhtä kuin, että on olemassa jotain, miltä tuntuu olla tämä olento. Tietoisesta olenolle on keskeistä, että sillä on *ominaisuuksia*, joiden avulla voidaan johtaa lopulta toteamus:

Olenolle olla tietoinen on, että on olemassa jotain, millaiselta tuntuu olla tämä olento, *tämän olennon tietoisuudessa*.¹⁰⁷

Honderich esittää tekstissään useita vastaavia väittämiä, joiden oleellinen yhteinen tekijä on teroittaa tietyn yksittäisen tietoisuuden tuntua tämän tietyn yksittäisen tietoisuuden tiedostajalle. Hän kuitenkin toteaa kehäpäätelmän ongelman vaivaavan näitä väittämiä. Loppujen lopuksi, tietoisuus pystytään hänen mukaansa näissä selittämään vain tietoisena olemisella. Keskeinen väite lauseissa on ”miltä jollekin tuntuu”. Tästä pääsemme Honderichin kuvaamaan seuraavaan ajatukseen tietoisuudesta; subjektiivisuuteen.¹⁰⁸

Honderich toteaa myös subjektiivisuutta ja tietoisuutta käsiteltäneen synonyyminomaisesti, hän myös toteaa jo edellä käsiteltyjen qualia-käsitteen ja ”millaista on olla jokin” -kysymyksen nivoutuvan läheisesti subjektiivisuuden olemukseen. Tietoisuus nähdään ensisijaisesti jonkun, yksittäisen henkilön ominaisuutena. Honderich kitenkin kysyy, voidaanko tietoisuutta lainkaan määrittää ominaisuutena, samoin kuin vaikka aistittavia, fyysisiä ominaisuuksia. Tietoisuus henkilön mahdollisena ominaisuutena on myös taipuvainen katoamaan. Voimme kaikki allekirjoittaa oletuksen, ettei tajuton henkilö ei ole tietoinen. Tietoisuutta ei niin ikään voi lukea koko (fyysisen) henkilön ominaisuudeksi. Vaikka henkilö olisi tietoinen, Honderich esittää epäilyksen, että kyseisen henkilön varpaat olisivat osallisena tästä tietoisuudesta. Niinpä Honderich päättelee, että tietoisuus lienee subjektiivista erityisellä metafysisessä itseen nivoutuvalla tavalla. Hän esittelee seuraavat tietoisuudesta seuraavat pääkohdat subjektiivisuuteen liittyen:

¹⁰⁶ Honderich 2014, 17–19, 21.

¹⁰⁷ ”For a thing to be conscious is for there to be something that is what it is like to be that thing *in the consciousness of that thing*.” Honderich 2014, 25, käännös oma. Kursiivi kuten alkuperäisessä tekstissä.

¹⁰⁸ Honderich 2014, 23–25.

(1) Tietoisena oleminen on subjektiivista siinä erityisessä mielessä, että se on jatkuvan ykseyden ominaisuus, tämä jatkuva ykseys ollen erityisessä mielessä itse, sisäinen subjekti, ego, ihmisen alku, sielu-substanssi, erillinen entiteetti tai mahdollinen ihmispersoonaa tai vastaavaa koskeva täydentävä tosiasia.¹⁰⁹

(2) Tietoiset tilat ovat subjektiivisia siinä mielessä, että ne ovat riippuvaisia olemassaolostaan kokemuksessa.¹¹⁰

(3) Jokaisen persoonan, tai vastaavan, tietoisuus on subjektiivista siten, että sen varmoissa otaksumissa, joita se otaksuu tai otaksuisi, henkilökohtaisen tiedon tai uskon uniikissa laajuudessa otaksumaan liittyen – on olemassa etuoikeutettu pääsy johonkin.¹¹¹

(4) Tietoisena oleminen on subjektiivista siinä mielessä, että jokin sisältää näkökannan tai perspektiivin johonkin.¹¹²

Honderich mainitsee ensimmäiseen kohtaan liittyvän jälleen kehäpäätelmän vaaran. Tietoisuus näyttäytyisi tässäkin tietoisesta persoonan ominaisuutena. Subjektiivisuuden ajatukseen liittyy myös muita, ajattelijoiden toimesta niin vastustettuja kuin puolustettujakin, näkökulmia. Honderich on kuitenkin sillä kannalla, ettei subjektiivisuuden käsite monitulkintaisuudestaan huolimatta olisi täysin vailla perustaa.¹¹³

Neljäs tapa pohtia tietoisuutta perustuu Franz Brentanon ajatuksiin. Hänen ajattelunsa pohjaa keskiajalla tehtyyn huomioon, että henkisillä tapahtumilla on kohde. Jotain joko rakastetaan, vihataan, hyväksytään ja niin edelleen. Ajatuksia kuvattiin keskiajan latinassa sanalla ”intentio” ja tästä johtaa Brentanon näkemys tietoisuudesta intentionallisuutena (intentionality). Tästä tietoisuuden määrittelystä puhutaan myös termein ”koskevuus” (aboutness) tai ”esitys” (representation). Honderich selvittää tämän ajatuksen mukaan tietoisuutta voivan näin kuvata jonakin, jolla on kohde tai sisältö (ajatuksia, käsitteitä, esityksiä ynnä muuta). Nämä kohteet eivät esiinny fyysisesti, mutta tietoisuus sisältää suuntautumisen näihin kohteisiin tai sisältöihin, ja ne sisältyvät tietoisuuteen. Näistä kohteista on tapana arkikielessä puhua esimerkiksi ajateltuina tai rakastettuina, haluttuina tai vihattuina. Honderich huomauttaa ja painottaa, että arkikielinen näkemys kohteista eroaa,

¹⁰⁹ (1) “Being conscious is subjective in the particular sense of being property of a continuing unity in the particular sense of a self, inner subject, ego, homunculus, soul substance, separate entity, or further fact in or about a human person or the like.” Honderich 2014, 29, käännös oma. *Homunculus = (tarinoissa esiintyvä) hyvin pieni ihmishahmo. Oxford Advanced Learner's Dictionary.

¹¹⁰ (2) “Conscious states are subjective in that they depend for their existence in experience.” Honderich 2014, 30, käännös oma.

¹¹¹ (3) “The consciousness of each person or the like is subjective in that on certain assumptions it does or would issue in a unique extent of personal knowledge or belief with respect to it – there is privileged access for something.” Honderich 2014, 31, käännös oma.

¹¹² (4) “Being conscious is subjective in the sense of something’s involving a point of view or perspective of something.” Honderich 2014, 32, käännös oma.

¹¹³ Honderich 2014, 28–32, 34.

tavanomaisesti ymmärrettynä, Brentanon muista näkemyksistä. Halutut (tai rakastetut tai vihatut) asiat, joita tässä tavanomaisessa ymmärryksessä tarkoitetaan, ovat tietoisuuden ulkopuolisia kohteita, ihmisiä, asioita, fyysisiä kohteita (esimerkiksi lasi piimää).¹¹⁴

Viidentenä ajatuksena tietoisuudesta Honderich esittelee neuro- ja kognitiotieteen sekä psykologian näkemyksen. Tämän mukaan tietoisuus on jakautunut kahteen, fenomenaaliseen (phenomenal) ja pääsy- tai psykologiseen tietoisuuteen (access /psychological consciousness). Honderich keskittyy keskustelemaan fenomenalisesta tietoisuuden ajatuksesta, kutsuen tätä fenomenalisuudeksi (phenomenality). Fenomenaalisen tietoisuuden ajatuksen keskeinen tekijä on aistihavainto (sensation), mikä linkittyy yhteen myös qualia-käsitteen kanssa. Fenomenaalinen tietoisuuden aate on keskeisesti kokemusta. Tätä voidaan myös kuvata kysymyksellä ”Millaista on kokea kokemus?”¹¹⁵

Mikä tekee tästä tietoisuuden aatteesta kiintoisan, on sen koettu suhde toiseen puoliskoonsa, pääsy tietoisuuteen (access/psychological consciousness). Honderich toteaa tätä pääsy tietoisuutta, hänen mukaansa huolettomasti, voitavan kutsua aivoiksi. Fenomenaalinen tietoisuus siis linkittyy hermosoluihin. Tässä vaiheessa selityksellisen juovan ongelma nostaa päätään. Mikä linkittää aivot, tai vastaavan perustavan materian, tietoisuuteen? Fenomenalisuudesta tietoisuuden teoriana ovat keskustelleet etenkin Ned Block ja David Chalmers ja he ovat työssään yhdistäneet tähän fenomenalisuuden teoriaan myös neljä jo edellä mainittua ajatusta tietoisuudesta. Osana fenomenalista tietoisuutta olisivat siis (1) qualia-käsite, (2) miltä tuntuu olla jokin, (3) subjektiivisuus ja (4) intentionallisuus. Honderich näkee näissä ajatuksissa tietoisuuteen olennaisena kuuluvana sen, että tietoisuutta ei ole johdettu mistään muusta, kuin toimijasta itsestään.¹¹⁶

Honderich näkee, että näistä keskustellusta viidestä tietoisuuden teoriasta yksikään kuitenkaan tarjoa tyydyttävää esitystä tietoisuuden luonteesta kehämäisyytensä ja epätäydellisyytensä perusteella. Nämä kuitenkin tarjoavat hänen mukaansa alustan kehittää tuoreita ajatuksia aiheesta. Kirjassaan Honderich on koostanut listan aatteita tietoisena olemisen sisällöstä, joista esitän seuraavassa muutamia. Tietoisuus on muun muassa jonkin omaksumista, kokemista jonkin kokemisen mielessä, tietoisuutta jostain (awareness), jonkin olemisen ilmeistä,

¹¹⁴ Honderich 2014, 35.

¹¹⁵ Honderich 2014, 43–45.

¹¹⁶ Honderich 2014, 45–46, 49.

jonkin olemassaoloa, jonkin oloa tiedettynä, jotain todellista ja välitöntä, ja niin edelleen. Honderichin lista on mittava ja esittää ikään kuin käsittelemätöntä tietoaaineistoa tavallisen (ordinary) tietoisuuden sisällyttämistä tekijöistä. Nämä Honderich niputtaa käsitteen ”todellinen tietoisuus” (actual consciousness) alle. Hän määrittää tämän ”todellisen” jonkin todellisena oloksi (consciousness as something’s being actual).¹¹⁷

Kysymys vaatii yhä vastaustaan. Mitä millekään on olla tietoinen tavanomaisessa mielessä? Honderich näkee tässä kaksi kysymystä. Mikä tietoisuudessa on todellista ja etenkin, mikä on todellista havainnollisessa, tietoisessa ja affektiivisessä tietoisuudessa kussakin? Edelleen, hän on sitä mieltä, ettei näihin kahteen kysymykseen voi vastata tyydyttävästi, ellei pystytä vastaamaan tyydyttävästi myös kolmanteen: miten tietoisuus ja aivot liittyvät toisiinsa.¹¹⁸

Honderich toteaa 2000-luvulla vallinneen laajan yksimielisyyden hallitsevasta tietoisuuden teoriasta. Tämä on funktionalistinen teoria tietoisuudesta. Funktionalismista tavataan puhua vastauksena siihen, mistä mielen sisäiset, eivät vain tietoiset, tosiasiat rakentuvat; miten mieli toimii. Honderich lähtee esittelemään funktionalismia kaksivaiheisen perustelun avulla. Ensimmäinen vaihe perustuu monikertaiseen ymmärrykseen (multiple realizability). Tämä liittyy henkisiin tiloihin, kuten kipu, verrattuna tiettyihin henkisiin tiloihin, kuten kipu juuri tietyllä hetkellä sormen jäädessä oven väliin. Moninkertainen ymmärrys koskee henkisten olotilojen suhdetta. Sama tietoisuuden tila (esimerkiksi kipu) syntyy useista erilaisista hermostollisista tiloista. Perustelun toinen vaihe kertoo, että tietoisien tilan tyyppi tulee kuvata sen kausaalisella yhteydellä muihin tapahtumiin, toisin sanoen tarkoituksensa avulla (function). Honderich kuvailee tätä esimerkillä:

- (1) Ulkona on metsikkö auringonvalossa, aamuauringon ensisäteissä.
- (2) Henkilön tietoisuus tästä näystä (metsiköstä) ikkunan kautta.
- (3) Henkilö kokee mieltänsä ylenevän ja kokee pilkahduksen mielihyvää. Mielihyvä on tietoinen tapahtuma c, funktionalismin mukaan merkki tietoisesta tapahtumasta C (mielihyvä yleensä).
- (4) Mielihyvän c kokemiseen liittyy hermostollinen tapahtuma n, joka on merkki hermostollisesta tapahtumasta N.
- (5) Henkilö muodostaa tietoisuudessaan (kokemuksesta c johtuen) aikomuksen mennä ulos hengittämään aamuilmaa.
- (6) Henkilö menee ulos puutarhaan (tapahtuma käytöksessä).¹¹⁹

¹¹⁷ Honderich 2014, 50, 67–69.

¹¹⁸ Honderich 2014, 81–83.

¹¹⁹ Honderich 2014, 97.

Funktionalismi on, Honderichia lainaten, vastaus kysymykseen c:n (merkki mielihyvystä) luonteesta, ”Mikä c on?”, ja vastaus annetaan syy-seuraus -suhteiden avulla. Esimerkin mukaan c on yksi esimerkki tietoisuuden tiloista C, joka on syy-seuraus -suhteessa muihin tapahtumiin; ulkoinen elementti (metsä), tietoisuus metsästä (näkeminen) ja niin edelleen.¹²⁰

Voivatko aivot saada aikaan tietoisuuden? Miten tämä tapahtuu? John Searle on kommentoinut, miten aivojen neurobiologiset prosessit saavat aikaan tietoisuuden, taannoisessa artikkelissaan ”Introduction: Addressing the Hard Problem”. Hän ei näe, että tietoisuus olisi käsitteenä vaikea määritellä. Tietoisuutta ovat kaikki tuntevat ja aistivan tietoisuuden (sentient awareness) tilat. Hän niin ikään näkee, että tiedämme jo maailmasta tarpeeksi voidaksemme todeta tietoisuuden olevan alempien hermosolujen toiminnan tuottamaa. Tietoisuus on neurosolujen ampumien sähköimpulssien tuottamaa ja se toteutuu hermosoluissa. Näin Searle ratkaisee mieli-keho ongelman (mind-body problem) biologiseen naturalismiin pohjautuvalla näkemyksellään. Tämän aatteen kannattajat olettavat, että mieli on luonnollinen biologinen ilmiö.¹²¹

Searle näkee edeltäneen keskustelun tietoisuuden luonteesta olevan suoranaisesti hämmentävä hänen näkemyksensä mukaan johtuen sekä dualismin että tieteellisen materialismin nimiin vannovista aatteista. Dualismi on hänestä väärässä olettaessaan tietoisuuden olevan jollain tavalla erillinen entiteetti kehosta ja tieteellinen materialismi olettaessaan tietoisuuden joko (1) olevan olematon (non-existent) tai (2) olevan johdettavissa johonkin havaittavaan, kuten käytös, hermosolujen impulssien ampuminen tai tietokoneohjelma. Searle kieltää tietoisuutta käsittelevinä aatteina substanssidualismin (substance dualism), ominaisuusdualismin (property dualism), materiaalisen reduktionismin (material reductionism), behaviorismin (behaviorism) ja komputationalismin (computationalism).¹²²

¹²⁰ Honderich 2014, 85, 93–98.

¹²¹ Searle 2014, vii–viii.

¹²² Substanssidualismin mukaan ihminen on yhtä kuin sielunsa ja aineettomina sieluina hetkellisesti yhdessä kehonsa kanssa. Ominaisuusdualismi taas näkee ihmisessä olevan kaksi metafysisesti erilaista ja erillistä ominaisuutta, sekä fyysinen että henkinen. Materiaalisen reduktionismin aate on, ettei tietoista ilmiötä ole olemassa. Pelkistymätöntä ontologista subjektiivisuutta ei yksinkertaisesti ole. Behaviorismissa tietoisuus pelkistyy käytökseen. Komputationalismissa tietoisuus rinnastetaan ohjelmaan; tietoisuus on ohjelma siinä missä aivot nähdään tätä ohjelmaa pyörittävänä laitteistona. Searle 2014, ix–x.

Voitaneen oikeutetusti kummastella, miksi tietoisuutta ei siis vielä ole tyhjentävästi perusteltu tieteellisesti biologiaan perustuen. Searle vastaa artikkelissaan kysymykseen. Hänen mukaansa ongelma on nykyisessä paljolti kuvantamiseen tai yhden solun toiminnan tallentamiseen perustuvassa tutkimustekniikassa. Tietoisuuden syntyä ei ole pystytty toteamaan näillä käytettävissä olevilla metodeilla. Searle pohtii tämän johtuvan siitä, että tietoisuus ei ole hänen mukaansa riippuvainen yksittäisistä hermosoluista eikä edes hermoimpulssikuvioista. Lisäksi, tietoisuuden ongelmaa on tähän asti tutkittu jo olemassaolevan tietoisuuden kentässä (conscious field), tutkimuskohteet ovat Searlen mukaan lähes kaikissa tapauksissa olleet jo tietoisia. Tämä ei hänen mukaansa palvele tutkimusta, kun tarkoituksena on selvittää, miten tutkittavasta tulee tietoinen ja kuinka aivot alun alkaen synnyttävät tietoisuuden kentän.¹²³

Tietoisuuteen siis liittyvät käsitteet joltain vaikuttavuus (qualia). Tähän nivoutuu myös kysymys, miltä jollekin tuntuu olla tämä jokin. Edelleen, tietoisuuteen liittyy vahvasti subjektiivisuuden oletus. Tietoisuudella on oletettu olevan kohde, jota kohti se suuntautuu. Lisäksi tietoisuuteen on liitetty fenomenalisuus, kokemus, ja tämä fenomenalisuuden on myös oletettu olevan yhteydessä fyysisyyteen, aivoihin. Tietoisuutta voidaan kuvata lukuisin eri tavoin, mutta vastausta kysymykseen, mikä tietoisuudessa on todellista, etsitään yhä. Hallitsevin teoria Honderichin mukaan on funktionalistinen teoria tietoisuudesta, siitä miten mieli ja tietoisuus voisi rakentua syy-seuraus –suhteista. Tämä teoria tukisi uskottavasti teoriaa, että aivot tuottavat tietoisuuden. John Searle on miltei jo selittänyt, miten tämä tapahtuu. Haaste on tietoisuuden mittareiden puuttuminen.

2.2 Näkökulma tieteelliseen materialismiin

B. Alan Wallace on kyseenalaistanut tieteen väitteitä tietoisuudesta aivojen tuottamana. Etenkin liittyen buddhalaiseen näkemykseen tietoisuuden urkenemisestä näen varsin mielekkääksi esitellä Wallacen tieteen paradigmoja haastavaa näkemystä tietoisuuden olemuksesta. Wallace lainaa Daniel Dennettia ”[t]ietoisuus yksin on nykyään aihe, joka usein jättää valistuneimmatkin ajattelijat hiljaisiksi ja hämmentyneiksi¹²⁴” ja John Searlea ”[m]illoin mieli on kyseessä,

¹²³ Searle 2014, x–xi.

¹²⁴ “Consciousness stands alone today as a topic that often leaves even the most sophisticated thinkers tongue-tied and confused.” (D. Dennett, *Consciousness Explained*, 1991.) Wallace 2007, 56, 175.

olemme tyypillisesti hämmentyneitä ja erimielisiä¹²⁵” todetessaan, että länsimaisen tieteen piirissä ei viimeaikaisen tiedon valossa osata sanoa tietoisuudesta juuri yhtään mitään ja mielen olemuksesta vallitsevat ristiriitaiset näkemykset. Wallace toteaa tieteentekijöiden oletettavan tietävänsä, ettei tietoisuus ole olemassa riippumatta aivoista, ja siten ainoa kysymys, johon täytyy löytää vastaus on miten aivot tuottavat tietoisuuden. Wallace viittaa myös Edward O. Wilsonin ajatukseen, että mieli koostuu aivojen toiminnasta, mutta Wallace painottaa, ettei tätä viimeaikaisen tutkimuksen valossa ole kuitenkaan vielä todistettu.¹²⁶

Wallace näkee luonnontieteen perustuvan viimeaikaisessa muodossaan paljolti 1800-luvun materialismille ja tiivistää tieteellisen materialismin maailmankuvan seuraavasti: ei-persoonallisista luonnollisista voimista juontuva fyysinen maailma on ainoa realiteetti. Fyysisessä maailmassa ei ole mitään moraalista järjestystä. Se toimii ilman minkäänlaisten hengellisten voimien väliintuloa. Elämä ja tietoisuus syntyivät täysin sattumanvaraisesti. Elämällä yleisesti, erityisesti ihmiselämällä, ei ole mitään muuta tarkoitusta, kuin mitä se lukee omaksi ansiokseen. Yksilön elämän aikana kaikki tunnetilat ja vastaavat tunteet ovat kehon aiheuttamia. Kuolemassa yksilö tuhoutuu.¹²⁷

Wallace toteaa tieteelle metodiselta kannalta olevan järkevää käyttää naturalisia oletuksia ja pyrkiä ymmärtämään maailmaa luonnollisessa, ei yliluonnollisessa, muodossa. Tieteelliset teoriat on hänen mukaansa voitava alistaa empiiriselle kokeellisuudelle ja havainnoille. Syy sille, miksi Wallace näkee tieteellisen materialismin epämielekkäänä, on se, että sen avulla ei tyydytä edistämään tietoa, vaan ajatusmalli on hänestä kehitetty maailmankatsomuksen asteelle. Wallacen mukaan tieteellisen materialismin maailmankuvalla ei välttämättä ole annettavaa tietoisuuden tieteelliselle tutkimukselle, mutta useat tieteentekijät kuitenkin näkevät tämän ainoana oikeana katsomuksena. Wallace huomauttaa, ettei tällainen vakaa usko eroa kovinkaan paljon uskonnollisesta uskosta. Tieteeseen kuuluvat myös uskomukset, hypoteesit, joita testataan, ja siten todistetaan uuden tiedon valossa oikeiksi tai vääriksi.¹²⁸

¹²⁵ “Where the mind is concerned we are characteristically confused and in disagreement.” (J. Searle, *The Rediscovery of the Mind*, 1992.) Wallace 2007, 56, 175.

¹²⁶ Wallace 2007, 56–57.

¹²⁷ Wallace ehdottaa Karl Marxin materialistisella ideologiallaan vaikuttaneen materialistisen näkemyksen syntyyn. Näkemys ujuttautui osaksi tieteen maailmaa Aleksander Oparinin 1938 englanninkielisenä käännöksenä julkaistussa *Origins of Life* -teoksessa. (Julkaistu venäjäksi vuonna 1936.) Wallace 2007, 29.

¹²⁸ Wallace 2007, 29–30.

Wallace toteaa tieteellisen tutkimusmetodin olleen erinomainen instrumentti maailmankaikkeuden tutkimuksessa. Tämä menestys on kuitenkin johtanut hänen mukaansa siihen, että ainoastaan tieteen tunnustamat tutkimusmenetelmät mielletään päteviksi. Koska tieteen menetelmät ovat osoittautuneet menestyksekkäiksi fyysisten ulottuvuuksien tutkimuksessa, on Wallacen mukaan tehty olettautunut, että nämä pätevät myös metafyyssisten ulottuvuuksien tutkimuksessa. Näillä alueilla ei kuitenkaan voi hänen mukaansa saada kokeellisia tuloksia. Tieteen menetelmien osoittautuneen menestyksekkäiksi fyysisten ulottuvuuksien tutkimuksessa on Wallacen mielestä johtanut tieteelliseen materialismiin, joka on ylittänyt tieteellisen teorian rajat ja saanut dogmin aseman.¹²⁹

Wallacen mukaan esimerkiksi moderni kognitiotiede perustaa löydöksensä käytökseen, aivojen toimintaan ja subjektiivisiin raportteihin normaalien ja epänormaalien ihmisten käytöksestä, mikä eittämättä johtaa käsitykseen psyykestä aivojen toimintana. Wallace kiittää, että tulokset tuskin tästä muuttuvat niin kauan, kuin tiede mittaa tuloksia edellä mainituilla tavoilla. Hän näkee, että materialistisiin oletuksiin perustuvat tieteelliset menetelmät tulevat todennäköisesti vahvistamaan näitä materialistisia oletuksia. Wallace toteaa Daniel J. Boorstin kutsuvan näkemystä, jonka mukaan psyyke todellakin olisi vain aivojen toimintaa, kirjassaan *The Discoverers: A History of Man's Search to Know His World and Himself* "tiedon illuusioksi". Tämä tiedon illuusio – Boorst väittää – eikä tietämättömyys, on ollut kautta aikain tieteellisten löytöjen suurimpana esteenä. Wallace onkin sitä mieltä, että tiedeyhteisö on konservatiivinen ja välttää hyväksymästä oletettuja tuloksia, jotka rapauttaisivat arvoja, joille tieteellinen maailmankuva on rakentunut.¹³⁰

Mainittu konservatiivisuus ei toki ole sääntö. Wallace kertoo aikanaan ensimmäisen neurotieteen laboratorion Harvardin yliopistoon perustaneen William Jamesin pohtineen kolmea teoriaa kuvaamaan mielen liikkeitä suhteessa aivot toimintaan. Hänen näkemyksensä mukaan aivot joko

(1) tuottavat ajatuksia samalla tavoin, kuin virtapiiri tuottaa valoa, (2) vapauttavat tai sallivat henkiset tapahtumat samalla tavoin, kuin jalkajousen liipaisin sallii jalkajousen jänteen vapautua poistamalla esteen sen tieltä tai (3) välittävät ajatuksia siten, kun prisma välittää valon osuessa siihen kaikki spektrin värit.¹³¹

¹²⁹ Wallace 2007, 28.

¹³⁰ Wallace 2007, 11, 22, 24.

¹³¹ "1) the brain produces thoughts, as an electric circuit produces light; 2) the brain releases, or permits, mental events, as the trigger of a crossbow releases an arrow by removing the obstacle that holds the string; and 3) the brain transmits thoughts, as light hits a prism, thereby transmitting a spectrum of colors." Wallace 2007, 12–13, käännös oma.

James itse oli kolmannen vaihtoehdon kannalla. Hänen teoriansa oli, että tietoisuus on eri ilmiö kuin aivot. Kuolemassa aivot menehtyvät, mutta tietoisuuden virta, jota ne välittivät, ei katoa. James toteaa: ¹³²

Ja viimein, kun aivot kokonaan lopettavat toimintansa, tai kuihtuvat, tämä erityinen tietoisuuden virta, jota ne palvelivat, katoaa kokonaan tästä luonnollisesta maailmasta. Mutta olemisen piiri, joka välitti tietoisuuden, olisi yhä koskematon. Siinä todellisemmassa maailmassa jossa, vaikka yhä ollessaan täällä, se oli jatkuva, tietoisuus voisi, meille tuntemattomilla tavoilla, jatkua yhä. ¹³³

Wallace toteaa, ettei luonnontiede ole toistaiseksi selittänyt tietoisuuden tai minkään muun henkisen ilmiön luonnetta tai alkuperää. Hän myöntää, että merkittäviä edistysaskeleita on otettu selvittäessä miten hermostolliset tapahtumat peilaavat henkisiä tapahtumia. Silti vielääkään ei osata selittää, miten subjektiivinen kokemus saa alkunsa aivojen toiminnasta. Viimeaikaisen tutkimuksen valossa tietoisuuteen liittyen, ei ole varmaa tietoa siitä, mitä tietoisuutta mitatessa tulisi mitata, eikä olemassa olevaa teknologiaa, jolla mitata tietoisuuden läsnäoloa tai poissaoloa. Wallace toteaa olevan yhä olemassa syitä epäillä vain aivojen olevan mielen takana. Tiede on menestyksekkäästi kehittänyt eri tutkimustapoja eri ilmiöiden tutkimiseen ja Wallacen näkemyksen mukaan myös mielen tutkimuksen metodeja tulisi kehittää. Hänestä mielen tapahtumien hermostollisten ja käytöksellisten korrelaattien tutkiminen ei ole riittävää. Niin kauan kuin tiede keskittyy näiden tutkimiseen, tietoisuus eittämättä tulee löytymään vain näistä tutkittavista prosesseista. ¹³⁴

Wallace kirjoittaa buddhalaisen Suuri täydellisyys -koulukunnan ajatuksista tietoisuuteen liittyen. Vaikka pyrinkin tutkimuksessani pidättäytymään theravāda-buddhalaisuudessa, näen mielekkääksi käydä läpi lyhyesti myös tämän tiibetiläisen koulukunnan näkemyksiä. Wallacen mukaan koulukunta näkee mielen toiminnan seuraamisen mahdolliseksi mietiskelyn, introspektion, avulla. Saavutettaessa meditatiivinen uinuvuus (śamatha) saavutetaan samalla tietoisuuden perustaso, joka tunnetaan myös kasvualustatietoisuutena (substrate consciousness, ālayavijñāna). Suuri täydellisyys -koulukunta tulkitsee tämän erilliseksi tietoisuuden virraksi, josta myös psyyke ja fyysiset aistit juontavat juurensa. On huomion arvoista, että

¹³² Wallace 2007, 12–13.

¹³³ “And when finally a brain stops acting altogether, or decays, that special stream of consciousness which it subserved will vanish entirely from this natural world. But the sphere of being that supplied the consciousness would still be intact; and in that more real world with which, even whilst here, it was continuous, the consciousness might, in ways unknown to us, continue still.” James 2008, 87; Wallace 2007, 13, käännös oma.

¹³⁴ Wallace 2007, 14, 39.

tämän näkemyksen mukaan keho ja vuorovaikutus fyysisen ympäristön kanssa vaikuttavat psyykeen, mutta tietoisuus juontuu kasvualustatietoisuudesta. Wallacen toteaa tämän väitteen käyvän yksin niin läntisen kulttuurin tunnustamien ajattelijoiden Pythagoraan¹³⁵, Sokrateen¹³⁶, Origeneen¹³⁷, Augustinuksen¹³⁸ ja jo yllämainitun William Jamesin teorioiden tietoisuuden luonteesta kanssa. Wallacen toteaa saman väitteen sopivan myös olemassaolevaan tietoon aivojen ja mielen vuorovaikutuksesta. Suuri täydellisyys -koulukunnan mukaan aivot ja ympäristö siis vaikuttavat henkisiin tapahtumiin mutta nämä tapahtumat lähtevät kasvualustatietoisuudesta ja myös palautuvat tähän. Wallace toteaa, ettei ”Hard Problem” -ongelmaa ole olemassa buddhalaisesta näkökulmasta. Ongelmaa siitä, miten aivot tuottavat subjektiivisia henkisiä kokemuksia, on tarpeetonta selittää. Nämä henkiset kokemukset urkenevat kasvualustatietoisuudesta.¹³⁹

Kasvualustatietoisuus saattaa vaikuttaa buddhalaiselta vastineelta muun muassa kristinuskon kuolemattomalle sielulle. Käsitteissä on silti muutamia keskeisiä eroja. Wallace kuvaa, miten mietiskelijät, jotka ovat saavuttaneet uinuvan meditatiivisen tilan, usein kuvaavat tämän tietoisuuden ulottuvuuden virtana nousevia ja poistuvia tietoisuuden hetkiä. Kasvualustatietoisuus ei siten ole jatkuva, muuttumaton entiteetti. Kasvualustatietoisuus vaikuttaa psyykeen ja kasvualustatietoisuuteen vaikuttavat vastaavasti fyysiset ja henkiset tapahtumat. Näin ollen kasvualustatietoisuus ei ole itsenäinen. Wallace lisää tiibetiläisen Suuri

¹³⁵ Pythagoraan (n. 570 – n. 495 eKr.) tiedetään perustaneen mietiskelyyn keskittyneen yhtä aikaa uskonnollisen ja tieteellisen yhteisön, jonka päämäärä oli vaalia pyhyttä kehon ja mielen puhdistamisen kautta, tieteen ollessa puhdistautumisen päämenetelmä. Puhdistautumiseen omistautuva oli asioita tarkasteleva (theorein, θεωρεῖν) ”tosi filosofi”. Pythagoras tunnettiin myös metempsychosis (μετεμψύχωσις) -teoriastaan, jonka mukaan kuolematon sielu syntyy uudelleen eri muodoissa. Wallace 2007, 12.

¹³⁶ Sokrateen taas tiedetään todenneen Platonin näytelmässä Phaedo, että vastoin yleistä uskomusta sielun menehtymisestä kuolemassa, sielu jatkaa kuoleman jälkeen. Lihallisten mielitekojen kieltämisen kautta ”kuolemaa harjoitellut” filosofin sielu jatkaa paikkaan, joka, kuten tämä sielunkin on kuolematon, pyhä ja viisas, ja jossa onnellisuus ja vapautus kaikesta inhimillisestä pahasta odottavat. Filosofiaa harjoittamattomien sielut taas jatkavat uudelleensyntymää edeltävään välitilaan. Sokrates toteaa, heidän lihallisten mielitekojensa vuoksi tulevan uudelleenvangituiksi kehoon, joka vastaa heidän edellisiä tekojaan. Wallace 2007, 12.; Platon, Phedo, 24.

¹³⁷ Origenes (185 – 254) ajatteli sielun voivan nousta kohti Jumalaa monen eri elämän ajan aikana. Mietiskelyn (theoria, θεωρία) avulla sielu kehittää tietoutta (gnosis) Jumalasta ja kehittyä, Platonin opetuksen mukaan, kunnes tulee pyhäksi (divine).

¹³⁸ Augustinus puolestaan esitti neljä hypoteesia sielun synnystä (1) yksilön sielu juontaa hänen vanhempiansa sieluista, (2) yksilön sielu luodaan ”tyhjältä” hedelmöityksen hetkellä, (3) sielut ovat jossain muualla, ja Jumala lähettää sielun asumaan ihmiskehoon, (4) sielut laskeutuvat ihmistiasolle omasta halustaan. Augustinuksen mielestä nämä kaikki sopivat kristinuskoon oikein hyvin, todellisessa filosofian hengessä, mitään näistä ei voi vahvistaa todemmaksi ilman hyvää syytä. Wallace 2007, 12.

¹³⁹ Wallace 2007, 15–17.

täydellisyys–koulukunnan tuntevan myös käsitteen alkuperäinen tietoisuus (primordial consciousness, jñāna) tietoisuuden absoluuttisena perustana. Tämän alkuperäisen tietoisuuden on sanottu olevan onnellisuuden, totuuden ja hyveen lähde. Kyseisen tietoisuuden ulottuvuudesta on lähtöisin myös perimmäinen kaipaus onnellisuuteen, totuuteen ja hyveeseen.¹⁴⁰

Tietoisuuden luonteesta vallitsee yhä vaihtelevia näkemyksiä. Tiede pystyy tarjoamaan teorioita, mutta ei selittämään tietoisuuden ilmiötä tai todistamaan sen perustumista fyysisiin ilmiöihin. Paradigmaa haastaa muun muassa Wallacen esittämä näkemys tietoisuudesta aivojen välittämänä ilmiönä, mikä sointuu yhteen myös buddhalaisen näkemyksen kanssa tietoisuuden luonteesta.

2.3 Tietoisuus theravāda-buddhalaisessa filosofiassa

Oman tutkimukseni kannalta on tietenkin keskeisen kiinnostavaa, miten tietoisuus ymmärretään theravāda-buddhalaisen koulukunnan piirissä. Tätä pohdittaessa on sekä mielestäni kiinnostavaa että – Collinsia lainaten – tärkeää muistaa ottaa huomioon buddhalaisen käsityksen tietoisuudesta periytyvän sanskritinkielisestä kulttuurista ja bramiiniuskosta. Bramiiniuskosta juontaa ajatus, että tuleva olomuoto on seurausta ja tulos toiminnasta. Bramiiniuskosta juontaa myös ajatus tietoisuudesta pelikenttänä valaistumisen saavuttamiseksi. Henkisen aikomuksen olemassaolosta kehittyi buddhalaisuudessa karmaan (kamma) ratkaisevasti vaikuttava tekijä, sillä onnellisuus ja kärsiminen eivät saa alkuaan toiminnasta, vaan niiden takana olevasta aikomuksesta.¹⁴¹

Bramiiniuskossa oli keskeisen tärkeää uhrin toimittaminen. Tämän uskottiin luovan jatkuvuutta elämälle ja mahdollisuuden (parempaan) uudelleensyntymään. Joskaan parempi uusi elämä ei ollut ehdoton ja välitön seuraus uhrin toimittamisesta. Buddhalaisessa ajatusmaailmassa tämä ajatus muuntautui. Itse aktuaalisen teon suorittamisen tärkeys himmeni teon aikomuksen rinnalla. Näin ollen toiminta muuttui ulkoisesta suorittamisesta sisäiseksi itseymmärrykseksi. Teon (aikomuksen) luonne muuttui, ei vain mahdollisesti seuraavaa olomuotoa määritteleväksi, vaan ehdottomasti seuraavan olomuodon määritteleväksi. Viisauden saavuttaminen tietoisesti suoritettun itsen havainnoinnin avulla tuli valaistumisen keskiöön. Valaistuksen rinnalla nähtiin maallikkouskovaisen elämän

¹⁴⁰ Wallace 2007, 17–21.

¹⁴¹ Collins 1982, 200–202.

pidennys saṃsāraassa, jolloin elämän pyörä jatkaa pyörimistään tekojen (aikomusten) seurauksena.¹⁴²

Theravāda-buddhalaiseen tietoisuuden käsitteeseen linkittyivät samalla Collinsin mukaan kaksi keskeistä käsitettä. Nämä olivat muodostelmat (formations, saṃskāra/pāli: saṃkhāra) karman ja uudelleensyntymisen peruskäsityksenä ja henkinen aikomus ja aikaansaaminen (organisation, abhisamskāra/pāli: abhisamkhāra¹⁴³). Saṃskāra ja abhisamskāra edustavat niin buddhalaisuuden ajatusmaailman ydintä kuin ajallisuuden ja jatkuvuuden tulkintaakin. Käsite saṃskāra esiintyy ihmisen muodostavan viiden elementin joukossa neljäntenä, ja tämä voidaan kääntää ”henkiset muodostelmat” tai ”perityt voimat”, mielihalut, toiveet tai taipumukset¹⁴⁴. Lisäksi saṃkhāra esiintyy ehdonvaraisen seuraavuuden listassa toisella paikalla tietämättömyyden ja tietoisuuden välissä.¹⁴⁵

Toimijassa erinäisten syiden seurauksena ilmenevä tietoisuus ilmaistaan theravāda-buddhalaisuudessa käsitteellä rakennetietoisuus (construction consciousness, abhisamkhāra-viññāṇa). Toimijan teot, hyveelliset, paheelliset tai neutraalit, muodostavat sen, miten kärsimys tai ansiot mahdollisesti koetaan uudelleensyntymässä. Toimijan teot ovat näin ehtoja ehdonvaraisen seuraamuksen periaatetta seuraavalle tulevalle tietoisuuden ilmenemiselle. Collinsin mukaan rakennetietoisuus on tapa määritellä, miten tietoisuus on muodostunut ilmiö. Siten rakennetietoisuus ei eroa tietoisuudesta. Rakennetietoisuus ei ole persoonallisen tietoisuuden pidentymä, vaan luomus, jonka ovat luoneet ei-persoonallisen tietoisuuden hetkellisten tapahtumien kertyminen. Collins kuvaa rakennetietoisuutta ikään kuin valkokankaaksi, jolle heijastetaan ihmisen eri elämät tapahtumiseen.¹⁴⁶

Buddhalaisuuden tavoitteena on, ettei uutta tietoisuutta enää muodostuisi. Collins tekeekin eron rakennetietoisuuden (construction consciousness) ja muodostuneen tietoisuuden (constructed consciousness) välille. Jälkimmäisen on saavuttanut valaistunut henkilö, joka ei enää koe uudelleensyntymää. Valaistuneella on siis enää muodostunut tietoisuus, hän ei enää luo uutta tietoisuutta. Hän on tavallaan saavuttanut vapautumisen (nirvāṇa/nibbāna), mutta

¹⁴² Collins 1982, 200–201.

¹⁴³ Termi juontaa sanskritin verbistä *abhisamskṛ*, ”saada jotain valmiiksi”, tämä verbi yhdistetään myös bramiiniuskossa keskeisen uhriteon valmiiksi saattamiseen.

¹⁴⁴ Gethin 1998, 135; Collins 1982, 202.

¹⁴⁵ Collins 1982, 201–203.

¹⁴⁶ Collins 1982, 205, 208.

elää vielä maan päällä – Collinsin mukaan – nirvāṇaa (nibbāna) elämässä (nibbāna-in-life), jolloin hänellä on nimi ja muoto, mutta kärsimys ja mielihalu ovat kaikonneet. Lopullinen vapautuminen saavutetaan vasta elämän kaikotessa. On huomion arvoista, että myös valaistunut on vielä eläessään edeltävän karmansa vaikutuksen alainen ja näin altis onnettomuuksille, joita edellisissä elämissä mahdollisesti luotu huono karma aiheuttaa.¹⁴⁷

Miten tietoisuuden virta sitten jatkaa soljumistaan? Tästä on buddhalaisessa perinteessä olemassa kansanomainen esimerkki, joka liittyy ravinnon käsitteeseen. Collins kertoo Buddhan opettaneen kaikkien olentojen tarvitsevan elääkseen ravintoa (āhāra). Theravāda-buddhalaisuus ymmärtää tämän neljän ravinnon käsiteellä: fyysinen ravinto, aistivaikutelmat, henkinen tahtominen ja tietoisuus. Nämä neljä ymmärretään jatkuvuuden välineiksi ja ne ylläpitävät olemassa olevia tai olemassaoloa etsiviä olentoja (beings that exist or are seeking to exist) kahdella tavalla. Collins toteaa elinajan olemassaolon ja sen kokemisen kestävät juuri niin kauan, kuin olenolla on ravintoa. Vertauksena, lamppu palaa juuri niin kauan kuin siinä on öljyä. Mielihalu ravintoon vaikuttaa tulevaan elämään ja sen syntyyn tavanomaisen ehdonvaraisen seuraamuksen virtauksen mukaan. Ehdonvaraisen seuraamuksen ilmiöissä ravinto (āhāra) liitetään yhdeksänteen, tarrautumiseen. Tämä tarrautuminen merkitsee Collinsin mukaan uudelleensyntymää kokonaisuudessaan. Collins kertoo myös, että buddhalaisissa teksteissä mainitaan kaiken syntyvän, ja siten ehdonvaraisen seuraamuksen mukaan etenevän, olevan lähtöisin ravinnosta. Mitä sitten on tietoisuus-ravinto? Collins toteaa Buddhan opettaneen se ”olevan syy uudelleensyntymälle¹⁴⁸”. Ehdonvaraisen seuraamuksen mukaan, milloin mielihalua mihinkään neljästä ravinnosta ilmenee, tietoisuus saa tällöin mahdollisuuden kasvaa. Tästä on seurauksena nimi ja muoto, edelleen aistit ja näin elämän virta jatkaa juoksuaan eteenpäin.¹⁴⁹

Toinen kansanomainen esimerkki tietoisuuden ilmenemisestä on uskomus Gandharva (pāli: Gandhabba) -hengen olemassaolosta ja mielikuva syntymän hetkellä laskeutuvasta nimestä-ja-muodosta, tietoisuudesta tai Gandharva-hengestä. Intian mytologiassa Veda-kirjallisuudessa Gandharva oli taivaallinen henki, muusikko, joka asui pilvien kaupungissa. Bramaaniuskon Upaniṣad-

¹⁴⁷ Collins 1982, 206–207.

¹⁴⁸ “Consciousness-food is the cause of the appearance of rebirth in the future.” Collins 1982, 210, käännös oma.

¹⁴⁹ Collins 1982, 209–210.

teksteissä veden kiertokulku pilvistä maahan nähtiin saṃsāran kierron vertauskuvana. Gandharva liitettiin myös mytologiaan kosmisista vesistä, jotka olivat olemassa ennen maailman luomista. Näistä ajatuksiin linkittyen Gandharva liitettiin elämän luomisen ajatukseen, syntymästä kohdusta. Theravāda-Buddhalaisissa teksteissä Gandharva esiintyy yhtenä monien henkien joukossa. Tietoisuuden laskeutuminen voidaan tavoittaa useista Buddhan puheista, muun muassa Visuddhimagga-teksti selittää tietoisuuden uudelleen linkittymisenä ja nimen ja muodon laskeutumisenä. Collins kertoo, että hieman modernimmassa tulkinnassa Nyanatiloka kertoo karmaenergiasta, joka lähtee kuolleesta ja asettuu uudestaan syntyvään. Buddhan vakaa mielipide oli, että tietoisuus ei pysy muuttumattomana, mutta on Collinsin mukaan selkeää, että theravāda-buddhalaisissa tulkinnoissa tietoisuuden erkaneminen tarkoittaa kuolemaa ja tietoisuuden laskeutuminen syntymää.¹⁵⁰

Tietoisuus käsitetään siis avaimeksi jatkuvuuteen ajassa. Collinsin mukaan tähän on piiloutunut paradoksi, koska sielun olemassaolo yhtä kaikki kielletään. Paradoksi kuitenkin ratkaistaan toteamalla, että tietoisuus muuttuu jatkuvasti, eikä pysy näin ollen pysy samana. Vaikka tietoisuus siis ei ole yhtä kuin oma ja persoonallinen sielu, se on silti jokin, josta jatkuvuus ja elämien jaksot uudelleensyntymän kierrossa (saṃsāra) rakennetaan. Tietoisuus tekee kodin neljään muuhun yksilön osatekijään ja näin tavallaan ankkuroituu näihin. Näin tehdessään tietoisuus säilyy (maintains, persists), muodostaen näin valaistumattoman, elämänkierrollisen ilmiön (samsaric phenomenon).¹⁵¹

Ihmispersonallisuuksien ketju (atthāva) on niin ikään tulosta seurauksesta tai ehdollistamisesta. ”Ihmisnukke” ilmenee, ei itseyden seurauksena, vaan seurauksena syystä. Buddhalaisuus käyttää paljon vertauksia luontoon, kasveihin ja siemenestä kasvamiseen. Samoin, kuin siemen kasvaa pellolla saadessaan vettä ja ravinteita, ihmiset kehittyvät erilaisiksi, koska karmassa on eriävyyksiä. Eri siemenistä kasvaa eri puita. Ajatus tietoisuudesta joko siemenellisenä, uutta tietoisuutta kasvattavana, tai siemenettömänä, uutta tietoisuutta kasvattamattomana, on hyvin tärkeä theravāda-buddhalaisuudessa. Theravāda -traditiossa käsite henkisestä ilmiöstä, joka varmistaa jatkuvuuden, tunnetaan bhavaṅga-mielenä.¹⁵²

¹⁵⁰ Collins 1982, 210–213.

¹⁵¹ Collins 1982, 214, 216–217.

¹⁵² Collins 1982, 223–224.

Bhavaṅga-käsite ja sen tärkeys voidaan kuvata theravāda-buddhalaisuudessa seuraavasti: Bhavaṅga-mielen tuokioiden olemassaolo takaa yhden karmisen yksikön (toimijan, persoonan) identiteetin ja jatkuvuuden läpi elämän ja elämien sarjan useiden uudelleensyntymien läpi. Nämä tuokiot ilmenevät silloin, kun muuta henkistä toimintaa tai tietoisia tapahtumia ei tapahdu.¹⁵³

Collins lainaa Nyanatilokaa:

Alitajuisen elämänvirtauksen tai bhavaṅga-sotan olemassaolo on välttämätön oletamus ajattelussamme . . . mikä sitten ikinä muodostaa toden ja perimmäisen ihmisluonnon tai minkä tahansa olennon luonnon, on tämä alitajuinen elämänvirta . . . [Se on] elämän *sine qua non*, jolla on prosessin, kirjaimellisesti juoksun tai virtauksen (sota) luonne.¹⁵⁴

Edelleen Collins lainaa Shwe Zan Aung'n toteamusta, miten olemassaolon virtaus on välttämätön edellytys tietoisuuden ilmenemiselle ja syy elämän olemassaololle. Hän lainaa myös E. Saratchandran toteamusta: bhavaṅgan käsite voi "antaa tyydyttävän teorian persoonallisuudesta ja eloonjäämisestä ilman käsitettä pysyvistä sielunoloisesta olomuodosta¹⁵⁵". Bhavaṅga selittää jatkuvuuden ja uudestisyntymisen yleisesti. Käsitteeseen ei kuulu vain kuoleman ja uudelleensyntymisen kierto, vaan – Collins lainaa F. Storeyn toteamusta – se myös sisältää seuraavan ajatuksen:

Kuoleman ja uudelleensyntymisen prosessi on itseasiassa jatkuva läpi elämän, koska tietoisuus sisältää sarjan ajatushetkiä, *citta-vīthi* (tiedostamisen toimintoja), jotka ovat kuin *bhavaṅgan*, tai tiedostamatonta elämänjatkumon, yhdistävän nauhan pujotettuja helmiä. Jokainen tiedostettu hetki, kun se syntyy ja hiipuu pois, on kuin pieni syntymä ja pieni kuolema.¹⁵⁶

Bhavaṅga-ajatusmallin kehittymisen taustalla theravada-ajattelussa on hetkellisyyden teoria. Jatkuvuus koostuu tuokioista, hetkellisistä materiahiukkasista, sekä tietoisuudesta että tiedostamattomuudesta. Mieli tietoisessa tilassaan tunnetaan käsitteellä *citta-vīthi* ja tiedostamattomassa tilassa käsitteellä *vīthi*-mutta. Jälkimmäinen käsite tunnetaan myös termillä bhavaṅga. Vaikka bhavaṅga on samoin kuin muutkin tekijät hetkellinen ja ilmenevä,

¹⁵³ Collins 1982, 247.

¹⁵⁴ "The existence of the sub-conscious life stream, or *bhavaṅga-sota*, is a necessary postulate of our thinking . . . whatever constitutes the true and innermost nature of a man or any other being is this subconscious life-stream . . . [it is] the *sine qua non* of life, having the nature of a process, lit. a flux or stream (*sota*)."¹⁵⁴ Collins 1982, 226, käännös oma. Pisteet, sulkeet ja kursiivit kuten alkuperäisessä tekstissä.

¹⁵⁵ "[g]ive a satisfactory theory of personality and survival without introducing a permanent and unchanging entity like a soul." Collins 1982, 226, käännös oma.

¹⁵⁶ "The 'dying and being reborn' process is actually continuous throughout life, for consciousness consists of a succession of thought-moments or *citta-vīthi* (Courses of Cognition), which are like beads strung on the connecting thread of bhavaṅga, or the unconscious life-continuum. Each conscious moment in its arising and passing away is a little birth and a little death." Collins 1982, 226, käännös oma.

olemassaoleva ja kuihtuva, sen oletetaan olevan läsnä kaikkina hetkinä, jolloin aktiivinen tietoisuus ei ole.¹⁵⁷

Bhavaṅga ymmärretään olemassaolon ehdoksi kahdella tavalla. Se nähdään olennaisena osana yksittäistä nimi-ja-muoto -entiteettiä, kausaalisesti niin syynä kuin seurauksena; rakentavana (construct-ive) ja rakennettuna (construct-ed). Bhavaṅga ymmärretään myös edellytyksenä mille tahansa tietoiselle kokemukselle. Henkiset ilmiöt voivat syntyä vain bhavaṅgan pohjalta. Siten on myös sanottu, että vaikka karma (kamma) edustaa minkä tahansa syntyneen mielen yleistä olotilaa, niin bhavaṅga on ehto elävälle mielelle. Bhavaṅgan virta mahdollistaa hetkittäisten henkisten ilmiöiden yhteenliittymisen tietoiseksi ilmiöksi. Bhavaṅga edustaa theravāda-buddhalaisuudessa useita eri tehtäviä niin uneen, havainnointiin, tiedostamiseen, kuoleman ja uudelleensyntymän prosessiin kuin meditaatiossa ”lakkaamisen” (cessation) tapahtumaan liittyen. Tunnetaan myös käsite bhavaṅga-citta, puhdas, valaistunut mieli. Käsitys mielestä alkuperäisessä olotilassaan puhtaana, valoisana ja kirkkaana tunnetaan läpi buddhalaisuuden historian kaikissa osissa Aasiaa, siinä missä mieli tavanomaisesti on likaantunut ja tuhrainen johtuen toimijan mielihaluista. Collins toteaa, että vaikka theravāda-koulukunta ei jalostanut tätä oppia mielestä yhtä hienopiirteiseksi kuin useat muut koulukunnat, on löydettävissä viitteitä bhavaṅga-mielestä, jonka tila on puhdas ja valaistunut. Yksi kuuluisimmista tekstikappaleista theravāda-kirjallisuudesta kertoo, että ”mieli, ollessaan puhdistunut tärvellyksistä, on sädehtivä¹⁵⁸”.¹⁵⁹

Tietoisuus on siis keskeinen elementti theravāda-buddhalaisessa ajattelussa. Tulevaan karmaan (kamma) nähdään vaikuttavan sisäinen itseymmärrys, aikomus. Kuten yllä totesin, valaistumisen keskiössä on viisauden saavuttaminen tietoisesti suoritettuna itsen havainnoinnin avulla. Keskeistä on se, että tietoisuus käsitetään avaimeksi jatkuvuuteen ajassa. Näin voidaan ymmärtää, että elämän kierto ja valaistuksen saavuttaminen ei ole mahdollista ilman tietoisuuden ilmenemistä.

¹⁵⁷ Collins 1982, 226, 238.

¹⁵⁸ “Mind, when freed from adventitious defilements, is luminous.” Collins 1982, 247, käännös oma.

¹⁵⁹ Collins 1982, 239–240, 246–247.

3. Robottien tietoisuus, moraalit ja vapaa tahto?

Johdannossa kerroin James Hughesin todenneen buddhalaisen opillisen ajattelun suhtautuvan verrattain positiivisesti ajatukseen mahdollisesta tietoisuuden ilmenemisestä koneessa. Hänen lauseensa kuuluu sananmukaisesti seuraavasti:

Toisin kuin uskonnot, joille ihmismuoto edustaa ainutkertaisuutta, mikä tekee keinotekoisesta mielestä mahdottomuuden, buddhalaiset ovat avoimempia ajatukselle tietoisuuden ilmenemisestä koneissa.¹⁶⁰

Tämä siis edellyttäisi, Dalai Lamaan viitaten, että koneen fyysinen olomuoto antaa mahdollisuuden tälle tietoisuuden ilmenemiselle. Edellytyksenä tietoisuuden ilmenemiselle, laskeutumiselle, buddhalaisen opin mukaan olisivat siis muoto, ajatukset ja tunteet, käsitteen muodostus sekä halut ja toiveet. Tässä luvussa pohdin, voiko koneen olomuoto olla valmis tähän tietoisuuden ilmenemiseen. Samoin pohdin edelleen, voiko robotti olla moraalinen toimija ja voiko robotilla olla vapaa tahto? Toimijuus lienee kuitenkin kehitystyön tahtotila. Lainaan euRobotics AISBL'n laatimasta strategisesta tutkimusagendasta,

Kasvaneen tietoisuutensa ja helppokäyttöisyytensä kanssa robotit edustavat uuden aikakauden koittoa, kaikkialla läsnäolevina apulaisina, jotka edistävät kilpailukykyä ja elämänlaatuamme. Robotiikasta on tulossa liikkeellepaneva teknologia, jolle perustuu kokonaan uusi sukupolvi autonomisia ja tiedollisia laitteita, jotka tarjoavat puuttuvan lenkin digitaalisten ja fyysisten maailmojen välillä.¹⁶¹

Tavoitteena on siis ”uusi sukupolvi autonomisia ja tiedollisia laitteita”. Pohdin tässä luvussa näiden laitteiden kehitystyötä robotin mahdollisen tietoisuuden, moraalin ja vapaan tahdon näkökulmasta. Keskustelen seuraavassa robotin älykkyydestä ja siitä miten tämä mahdollisesti linkittyy tietoisuuden käsitteeseen. Tutkin, miten moraalit on pyritty implementoimaan koneeseen, robottiin. Lopuksi keskustelen autonomiasta, vapaasta tahdosta, ja mitä tämä voisi olla, kun toimijoina olisi jo mainittu ”uusi sukupolvi autonomisia ja tiedollisia laitteita”.

¹⁶⁰ “Unlike faiths that posit some uniqueness to the human form that would make artificial minds impossible, Buddhists are more open to the possibility of consciousness instantiated in machines.” Hughes 2012, 69, käännös oma.

¹⁶¹ “With their increased awareness and ease of use, robots represent the dawn of a new era, ubiquitous helpers improving competitiveness and our quality of life. Robotics is set to become the driving technology underpinning a whole new generation of autonomous devices and cognitive artefacts, providing the missing link between the digital and physical worlds.” EC, Robotics 2020, 15, käännös oma.

3.1 Älykkäät koneet

Lienee mielekästä aloittaa pohtimalla, mitä on keinoäly ja millaista koneen älykkyys nykytutkimuksen tuntemassa muodossa on. Keinoäly on nykyisessä kehitysvaiheessaan – kuten Wallach ja Allen ovat todenneet – sidottu digitaaliseen laskentaan ja sitä kehitetään digitaalisen alustan päällä toimivana ohjelmistona. Kehitystyö etenee myös tutkimuslaboratorioissa, eikä ehkä ole poissuljettua, että tulevaisuuden keinoäly syntyisi petrialjoissa keskusyksiköiden sijaan. Tutkimuksessani keskityn kuitenkin keskustelemaan kehollisesta keinoälystä, joka on fyysisen rakenteensa perusteella selkeästi kone, ei organismi. Mikäli keinoälyn älykkyys rinnastetaan laskentakykyyn, tietokoneet kehittyvät yhä älykkäämmiksi. Ne eivät ole toistaiseksi ihmisaivojen tasalla, mutta rohkeimmat ennusteet väittävät, että ihmisaivojen hermosolumäärää vastaava määrä bittejä on saatavilla kotitietokoneisiin 2020 mennessä. Tätä näkemystä edustavat niin kutsutun vahvan AI:n kannattajat, jotka näkevät oikein ohjelmoidun koneen olevan mieli. Ray Kurzweil on ennustanut tietokoneen ylittävän ihmisen älykkyuden noin vuonna 2029, jolloin ihmiskunta siirtyy Singulariteettiin.¹⁶²

Nick Boström toteaa singulariteettiin läheisesti liittyvän idean olevan myös älykkyysräjähdys (intelligence explosion), mahdollisuus koneiden superälykkyteen. Koneiden ihmisentasoista älykkyyttä on hänen mukaansa pohdittu koneiden keksimisestä 1940-luvulla lähtien. Silloin ihmisentasoinen älykkyys ounasteltiin tavoitettavan kahdenkymmenen vuoden aikajänteellä. Ennuste on tämän jälkeen elänyt vuoden kerrallaan siten, että tänäkin päivänä keinoälyn ennustetaan olevan todellisuutta kahden vuosikymmenen kuluttua. Boströmin esittämän tutkimuksen mukaan keinoälyä tutkivista asiantuntijoista uskoi koneiden ihmisentasoisen älykkyuden olevan saavutettavissa 10 % mahdollisuudella vuoteen 2022 mennessä, 50 % mahdollisuudella vuoteen 2040 mennessä ja 90 % mahdollisuudella vuoteen 2075 mennessä.¹⁶³

Koneet kehittyvät siis väijäämättä älykkäämmiksi. Mutta onko älykkyys riittävä tekijä tietoisuuteen? Jos älykkyys olisi riittävä tekijä tietoisuuteen, onko

¹⁶² Wallach & Allen 2009, 55–57.

¹⁶³ Kyselyt suoritettiin Philosophy and Theory of AI (Thessalonki, 2011, vastausaste 43/88) sekä Artificial General Intelligence ja Impacts and Risks of Artificial General Intelligence (Oxford, 2012, vastausaste 72/111) -konferenssin osallistujien parissa, sekä Greek Association of Artificial Intelligence –järjestön jäsenien (2013, vastausaste 26/250) ja 100:n eniten viittauksia saaneen keinoälyä käsitelleen kirjoittajan joukossa (2013, vastausaste 29/100). Boström 2014, 1–4, 19, 264.

tämä tietoisuus kuitenkin rinnastettavissa fenomenologiseen tietoisuuteen sellaisena kuin ihmiset biologian organismeina sen kokevat. On kuitenkin muistettava, että kone ja ihminen ovat keskeisesti erilaisia. Koneet ovat nopeampia päätöksenteossa ja pystyvät käsittelemään suurempaa tietomäärää, olettaen, että tieto on oikeassa muodossa. Keinoälyllä ei toistaiseksi ole tunteita mahdollisesti vääristämässä moraalisen päätöksenteon tulosta, mistä voi olla etua. Ihminen on kuitenkin parempi toimimaan vaillinaisen tiedon varassa ja tekemään päätöksiä epäselvissä tilanteissa. Mikä on Wallachin ja Allenin mukaan keskeistä, on se, että ihmisen loogisuus rakentui aistinvaraisella ja tunteellisella, biokemiallisella alustalla, jolloin olennaista oli vain henkiinjääminen ja lisääntyminen. Järki syntyi tunneivojen pohjalle, kun taas nykyteknologia yrittää kehittää mahdollisesti biokemiallista, mahdollisesti tuntevaa, entiteettiä digitaalisella ja loogisen toiminnan alustalla.¹⁶⁴

Koneet eittämättä pystyvät käsittelemään suurempaa tietomäärää, kuin ihmiset. Voitaneen kuitenkin oikeutetusti pohtia, riittääkö tämä. Siirrynkä nyt pohtimaan, miten viimeaikainen tutkimus on pyrkinyt muuttamaan tiedon tietoisuudeksi ja miten tietoisuus käsitteenä ymmärretään keinoälyä kehitettäessä.

3.2 Robotin tietoisuus

3.2.1 Tietoinen kone

Esitän tässä luvussa viimeaikaisessa tutkimuksessa urjenneita ajatuksia tietoisien robotin olemuksesta. Pentti Haikonen on pohtinut robotin tietoisuutta kirjassaan *Robot Brains: Circuits and Systems for Conscious Machines*. Hän kuvaa tietoisien toimijan olevan ”tietoinen olemassaolostaan mielekkäästi havaitussa maailmassa¹⁶⁵” ja kuvaa tietoisien koneen (conscious machine) pystyvän muun muassa tuntemaan ja havaitsemaan sekä käsittelemään tietoa ja ajattelemaan ihmisen lailla. Keskustellessa koneen tietoisuudesta – Haikonen toteaa – on kuitenkin tehtävä ero kahden käsitteen välillä: tietoisuuden sisältö ja mekanismit. Koneen tietoisuuden tutkimus keskittyy pääosin tietoisuuden mekanismeihin, ei tietoisuuden sisältöön.¹⁶⁶

¹⁶⁴ Wallach & Allen 2009, 71.

¹⁶⁵ “[A conscious agent is] aware of its existence in a meaningfully perceived world.” Haikonen 2007, 185, käännös oma.

¹⁶⁶ Haikonen 2007, 185.

Haikonen korostaa erityisesti kahta asiaa olennaisina tietoisuuteen liittyen, kykyä käsitellä merkityksiä ja havaintokykyä. Havainnot ja tietoisuus ovat Haikosen mukaan toisiinsa liittyneitä ja tietoisuus itsestä syntyisi robotin kyetessä havainnoimaan oman kehonsa tiloja.¹⁶⁷ Havainto kehollisista ja henkisistä tapahtumista ei kuitenkaan itsessään ole riittävä. Haikonen keskustelee tietoisuuteen liittyvän myös tietoisuus itsestä. Tähän tarvitaan hänen mukaansa käsitys kokevasta itsestä, minästä. Haikonen toteaa minän viittaavan tässä yhteydessä itseyteen. Hän kuitenkin jatkaa pohdintaa, tarvitseeko kone itseyttä ollakseen itsetietoinen ja tulee päätökseen, ettei näin ole. Kokijana ei hänen mukaansa koneen ollessa kyseessä ole ”minä itse”, vaan koko systeemi – kokonaisuus. Haikonen kertoo koneen ollessa kyseessä itseyden olevan kielellinen illuusio ja tiedollisen systeemin voivan toimia hyvin ilman tällaista erillistä itseä. Jotta robotti voisi mielekkäästi käsitellä tuntemuksiaan, sen täytyy kuitenkin saada käsite, jonka avulla kommunikoida omasta olotilastaan. Toisin sanoen, robotille olisi annettava nimi, jonka se voisi yhdistää kokemuksiinsa. Haikonen toteaa kuitenkin abstraktien käsitteiden juontavat konkreettisista käsitteistä.¹⁶⁸

Voisi siis olla mahdollista, että robotti esitellessään itsensä voisi toimia kuten buddhalainen munkki Nāgasena esitellessään itsensä kuningas Milindalle. Esitellessään itsensä tietoinen robotti voisi todeta nimensä olevan vain nimilappu, joka viittaa osatekijöistään koostuvaan olentoon Itsenäistä itseä nimilapun takana ei kuitenkaan olisi. Robotille olisi näin verrattain helppoa kieltää kuvitelma ”minä olen”, asmi-māna.

Haikonen erottelee tietoisuudesta käsitteet sisältö ja mekanismit ja keskittyy keskustelemaan tietoisuuden mekanismeista. Tietoisuuteen liittyy kuitenkin myös sisältö. Voidaan pohtia, onko tämä sisältö tavoitettavissa laskennallisesti. Voidaanko tietoisuus sieventää laskennaksi? Esimerkiksi John Searle on nähnyt, ettei ohjelman suorittaminen voi olla yhtäläistä ymmärryksen kanssa. Hän esitti tätä kantaa mukailevan kiinalaisen huoneen argumenttinsa vuonna 1980. Kiinalaisen huoneen argumentissa henkilö istuu huoneessa edessään kiinalaisia merkkejä. Nämä edustavat tietokoneen tietokantaa. Hänen edessään on myös kirja, jossa on ohjeita erilaisten kiinalaisten merkkien yhdistämiseen ja tuottamiseen toisten merkkien pohjalta. Tämä edustaa tietokoneen ohjelmaa. Henkilö ei kuitenkaan ymmärrä kiinaa ja ei siten tiedä, mitä merkit tarkoittavat. Sitten hän saa huoneen

¹⁶⁷ Haikonen 2007, 186–187, 191.

¹⁶⁸ Haikonen 2007, 192–193.

ulkopuolelta viestissä merkkijonon, joka on kiinankielinen kysymys. Tämä merkkijono edustaa tietokoneen syötetietoja. Henkilön on nyt ohjeiden pohjalta muutettava nämä merkit toisiksi merkeiksi, vastaukseksi kysymykseen. Hän toimii näin kirjan ohjeiden perusteella ja lähettää muodostamansa uudet merkit vietäväksi huoneen ulkopuolelle. Uudet merkit edustavat tietokoneen tulostustietoja. Henkilö huoneessa ei tiedä, että merkit ovat kiinaa. Hän ei tiedä, että merkit ovat täysin ymmärrettäviä ja järkeenkäypiä kiinaa äidinkielenään puhuvalle. Henkilön tuottamat viestit ovat oikein, mutta hänellä itsellään ei ole tästä mitään tietoa. Hän vain yksinkertaisesti sijoittaa merkkejä ohjeiden perusteella toistensa eteen. Voisi kuvitella, että henkilö huoneessa puhuu sujuvaa kiinaa, vaikka hän todellisuudessa ei näin tee. Tällä argumentilla Searle perustelee, että vaikka kone osaisi ohjeiden perusteella tuottaa oikeita merkkejä, sillä ei silti ole ymmärrystä toiminnastaan. Tämä johtuu siitä, ettei tietokone (henkilö huoneessa) tunne semanttisia sääntöjä, jotka antavat merkeille merkityksen. Ymmärrys edellyttää semanttisten, ei vain syntaktisten, lauseopillisten, sääntöjen käyttöä, ja ihmisten tulee hyödyntää näistä molempia voidakseen ymmärtää koko asiasisällön.¹⁶⁹

Searle muotoili argumenttinsa jo vuosikymmeniä sitten, mutta hän ei kuitenkaan ole muuttanut kantaansa koskien komputationalismia – teoriaa, jonka mukaan tietoisuus olisi tietokoneohjelma, tai useita samaan aikaan käynnissä olevia ohjelmia. Searlen argumentti oli, ja on, että ohjelman suorittaminen ei luo tietoisuutta, koska ohjelman sisältämät syntaktiset säännöt eivät tuo koneelle mahdollisuutta ymmärtää semanttisia sisältöjä ja merkityksiä.¹⁷⁰ Kone toimii parhaimmillaankin ”ikään kuin” se ymmärtäisi asioita ja olisi tietoinen. Myös Steve Torrance toteaa ehkä voitavan olettaa, että hienostuneinkin keinoäly tulee aina olemaan tietoisuutta vailla oleva ”käyttäytyjä” (behaveer). Keinoäly ei siis vielä voisi saavuttaa aitoa, fenomenaalista ja laadullista, tietoisuutta.¹⁷¹

Vaikka keinoäly ei saavuttaisi fenomenaalista ja laadullista tietoisuutta, Torrance kertoo silti esitetyn ajatusta, että keinoäly voisi olla kognitiivisesti tai toiminnallisesti tietoinen¹⁷². Myös Wallach ja Allen esittävät ajatuksen tietoisuuden pohdinnan yksinkertaistumisesta. Tietoisuuden voisi käsittää samoin kuin esimerkiksi lentäminen, toiminnallisena tapahtumana. Lentämisen ydinkysymys ei

¹⁶⁹ Honderich 2014, 306–307.

¹⁷⁰ Honderich 2014, 306–307; Wallach & Allen 2012, 57–58.

¹⁷¹ Torrance 2008, 499.

¹⁷² Torrance 2008, 515.

näet ole toimia samoin kuin linnut. Keskeinen idea on pysyä ilmassa. Wallace ja Allen toteavat, että sama periaate voidaan laajentaa koskemaan tietoisuutta ja lainaavat Stan Franklinin toteamusta keinotekoisen toimijan (artificial agent) olevan toiminnallisesti tietoinen, jos se pystyy suorittamaan useita niistä tehtävistä, joita ihminen pystyy suorittamaan tietoisuutensa perusteella. Vaikka robotit eivät olisi täysin samalla tavalla tietoisia kuin ihmiset, olisi ehkä mahdollista suunnitella ne siten, että niillä on silti samat kyvyt.¹⁷³

Koneen tietoisuutta ovat pohtineet myös Christof Koch ja Giulio Tononi artikkelissaan ”Can Machines Be Conscious?” He toteavat, että vaikka vielä ei pystytäkään toteamaan, mitä tietoisuus on, pystytään silti toteamaan, mitkä eivät ole tietoisuuden edellytyksiä. He näkevät asian seuraavasti: tietoisuus ei edellytä kykyä kanssakäymiseen ympäristön kanssa, muistia, itsereflektiota, tunteita, huomiokykyä tai kieltä. Tämän he perustelevat siten, että todistettavasti tietoiset ihmiset ovat olleet tietoisia huolimatta jonkun näistä tekijöistä puutteesta. Tietoisuuden edellytys on heidän mukaansa kyky tuottaa *yhdistettyä tietoa* (integrated information). Tietoiselle toimijalle on ominaista omata laaja valikoima tiloja.¹⁷⁴

Mitä sitten on yhdistetty tieto? Koch ja Tononi esittävät esimerkin: kun tietoinen toimija näkee pimeän ruudun, hän ei näe vain valoisan ruudun vastakohtaa. Samalla kun tietoinen toimija näkee pimeän ruudun, hän ei näe ”lukemattomia asioita, joita hän on joskus nähnyt tai tulee näkemään¹⁷⁵.” Kochin ja Tononin mukaan on mahdollista muodostaa teoreettinen kehys, jolla pystytään mittaamaan miten sujuvasti eri neuronirakenteet pystyvät tuottamaan yhdistettyä tietoa. Tämä, yhdistetyn tiedon teoria (intergrated information theory, IIT) perustuu informaatiolaskentaan (mathematics of information) ja kompleksisuusteoriaan (complexity theory) ja sen avulla pystytään päättelemään, minkä määrän yhdistettyä tietoa mikin keskenään kommunikoivia osia sisältävä järjestelmä tuottaa. Määrän merkinä on Φ ja tämä ilmaistaan tavuina (bits).¹⁷⁶

Mitä suurempi arvo Φ :llä on, sitä laajempi tietoisuus mitattavalla entiteetillä on. Koch ja Tononi toteavat yhdistetyn tiedon teorian olevan vakuuttava tapa, jolla mahdollisesti pystyttäisiin jatkossa mittaamaan koneen tietoisuutta. Tämä

¹⁷³ Wallach & Allen 2009, 67–68.

¹⁷⁴ Koch & Tononi 2008, 57–58.

¹⁷⁵ “... uncountable things that you have ever seen or could ever see.” Koch & Tononi 2008, 58, käännös oma.

¹⁷⁶ Koch & Tononi 2008, 58.

tapahtuisi käytännössä pyytämällä konetta kuvailemaan kohtaus kuvassa. Ihminen pystyy heidän mukaansa päättelemään tilanteen eriskummallisistakin kuvista, mutta koneelle tämä haaste. On huomioitava, että vielä artikkelin kirjoitusaikaan – Koch ja Tononi toteavat – kuvia tunnistavaa konetta ei oltu pystytty toteuttamaan.¹⁷⁷

Koch ja Tononi toteavat, ettei tietoisuus edellytä kykyä kanssakäymiseen ympäristön kanssa, muistia, itsereflektiota, tunteita, huomiokykyä tai kieltä. He perustelevat tämän esimerkin avulla esittelemällä todistettavasti tietoisin henkilön, jolta kyseessä oleva yksi ominaisuus puuttuu. Esimerkiksi otsalohkonsa vaurioittaneet henkilöt eivät pysty tuntemaan tunteita, mutta ovat silti tietoisia¹⁷⁸. Kochin ja Tononin argumentaatio vaikuttaa ontuvan, koska he eivät pohdi lainkaan tilannetta, jossa kaikki nämä tietoisuuden edellytykset puuttuvat. He pohtivat kyvyn kanssakäymiselle kenties olevan kenties tarpeellista tietoisuuden kehittymiselle varhaisessa vaiheessa, mutta kokevat tämän olevan jo kokonaan toinen kysymys¹⁷⁹. Koch ja Tononi eivät myöskään tarjoa vakuuttavaa perustetta, että koneen kyky tuottaa yhdistettyä tietoa antaisi sille valmiuden ymmärtää semanttisia merkityksiä, joiden tärkeyttä muun muassa John Searle on tietoisuudessa painottanut.

Olen edellä keskustellut robotin tietoisuuden edellytyksistä. Linaan edelleen Pentti Haikosen toteamusta: ollakseen tietoinen koneen täytyy pystyä käsittelemään merkityksiä. Ei liene liikaa sanottu, että hän on tässä samoilla linjoilla John Searlen kanssa. Heidän näkemyksensä eroavat tämän merkityksen käsittelyn mekanismeissa, miten merkitystä käsitellään. Searle toteaa, ettei ohjelman suorittaminen luo tietoisuutta. Haikonen taas näkee oikein rakennetun ohjelman riittävään tietoisuuteen. Hän toteaaakin tietoisuuden itsestä syntyvän robotin kyetessä havainnoimaan oman kehonsa tiloja. Haikonen toteaa, että tietoinen kone pystyisi imitoimaan ihmisen tietoisuuden prosesseja vaikuttaen tietoiselta¹⁸⁰. Kysymys on, onko tietoiselta vaikuttava tietoinen. Wallach ja Allen totesivat, ettei lentämisenkään tarkoitus ole imitoida lintuja jokaisessa piirteessä, vain pysyä ilmassa.

¹⁷⁷ Koch & Tononi 2008, 58–59.

¹⁷⁸ Koch & Tononi 2008, 57.

¹⁷⁹ Koch & Tononi 2008, 57.

¹⁸⁰ Haikonen 2007, 2.

3.2.2 Tunteet ja aistit tietoisuuden mekanismeina

Ihmisen tietoisuudessa ja koneen mahdollisessa tietoisuudessa on siis yhä viimeaikaisen tutkimuksen valossa eroja. Erot eivät rajoitu vain tähän. Pohdittaessa esimerkiksi (moraalisen) toiminnan luonnetta ja rinnastettaessa ihmisen ja koneen, robotin, toimintaa keskenään, on hyvä muistaa, etteivät ihmisten moraaliset arviot ole suoraviivaisia ja ne ovat harvemmin täysin järkeiltyjä. Usein päätöksenteko on sotkuista ja siihen vaikuttavat muun muassa tunteet, intuitiot, refleksinomaiset vastausmallit, velvollisuudet ja kenties arvio siitä, onko toiminnan tuoma hyöty merkittävää. Wallach ja Allen toteavatkin myös tunteilla ja sosiaalisilla taidoilla on merkitystä mielekkääseen päätöksentekoon järjen rinnalla¹⁸¹. Pohdin seuraavassa tunnetta osana päätöksentekoa.

Filosofia on useissa yhteyksissä korostanut sitä, miten moraalisen päättelyn tulisi olla vapaata tunteiden vaikutuksesta. Tunteiden arvo on silti myös tunnustettu. Myös usein tunteet kieltäviksi mielletyt stoalaiset filosofit pitivät osaa ihmisen tunnekirjosta hyvinä tunteina. Ilo, varovaisuus, tahtominen ja toivominen soveliaissa määrin ja järjen uskomuksia seuraten olivat ihmiselle tarpeellisia¹⁸². Aristoteleen näkökanta oli, että tunteet auttavat määrittelemään mikä on hyveellistä, tosin hyveellisen toimijan tuli hänen mukaansa myös hallita tunteensa. David Hume ajatteli, että ainakin jotkut tunteet ovat mielekkäitä moraalisen elon kannalta, nimittäin myötätunto, sääli, välittäminen ja rakkaus.¹⁸³ Linaan David Humea: ”järki on, ja sen pitäisikin olla, passioiden orja, eikä se koskaan voi teeskennellä muuta virkaa itselleen, kuin palvella ja totella näitä¹⁸⁴.”

Wallach ja Allen keskittyvätkin pohtimaan, miten tunteet ja sosiaaliset taidot voivat toimia lisätietoa antavina kanavina moraalipäätöksiä koskevassa pohdinnassa. Vaikkakin tunteet voivat myös – Wallachia ja Allenia lainaten – hidastaa järkevää päätöksentekoa, ne eittämättä tarjoavat lisäkanavan tiedolle, jota ei muuten voi saada. Tunteet ja tuntemukset, usein tieteessä affektiivisiin niputetut tilat, ovat huomattavia haasteita keinoälylle. Wallach ja Allen toteavatkin, että nykyisen tiedon ja teknisen osaamisen valossa robotit eivät kykene aidosti

¹⁸¹ Wallach & Allen 2009, 139, 177.

¹⁸² Remes 2004, 76.

¹⁸³ Wallach & Allen 2009, 143–144.

¹⁸⁴ “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.” Hume 1739–40, SBN 415, käännös oma.

ymmärtämään ja kokemaan ihmissuhteita määritteleviä ja eettisiä sääntöjä muovaavia tunteita¹⁸⁵.

Voidaan siis pohtia, miten paljon koneen sitten pitäisi tunteiden osalta jäljitellä ihmistä, jotta sitä voitaisiin nimittää moraalitoimijaksi. On tosin – Wallachia ja Allenia lainaten – aivan mahdollista, että kone mahdollisesti tehdessään moraalista ratkaisua, turvautuisi aivan erilaisiin tiedollisiin ja affektiivisiin välineisiin, kuin ihminen. Wallach ja Allen korostavatkin, että vaikka tunteista on hyötyä ihmisten toiminnalle, on harkittava, ovatko tunteiden soveltamisen haitat hyötyjä suurempia keinoälyn ollessa kyseessä. He näkevät, ettei robotin valmistuksen päämäärä ole monistaa jokaista ihmisen piirrettä, vaan kehittää piirteet, jotka mahdollistavat eettisen toiminnan.¹⁸⁶

Aistit ovat tärkeitä ihmisen tuntemuksille ja Wallach ja Allen toteavat mielekkään tavan robotille ymmärtää ihmisen herkkyyttä olisi omata sama herkkyyks. Keinoäly käsittelee aistimuksia toistaiseksi kuitenkin kognitiivisina esityksinä, joten se ei ole kykenevä ihmisen herkkyyteen. Kyky tuntea mielihyvää ja kipua on olennainen osa ymmärrystä, miten mitkään toiminnat vaikuttavat ihmisiin. Koneille tulee olemaan vaikea havaita ja ymmärtää moraalisesti tätä. On kuitenkin mahdollista rakentaa koneita, jotka pystyvät lukemaan ihmisten tunteita ja vuorovaikuttamaan ikään kuin ne ymmärtäisivät ihmisten aikomukset ja odotukset.¹⁸⁷

Vaikuttaa siltä, että koneen tietoisuuden käsitettä lähestyttäessä olisi esitettävä kysymys, keskustellaanko tietoisuuden käsitteestä merkityksessä tietoisuuden sisältö vai tietoisuuden prosessit. Mikäli tietoisuus ymmärretään prosesseina, viimeaikaisen tutkimuksen valossa kyettäneen jäljittelemään näitä prosesseja. Haaste lienee ihmisen ja koneen ero ja etenkin affektiivisten tilojen mallinnus. Mutta ovatko tietoisuus tai tunteet kuitenkaan kriteeri keinotekoiselle moraalille? Keskustelenkin seuraavassa miten keinotekoista moraalialia on lähestytty viimeaikaisessa tutkimuksessa.

¹⁸⁵ Wallach & Allen 2009, 140–141.

¹⁸⁶ Wallach & Allen 2009, 141–142, 149, 178.

¹⁸⁷ Wallach & Allen 2009, 151–152.

3.3 Keinotekoinen moraal

3.3.1 Moraalisen toiminnan kategoriat ja ohjelmoitu moraal

Pureudun tässä alaluvussa yhteen tutkimukseni keskeiseen kysymykseen, voiko moraal

 ilmetä kehollisessa kenoälyssä. Miten tämä nähdään viimeaikaisen tutkimuksen valossa? Wallach ja Allen painottavat moraalisen toimijan ohjelmoinnissa kahden osatekijän tärkeyttä, autonomiaa ja eettistä herkkyyttä. Nykyisen koneiden moraalin he jakavat kahteen segmenttiin; operatiiviseen ja toiminnalliseen moraaliin. Operatiivinen moraal on täysin esineen suunnittelijan ja käyttäjän hallinnassa. Tällainen voi olla vaikkapa turvajärjestelmä, niinkin yksinkertainen asia, kuin esimerkiksi aseen varmistin. Aseen varmistin on hyvä toimija, jos se on asianmukaisesti suunniteltu ja jos sitä käytetään. Toiminnallisen moraalin piiriin taas kuuluvat systeemit, jotka toimivat niin sanotusti autonomisesti, mutta eivät ole millään tavoin tai vain rajoitetusti eettisesti tiedostavia, tai jotka toimivat niin sanotusti epäautonomisesti, mutta Wallachin ja Allenin mukaan eettisesti tiedostavasti.¹⁸⁸

Hyvin pienellä valvonnalla lentoa suorittava lentokoneen autopilotti on heidän mukaansa kel

po nykypäivän esimerkki toiminnallisesta moraalitoimijasta. Se lentää ohjelmansa perusteella itseohjautuvasti. Arkipäiväisesti yleistäen voisi käyttää jopa termiä autonomisesti. Autopilotin niin kutsuttu eettinen tiedostavuus on kuitenkin käytännössä mitätön, rajautuen tarkasti ennalta ohjelmoituihin sääntöihin. Toisentyyppisenä esimerkkinä Wallach ja Allen esittelevät niin kutsutut eettiset tietojärjestelmät, joita hoitohenkilökunta käyttää apunaan hoitopäätöksiä pohtiessa. Näihin on ohjelmoitu laaja tietokanta eettisiä sääntöjä ja edeltäviä ratkaisuja, joiden perusteella ratkaisua etsimään ohjelmoitu kone pystyy antamaan vastauksen eettiseen kysymykseen. Ohjelmalla on näin erittäin laaja, ohjelmoitu, eettinen tietous, mutta se osaa antaa vastauksen vasta esitetyn kysymyksen perusteella, riippuen sille annetusta informaatiosta ja muodosta, jossa tämä informaatio on annettu.¹⁸⁹

Näin ollen, nykyiset koneet ovat Wallach ja Allenin mukaan joko ”eettisesti tiedostavia”, mutta autonomisesti toimintakyvyttömiä, tai eettisesti sokeita, mutta hyvin autonomisesti toimivia. Wallach ja Allen näkevätkin, että keinotekoinen

¹⁸⁸ Wallach & Allen 2009, 25–26.

¹⁸⁹ Wallach & Allen 2009, 26–27.

moraalitoimija toteutuu, kun molemmat ominaisuudet, autonomisuus ja eettinen herkkyys, kasvavat koneessa yhtämittaisesti.¹⁹⁰

Edellä mainitun jaottelun lisäksi Wallach ja Allen viittaavat James Moorin kehittämään hierarkiseen malliin keinotekoisien moraalitoimijoiden luokittelussa. Moorin ensimmäisenä kategoriana on eettinen vaikutus -toimija (ethical impact agent), kone, jota voidaan arvioida sen perusteella, millaisia eettisiä vaikutuksia sillä on tai ei ole. Kone on välineellisesti hyvä, jos sillä tai sen käytöllä on hyvä seuraus. Seuraava kategoriana on epäsuora eettinen toimija. Epäsuora eettinen toimija on kone, joka on ohjelmoitu siten, ettei sillä ole negatiivisia eettisiä vaikutuksia. Toisin sanoen, se ei aiheuta vahinkoa. Tämä toteutetaan panostamalla koneen luotettavaan toimimiseen ja turvallisuuteen. Kolmas kategoriana on selkeä moraalinen toimija (Explicit ethical agent), kone, joka tekee eettisiä ratkaisuja ohjelmaansa, usein käskyihin nojaavaan logiikkaan, perustuen. Kaikkien näiden jälkeen seuraa neljäntenä ja viimeisenä kategoriana täysi moraalinen toimija (Full ethical agent).¹⁹¹

Täysi moraalinen toimija -malli kuvaa täysivaltaisen aikuisen ihmisen toimintaan verrattavissa olevaa toimintaa. Oletus on, että täysivaltaisen moraalinen toimija osaa tehdä selkeitä moraalisia ratkaisuja ja selittää toimintansa. Kategorian kuvaaman toiminnan oletetaan yleensä edellyttävän Wallach ja Allenin mukaan tietoisuutta, tarkoituksellisuutta ja vapaata tahtoa. On huomioitavaa, että jos ihminen ei täytä edellä mainittuja kriteereitä, hänen moraalinen toimijuutensa kyseenalaistetaan. Moorin argumentti on, että koneiden kehityksessä tulee pyrkiä kolmanteen kategoriaan, koneisiin, jotka ovat selkeitä moraalisia toimijoita ja tämän pidemmälle ei Moorin mukaan voida lähitulevaisuudessa pyrkiä.¹⁹²

Moraalisen toiminnan kategorioita pohtien, Wallachin ja Allenin määrittelyt ovat hyvin suoraviivaisia. Tässä voitaneen todeta, etteivät heidän toiminallisen moraalien piiriin määrittelemänsä laitteet täytä fenomenalisessa merkityksessä heidän käyttämiensä käsitteiden ”autonominen” ja ”eettisesti tiedostava” määritelmiä. Autopilotti osaa eittämättä lentää kohteesta A kohteeseen B ilman ihmisen väliintuloa. Huomauttaisiin kuitenkin, ettei tämä täytä kuin hyvin rajatun autonomian tunnusmerkkejä reitin ollessa täysin ohjelman käskyihin perustuva. Edelleen, eettinen tietojärjestelmä ei sekään nykypäivän teknologiaan perustuen

¹⁹⁰ Wallach & Allen 2009, 27, 34.

¹⁹¹ Wallach & Allen 2009, 33–34.

¹⁹² Wallach & Allen 2009, 33–34.

ymmärrä antamiensa vastausten sisältöä eikä näin niiden eettistä arvoa fenomenalisessa merkityksessä. Kyseisiin laitteisiin ohjelmoitu lauseoppi, syntaksi, ei viimeaikaiseen tutkimukseen perustuen pysty tuottamaan tai käsittämään merkityksiä.

Olen edellä esitellyt keinotekoisien moraalin ehtoja sekä keinotekoisien moraalin kategorioita. Seuraavaksi keskustelen tavoista, joilla moraalialia on implementoitu keinotekoiseen moraalitoimijaan. Näitä ovat käskyihin perustuva ylhäältä-alas –moraali, oppimiseen perustuva alhaalta-ylös –moraali sekä näiden implementointitapojen yhdistelmä.

Ylhäältä alas suuntautuva, eettistä toimintaa säätelevä, moraalialia perustuu käskyihin kuten esimerkiksi kymmenen käskyä tai vaikkapa seuraavat Isaac Asimovin tieteiskirjallisuudessa kehittämät robotiikan lait.

(0) Robotti ei saa vahingoittaa koko ihmiskuntaa tai toiminnasta pidättäytymisen kautta sallia ihmiskunnan vahingoittumista; (1) Robotti ei saa tehdä mitään, minkä se tietää vahingoittavan ihmistä tai toiminnasta pidättäytymisen kautta tietäen sallia ihmisen vahingoittumista; (2) Robotin täytyy totella ihmisen sille antamia ohjeita lukuun ottamatta tilanteita, joissa tämä toiminta on ristiriidassa ensimmäisen lain kanssa; (3) Robotin tulee suojella omaa olemassaoloaan lukuun ottamatta tilanteita, joissa tämä toiminta on ristiriidassa ensimmäisen lain kanssa.¹⁹³

Pohtiessa sääntöihin perustuvaa moraalialia ja sen implementointia, olisi hyvä esittää kysymys: ”kenen moraalialia?” Jos etiikka ymmärretään sääntöinä, joita ei tule rikkoa, kohdataan ongelma. Mitkä nämä säännöt ovat ja ovatko ne sellaisia, että niitä saa joissakin olosuhteissa rikkoa, vai sellaisia, joita ei saa milloinkaan rikkoa? Moraalisen robotin rakentaminen ylhäältä alas suuntautuvan moraalin perustalle olisi siten riippuvaista oikeiden sääntöjen ja oikeiden ongelmanratkaisukaavojen löytämisestä. Ylhäältä alas suuntautuvan moraalin lopputuloksena on moraalialia toiminta, mutta on huomioitavaa, että tämä on niin kutsuttua rajattua moraalialia.¹⁹⁴

Perinteiset filosofiset koulukunnat, joille ylhäältä alas suuntautuvaa moraalialia on perustunut, ovat utilitarismi, deontologia ja Aristoteleen hyve-etiikka. Ohjelmoijan tulisi siis tehdä valinta. Ovatko käskyt sellaisia, että lopputuloksena olisi suurin mahdollinen hyvä (Kenelle? Kaikille?), noudattavatko ne ennalta määriteltyjä sääntöjä, kuten Asimovin neljä lakia vai onko käskynä toimia hyveellisesti. Kysymys, voidaanko lainalaista systeemiä noteerata todelliseksi

¹⁹³ “(0) a robot may not harm all humanity or, through inaction, allow humanity to come to harm; (1) a robot may not injure a human being or, through inaction, allow a human being to come to harm; (2) a robot must obey orders given to it by human beings, except where such orders would conflict with the first law; (3) a robot must protect its own existence as long as such protection does not conflict with the first or second law.” Abney 2012, 42–43.

¹⁹⁴ Wallach & Allen 2009, 62, 77–78.

moraaliseksi toimijaksi, on yhä avoin. Tämä on Wallachin ja Allenin mukaan yhtä vastauksen ulottumattomissa kuin se, onko ihmisellä todella vapaa tahto¹⁹⁵.

Wallach ja Allen mukaan James Gips esitti ehkä ensimmäisenä määritelmän utilitaristisen ja seurauseettisen robotin ominaisuuksista vuonna 1991 artikkelissaan ”Towards the Ethical Robot”. Voidakseen toimia seurauksien kannalta mielekkäästi robotilla pitäisi olla (1) kyky kuvata ympäröivä tilanne, (2) saada kuva mahdollisista toimintatavoista, (3) ennustaa seuraava tilanne, mikäli toimintaa sovellettaisiin olevaan tilanteeseen, ja (4) tapa arvioida tilanne hyvyyden ja haluttavuuden kannalta. Wallach ja Allen toteavat, että ylläolevan listan toteuttava ohjelma olisi varsin haasteellinen kirjoittaa. He pohtivat, miten ohjelma voisi ennustaa tulevia tapahtumia ja miten arvottaa eri tilanteita hyvyyden ja haluttavuuden kannalta.¹⁹⁶

Gips esittää artikkelissaan kaavan utiliteettiin: $\sum w_i p_i$, jossa w_i on kunkin henkilön kerroin ja p_i kunkin henkilön hyvä tai onnellisuus. Jo Jeremy Bentham tavoitellessaan moraalista aritmetiikkaa pyrki siihen, että utiliteetti, suurin mahdollinen suurimmalle mahdolliselle joukolle, voitaisiin laskea antamalla numeroarvo mielihyvälle tai -harmille. Mutta hyvyyden ja haluttavuuden määrittely voi osoittautua haasteelliseksi tehtäväksi. Esimerkiksi filosofit Bentham ja Mill olivat esimerkiksi täysin eri mieltä, korreloiko toisen asiasta A saatu mielihyvä jonkun toisen asiasta B saatuun mielihyvään.¹⁹⁷

On pohdittu, että keinotekoisien moraalitoimijain kehityksessä voitaisiin seurata ihmislapsen moraalista kehitystä, jossa oikean ja väärän ymmärrys liittyvät ensin omaan hyvinvointiin, sitten käytökseen toisten ihmisten kanssa toimien, lopulta päädytään niin kutsutun sosiaalisen sopimuksen ymmärrykseen, totellaan lakeja ja tapoja. Tähän liittyy myös huoli ihmisistä, joka voi taas johtaa universaaleihin moraaliprinssiipeihin.¹⁹⁸ Tätä lähestymistapaa kannattavat tutkijat soveltavat keinotekoisien moraalitoimijain rakentamiseen lähestymistapaa, jota kutsutaan alhaalta ylös suuntautuvaksi moraaliksi.

Alhaalta ylös suuntautuvaa moraalialueita lienee ajatellut ensimmäisenä Alan Turing. Hän pohti 1900-luvun puolivälissä mahdollisuuksia kehittää tietokoneelle lapsen mieli, jota kouluttamalla päädyttäisiin aikuisen mieleen. Geneettinen

¹⁹⁵ Wallach & Allen 2009, 83–86.

¹⁹⁶ Wallach & Allen 2009, 86–87; Gips 1991, 2–3.

¹⁹⁷ Wallach & Allen 2009, 86–87; Gips 1991, 2–3.

¹⁹⁸ Wallach & Allen 2009, 108.

algoritmin keksiminen 1975 joudutti edelleen keinotekoisien mielen kehittämisen pohtimista. Tämä johti ideaan keinotekoisesta elämästä, Alife. Alifessa oli alkuun kyse evoluution simuloimisesta virtuaalisessa ympäristössä. Sitten huomattiin, ettei virtuaalinen ympäristö ollut verrattavissa todellisen elämän asettamiin haasteisiin. Evolutionistiseksi robotiikaksi kutsuttu tieteenala on nyttemmin lainannut Alifen käyttämät menetelmät oikeissa ympäristöissä toimivien robottien kehitystyöhönsä. Sekä keinotekoinen lapsen mieli että Alife ovat esimerkkejä moraalin rakentamisesta alhaalta ylöspäin.¹⁹⁹

Samaan aikaan kun geneettinen algoritmi keksittiin, esitti myös E.O. Wilson teorian etiikan evolutiivisesta taustasta. Sosiobiologit ovat pyrkineet näyttämään, että arvot syntyvät evoluution tuloksena. Tämä aate pohjaa osaltaan peliteoriaan vuodelta 1944. Tutkijat John Maynard Smith ja William Hamilton huomasivat vuonna 1973, että peliteoriaa voidaan soveltaa geenitasolla. Keskeinen hämmennyksen aiheuttaja oli näet ollut, miten yhteistyöhaluinen käytös on voinut kehittyä, sillä yhteistyö ei aina ole välttämättä suoraan hyväksi yksilön omien geenien säilymiselle. Peliteorian soveltaminen osoitti, että kaiken ei välttämättä tarvitse olla hyväksi organismille ollakseen hyväksi geenille. Yhteistyön tekeminen ja sosiaalinen käytös edesauttoi geenien säilymistä ja tämä antoi mahdollisuuden olettaa, että sosiaalinen käytös voi olla tulosta evoluutioprosessista. Tästä johtuen on teoretisoitu, että antamalla keinoälyn kehittyä keinotekoisessa ympäristössä, tuloksena olisi ennen pitkää moraalinen toimija. Tätä on kuitenkin kritisoitu siten, että ymmärrys olosuhteista, joissa ihmisen moraalit kehittyi, on täysin vajavaista. Moraali, joka on tulosta keinoälyn kehittymisestä keinotekoisessa ympäristössä, voi olla tyystin erilaista ihmisten moraalisiin verrattua.²⁰⁰

Mutta onko kumpikaan ratkaisu, ylhäältä-alas ohjelmoitu moraalit tai alhaalta-ylös opittu moraalit kestävä ratkaisu moraalin implementoinnissa keinotekoiseen toimijaan? Wallach ja Allen toteavat että myös molempia lähestymistapoja voisi hyväksikäyttää yhtä aikaa. He viittaavat Aristoteleeseen, joka erottelee moraalihyveet erilleen käytännön järjestä ja älyllisistä hyveistä. Älylliset hyveet saattoi Aristoteleen mukaan opettaa, kun taas moraaliset hyveet oli opittava tapojen ja käytännön kautta. Wallach ja Allen päättävät, että keinotekoisien moraalitoimijain ollessa kyseessä olisi täysin mielekästä käyttää eri lähestymistapoja eri hyvelajien implementointiin. Älyllisten hyveiden opetus

¹⁹⁹ Wallach & Allen 2009, 99–101.

²⁰⁰ Wallach & Allen 2009, 101–104.

rinnastuisi ylhäältä-alas -teorianmukaiseen sääntöjen ohjelmointiin, kun taas moraalisten hyveiden ollessa kyseessä voitaisiin käyttää hyväksi alhaalta ylös lähestymistapaa, oppimista käytännön kautta.²⁰¹

Sovellettaessa robottien kehitykseen tapaa, jolla ihmislapset oppivat, puhutaan multimodaalisesta lähestymisestä. Tällöin järjestelmä pystyy samanaikaisesti rekisteröimään ääntä, kuvaa ja kosketusta ja oppimaan puheen ja toiminnan vastavuoroisuutta, jolloin oppiminen on sidoksissa fyysiseen ympäristöön, omaan ja toisten toimintaan. Tähän liittyy keskeisesti ajatus kehollisuudesta. Tämä tarkoittaa, että entiteetin tila minä tahansa hetkenä on määritelty sen suhteessa toisiin objekteihin, entiteetteihin ja toimijoihin ympäristössä. Useat tutkijat uskovat, että tämä oppimistapa pienentää eroa ihmisen älyn ja keinoälyn ymmärtämistavan välillä.²⁰²

Tieteentekijät uskovat, että konnektionismi, samanaikainen jaettu prosessointi, voisi hyvinkin kuvata keinoälyssä sitä, miten hyveet saavutetaan. Jo Aristoteles aikoinaan ajatteli, että hyvän merkitys tiedetään intuitiolla ja opitaan kokemuksen kautta. Wallace ja Allen viittaavat James Gipsin toteamukseen – niin ikään artikkelissaan ”Towards the Ethical Robot” – että konnektionistinen lähestyminen keinoälyyn heijastaa erityisesti Aristoteleen ajatuksia hyveestä. Molemmat korostavat välitöntä, havainnoivaa, ei-merkkeihin sitoutunutta ymmärrystä, joka voidaan saavuttaa harjoituksen kautta. On osoitettu, että keinotekoiset hermosoluverkot ovat kykeneviä oppimaan asioita itsestään paikantamalla tilastollisia säännönmukaisuuksia saamastaan tiedosta. Esimerkiksi Paul Churchland on ehdottanut, että konnektionismi voi olla vastaus myös moraalisen tiedon kasvamiseen. Hän toteaa, että moraalista oppimista ohjaa sosiaalinen kokemus. On kuitenkin huomattava, että konnektionistinen teoria ei selitä sitä, miten hyppy tietoisuuteen mallin olemassaolosta tapahtuu tiedostamattomasta mallien tunnistuksesta.²⁰³

Haasteen tuo myös se, miten hyveet määritellään. Platonin mukaan nämä olivat viisaus, rohkeus, kohtuullisuus ja oikeudentaju. Aristoteles laajensi ja jakoi älyllisiin ja moraalisiin hyveisiin. Akvinolainen lisäsi Paavalia mukaillen teologiset hyveet uskon, toivon ja rakkauden. Hyveidenkään luonteesta tai määrästä, eikä siitä, voivatko ne muuttua yhteisöstä yhteisöön, ei siis olla yhtä mieltä. Wallach ja

²⁰¹ Wallach & Allen 2009, 117–118.

²⁰² Wallach & Allen 2009, 63, 167.

²⁰³ Wallach & Allen 2009, 121–123; Churchland 1993, 302; Gips 1991, 8.

Allen myös kritisoivat ajatusta hyveiden jaosta siisteihin kategorioihin. Haasteet hyveiden listauksessa vaikeuttavat mahdollisuutta soveltaa niitä ylhäältä alas annettuina käskyinä. Wallach ja Allen jatkavat toteamalla, että vaikka hyveet vaikuttavatkin siihen, miten ihmiset pohtivat asioita ja motivoivat tekojaan, itse toimimisen hetkellä harvat päättävät tehdä asian X, koska kokevat omaavansa hyveen Y. Wallach ja Allen huomauttavat toimintaa perusteltavan pikemminkin motivaatiolla, ”on hyvä toimia näin”. Hyveiden toteuttaminen edellyttäisi näin sekä motivaation, että halun toimia.²⁰⁴

Keinotekoisien moraalien implementointia voidaan siis lähestyä ylhäältä alas suuntautuvien sääntöjen, alhaalta ylös suuntautuvan oppimisen tai näiden yhdistelmän kautta. Tosin on haasteellista määrittellä, mitkä ovat ne hyveet, jotka robotille olisi tarpeellista ohjelmoida tai opettaa. Jos moraalit onnistuttaisiin implementoimaan, millaista keinotekoinen moraalit olisi ja milloin robotit olisi keinotekoinen moraalitoimija. Käsittelen tätä kysymystä seuraavassa luvussa.

3.3.2 Keinotekoinen moraalitoimija

Wallach ja Allen toteavat ohjelmistoperustaisen älykkyyden rajojen olevan yhä kiistanalaisia, mutta argumentoivat tästä huolimatta, ettei tämä ole syy hylätä keinotekoisia moraalitavoitteita. Kriteeri moraalitavoitteelle on Wallachin ja Allenin mukaan haasteellinen määrittettävä, mikä tekee keinotekoisien moraalitavoitteiden rakentamisesta niin ikään haasteellista. Wallach ja Allen viittaavat Drew McDermottin ajatukseen, eroista eettisen päättelijän ja eettisen päätöksentekijän välillä.²⁰⁵

Koneiden etiikka (machine ethics) voi viitata McDermottin mukaan kahteen asiaan. Se on joko ihmisten päätöksenteon imitoimista tai ihmisten eettisen päättelyn mallintamista. McDermott kutsuu ensimmäistä *eettisen päätöksenteon ongelmaksi* ja jälkimmäistä *eettisen päättelyn ongelmaksi*. Eettinen päättely ei – McDermottia lainaten – eroa merkittävästi muusta päättelystä, mutta eettinen päätöksenteko on merkittävästi erilaisista verrattuna muunlaiseen päätöksentekoon. Eettinen päätöksenteko edellyttää McDermottin mukaan kykyä tunkea eettinen konflikti, ristiriita omien halujen ja eettisten vaatimusten välillä. Eettinen päätöksenteko edellyttää myös vapaata tahtoa, toimijalla on mahdollisuus olla aito

²⁰⁴ Wallach & Allen 2009, 118–120.

²⁰⁵ Wallach & Allen 2009, 35, 56.

eettinen päätöksentekijä vain, jos sillä olisi tarpeeksi vapaata tahtoa toimia myös vastoin moraalista vaatimusta.²⁰⁶

Wallach ja Allen huomioivat ihmisillä moraalisisina toimijoina olevan vapaus tehdä yhtä hyvin oikein kuin väärinkin ja ihmisten moraalisisina toimijoina tuntevat valmiutta sekä itsekkääseen että epäitsekkääseen käytökseen. He pohtivat silti, onko tämä valinnanmahdollisuus olennaista eettistä toimintaa koskien. Heidän näkemyksensä on seuraava: on vaikeaa määrittää, onko ihmisillä vapaa tahto. Siksi heidän näkemyksensä mukaan keinotekoisen moraalitoimijan mahdollisen vapaan tahdon puute on heikko argumentti kykyä eettiseen päätöksentekoon koskien. He toteavat, ettei se, ettei tätä vapaata tahtoa voisi implementoida keinotekoisen moraalitoimijaan, millään tavalla estä keinotekoisen moraalitoimijan rakentamista.²⁰⁷

Wallachin ja Allenin näkemyksen mukaan keinotekoisen moraalitoimijan luomisessa riittäisi, että ohjelma saataisiin ensin toimimaan *ikään kuin* se omaisi ymmärryksen moraalista. Wallach ja Allen, toteavat niin ikään, että mikä ei vaikuta ulospäin havaittavaan käytökseen, on irrelevanttia.²⁰⁸

Toiminnallinen samankaltaisuus käytöksessä on ainoa asia, joka voi mahdollisesti merkitä jotain keinotekoisista moraalitoimijaa suunnitelmassa²⁰⁹.

Wallachin ja Allenin mukaan kysymyksiin ”[v]oiko (ro)botti, olla moraalinen toimija?” ja ”[m]istä tämän tietää?” ei voi vastata tyydyttävästi, mutta tämä ei toisaalta ole heistä tarpeellista. He toki myöntävät keskustelun asiasta olevan mielekästä, koska epäilykset keinoälystä voivat innoittaa entistä enemmän pohtimaan asioita, kuten. ”Mitä moraaliseen toimijuuteen todella tarvitaan?”. Jos moraalisen ymmärryksen käsitetään tarkoittavan valmiutta toimia asianmukaisesti ja mukautuen fyysiseen ja sosiaaliseen ympäristöön Wallach ja Allen eivät näin näe syytä, miksi kone ei pystyisi tällaiseen toimintaan. Näin ei ole poissuljettua, etteikö tulevaisuus todistaisi moraalisisella tavalla toimivia koneita.²¹⁰

Wallach ja Allen eivät pysähdy pohtimaan moraalista toimijuutta tai toimijuutta yleensä käsitteinä. Näitä käsitteitä on pohtinut esimerkiksi David Gunkel kirjassaan *The Machine Question*. Hän näkee moraalisen toimijan

²⁰⁶ Wallach & Allen 2009, 35; McDermott 2008, 2, 7.

²⁰⁷ Wallach & Allen 2009, 56, 60–62.

²⁰⁸ Wallach & Allen 2009, 58, 68.

²⁰⁹ “Functional equivalence of behaviour is all that can possibly matter for the practical issues of designing AMAs.” Wallach & Allen 2009, 68, käänös oma.

²¹⁰ Wallach & Allen 2009, 58, 66.

edustavan käsitteen toimija alakategoriaa, toimijuudelle ollessa olennaista toimia samalla ollen kykenevä aikomukseen (intention). Moraalinen toimijuus taas voidaan nähdä kykynä toimia moraaliperiaatteiden mukaan, minimitasolla toimia hyveellisesti.²¹¹

Voiko kone olla moraalinen toimija? Gunkel lainaa Kenneth Einar Himman moraalisen toimijan määritelmää artikkelista "Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent?"

X:ää koskee seuraava: X on moraalinen toimija jos ja vain jos X on (1) toimija, jolla on kyky (2) tehdä vapaita valintoja, (3) pohtia, miten tulisi toimia ja (4) ymmärtää ja soveltaa moralisääntöjä oikein tyypillisessä tilanteessa.²¹²

Gunkel kuitenkin huomauttaa, että tämä on vain yksi mahdollisista määrittelyistä.²¹³

John Sullins esittelee artikkelissaan "When Is a Robot a Moral Agent?" neljä eriävää näkökantaa robotin mahdollisesta moraalisesta toimijuudesta. Daniel Dennett esittää ajatuksen (1), jonka mukaan robotit eivät ole toistaiseksi ole moraalisia toimijoita, mutta voivat olla tätä tulevaisuudessa. Vastakkaista näkökulmaa edustaa Selmer Bringsjord. Hän näkee (2), että robotit eivät voi olla milloinkaan moraalisia toimijoita. Hänen mukaansa robotit toimivat aina deterministisesti ohjelmansa ehdoilla. Koska robotilla näin ei ole vapaata tahtoa, se ei Bringsjordin mukaan ole moraalitoimija. Joseph Emile Nadeau puolestaan esittää väitteen (3), että robotit ovat moraalisia toimijoita, mutta ihmiset eivät. Nadeau väittää, että toiminta on vapaata vasta, kun se on täysin järjen tulosta. Hän näkee, etteivät ihmiset pysty toimimaan puhtaan järjen pohjalta. Robotit jälleen voivat tehdä näin. Neljäntenä Sullins esittelee Luciano Floridin ja J.W. Sandersin näkemyksen. He argumentoivat (4), että moraaliteorian monet paradoksit voidaan ohittaa, mikäli jätetään huomiotta vaikeasti filosofiassa todennettavat käsitteet "vapaa tahto" (free will) ja "tarkoituksellisuus" (intentionality). Floridi ja Sanders käyttävät käsitettä "mindless morality". Tämän käsitteen avain on se, että toimijuuteen riittää se mikä näkyy ulospäin.²¹⁴ Wallach ja Allen vaikuttavat myös kannattavan Floridin ja Sandersin näkökulmaa.

²¹¹ Gunkel 2012, 18–19, 22–23.

²¹² "for all X, X is a moral agent if and only if X is (1) an agent having the capacities for (2) making free choices,(3) deliberating about what one ought to do, and (4) understanding and applying moral rules correctly in paradigm cases." Gunkel 2012, 22; Himma 2009, 24, käännös oma.

²¹³ Gunkel 2012, 22.

²¹⁴ Gunkel 2012, 69–70; Sullins 2006, 26–27.

Sullinsin mukaan pohdittaessa robotin moraalista toimijuutta, on pohdittava kysymyksiä onko robotti merkittävän autonominen, onko robotin toiminta tarkoituksellista ja onko robotti vastuullisessa asemassa. Sullins argumentoikin, että robotin moraaliseen toimijuuteen riittää autonomia koneteknisessä mielessä. On riittävää, jos robotti ei ole suoraan käyttäjän hallinnassa. Tarkoituksellisuuden määrittelyyn riittää se, jos robotin toiminta *vaikuttaa* tarkoitukselliselta. Toisin sanoen, jos robotti toimii tavalla, joka voitaisiin tulkita tarkoitukselliseksi. Samoin – Sullins väittää – robotti on vastuullinen, jos voidaan *olettaa* sen olevan vastuullinen. Moraaliseen toimijuuteen riittää siis, että robotin ominaisuuksiksi voidaan uskoa moraalisen toimijuuden piirteet. Sullins huomauttaa, ettei näiden uskomusten pitä olla tosia. Riittää, että ne ovat ilmeisiä.²¹⁵

Vastakkaista mielipidettä edustavat niin kutsutun orgaanisen näkökulman eettiseen asemaan (organic view on ethical status) ottavat. Steve Torrance esittelee tätä näkökulmaa artikkelissaan ”Ethics and consciousness in artificial agents”. Hänen mukaansa orgaaninen näkökulma sisältää seuraavat teesit: (1) koneiden ja biologisten olentojen välillä on kahtiajako, (2) vain aidoilla organismeilla voi olla moraalinen status ja (3) moraalinen ajattelu, tunne ja toiminta ilmenevät orgaanisesti ihmisen biologiaan pohjaavasta historiasta. Orgaaninen näkökulma nojaa fenomenaaliseen tietoisuuteen. Torrance lisää edellisiin kolmeen näkökohtaan kaksi: (4) vain tunteeseen ja fenomenaaliseen tietoisuuteen kykenevät olennot voivat olla moraalisen huolen tai arvioinnin kohteena ja (5) vain biologiset organismit ovat kykeneviä olemaan ”aidosti tuntevia ja tietoisia”²¹⁶.²¹⁷

Keskeinen ajatus on seuraava: moraalit on keskeisesti itsesäilyttävien (self-maintaining) olentojen aluetta, olentojen joilla on luontainen, vaikuttava *telos*. Moraali ei ole ulkopuolisen kokoonpanijan suunnitteleman, rakennetun artefaktin ominaisuus. Orgaanisen näkökulman kannattajat puolustavat voimakkaasti (1) tietoisuuden, elämyksen kokemuksen, (2) biologisten organismien autonomisen itsesäilytyksen ja (3) moraalinormien kytköstä. Tietoisuus on moraalisen aseman ytimessä.²¹⁸

Orgaanisen näkökulman mukaan täysin moraalisesti toimintakykyiset keinotekoiset toimijat eivät ole lähitulevaisuutta. Torrancen mukaan näkemys ei

²¹⁵ Sullins 2006, 28–29.

²¹⁶ “... the ability to be genuinely sentient or conscious.” Torrance 2008, 503, käännös oma.

²¹⁷ Torrance 2008, 502.

²¹⁸ Torrance 2008, 512, 516.

kuitenkaan kiellä, etteivätkö näennäismoraaliset (quasi-moral) keinotekoiset toimijat voisi olla todellisuutta teknologian edetessä. Nämä keinotekoiset toimijat eivät olisi täysiiä moraalisia toimijoita, mutta niihin olisi silti mahdollista rakentaa moraalisen toimijuuden ominaisuuksia. Torrance kirjoittaa, että vaikka orgaaninen näkökulma olisi paikkansapitävä, se jättää silti mahdollisuuksia koneiden etiikan (machine ethics) kehittämiseksi. Torrance mainitsee eettisen päätöksenteon mallinnukset, eettiset neuvonantajat ja näennäispersoonallisen vuorovaikutuksen.²¹⁹

Koneen hyvyttä mitataan toistaiseksi suhteessa toimintaan, johon ne on tarkoitettu. David Gunkel lainaa Jean-François Lyotardia kirjasta *The Postmodern Condition*, tekniikka on ”hyvää”, kun sen avulla saavutetaan tehokkaalla tavalla haluttu lopputulos²²⁰. Pohdittaessa robottia esimerkiksi eettisenä neuvonantajana on kiintoisaa pohtia, miten keinoäly voisi toimia esimerkiksi rawlsilaisena tietämättömyyden verhon takana olevana päättelijänä yhteiskunnassa. Ville Päivänsalon mukaan Rawls näkee laajan ymmärryksen (reason) käytön tärkeänä moraaliseen toimimiseen liittyen. Oikein toimiminen, moraalisuus, perustuu ”vapaiden ja älykkäiden miesten ja naisten jaettuun käsitykseen²²¹”. Nämä toimijat, sellaisina kuin Rawls konstruoi ne varhaisessa tuotannossaan, ovat oleellisella tavalla (relevant sense) järkeviä kyeten harkitsemaan ehdotettuja periaatteita päätöksiä ilman että he vähentäisivät moraalialia hiusten halkomiseen. Rawls on määritellyt näille toimijoille ominaisuudet olla kypsiä, heillä täytyy olla tietoa ja koulutusta, he ovat kykeneviä ja halukkaita loogiseen päättelyyn ja omaavat omatunnon sekä ”sympatiaa ja ymmärrystä” muiden ihmisten intresseille. On huomionarvoista, että päätöksenteon ei hänen mukaansa tule olla kaavamaista, vaan toimijoiden on kyettävä arvelemaan oikean ratkaisun luonnetta ja tähän perustuen edettävä kohti oikeaa ratkaisua. Moraaliperiaatteet tulisi Rawlsin mukaan perustella periaatteella, jota Päivänsalo kutsuu termillä ”laaja ymmärryksen käyttö” (broad use of reason).²²²

Päivänsaloon viitaten, artikkelissaan ”Outline for a Decision Procedure for Ethics” Rawls esittelee käsitteen ”pätevä moraalipäätätjä” (competent moral judge). Rawls näkee tällaisina pätevinä moraalipäätätjinä joukon miehiä ja naisia, jotka

²¹⁹ Torrance 2008, 516–518.

²²⁰ Gunkel 2012, 25.

²²¹ ”in the collective sense of right of free and intelligent men and women”, Päivänsalo 2007, 1, käänös oma.

²²² Päivänsalo 2007, 1–3.

kykenevät tuottamaan oikean ja hyväksyttävän ratkaisun, harkitun päätöksen (considered judgement).²²³

Päivänsalo hahmottelee Rawlsin laajan ymmärryksen käytön periaatetta seuraavasti, periaate P9:

P9 Moraalisen päätöksen tulisi olla harkittu päätös.

Kun otetaan huomioon määritelmät D1 ja D2:

- D1 Harkittu päätös on tulosta prosessista, jossa järkevät ihmiset pohtivat moraalista kysymystä.
- D2 Järkevä henkilö on omaa laajan sympaattisen tiedon ihmisten intresseistä, laajan tiedon maailman tosiasioista, kypsät päättelytaidot ja laajasti tunnustetun näkemyksen moraaliin.²²⁴

Näistä lauseista puuttuvat moraaliset määreet miltei kokonaan. Maininta moraaliarvostuksesta löytyy vain määritelmästä D2. ”Laajasti tunnustettu näkemys moraalin” yhdistää määritelmällisesti järkevän ihmisen laajemman joukon jaettuun moraalitajuun. Rawls ei määrittele, kuinka mitata ymmärrystä omaavan toimijan (persoonan) tietoja, taitoja tai mainetta.²²⁵

Vain laaja ymmärryksen käyttö ei yksin riitä Rawlsille perustelevaan oikeita päätöksiä. Olennaista ovat myös oikeat olosuhteet. Omien etujen ei tulisi tulla järkevän, ymmärryksellä tehdyn, ja vapauden, tasa-arvon ja palkkion huomioivan (F3), päätöksen (P11) tielle. Periaate P11 ja totuudellinen väite F3 alla:

- P11 Järkevän pohdinnan prosessin pitäisi tapahtua olosuhteissa, joissa järkevän toimijan omat edut eivät vaikuta heidän päätökseensä (i.e. tarkoituksenmukaisissa olosuhteissa).²²⁶
- F3 Järkevä henkilö on omaa laajan sympaattisen tiedon ihmisten intresseistä, laajan tiedon maailman tosiasioista, kypsät päättelytaidot ja laajasti tunnustetun näkemyksen moraaliin. Vapauden, tasa-arvon ja palkkion yhteiseen hyvään tehdystä työstä elementit ovat olennaisia ymmärryksellä tehdyssä, järkevässä pohdinnassa (i.e. ne ovat prima facie oikeutettujen periaatteiden elementtejä).²²⁷

²²³ Päivänsalo 2007, 4.

²²⁴ P9; “A moral judgement should be a considered judgement, D1; A considered judgement is an outcome of a process where reasonable persons consider a moral question (i.e. a process of reasonable consideration, D2; A reasonable person is a person who has wide sympathetic knowledge of human interests, wide knowledge of facts of the world, mature skills of reasoning, and widely recognized moral insight.” Päivänsalo 2007, 4, käännös oma.

²²⁵ Päivänsalo 2007, 4–5.

²²⁶ P11; “The process of reasonable consideration should take place in circumstances where reasonable person’s own interests do not bias their judgements (i.e. in appropriate circumstances).” Päivänsalo 2007, 8, käännös oma.

²²⁷ F3; “Elements of liberty, equality, and reward for services contributing to the common good are implicit in the process of reasonable consideration (i.e. they are elements in the prima facie justified principles).” Päivänsalo huomioi, että edetessään pohdinnoissaan Rawls jättää ajatuksen palkkiosta ja keskittyy tasa-arvoon keskeisenä elementtinä (vrt. Outline for a Decision Procedure for Ethics, 1951 ja Theory of Justice, 1971). Päivänsalo 2007, 14, 15, käännös oma.

Rawls näkee oikeat olosuhteet järkevälle päätöksenteolle alkuasetelmana (original position), jossa päätöksiä tekevä toimija on tietämättömyyden verhon (veil of ignorance) takana. Tätä käsitystä on myös kritisoitu muun muassa siksi, että ihmisen alkuasetelmassa lienee vaikeaa ellei mahdotonta, irrottautua identiteetistään ja kokemuseräisistä piirteistään.²²⁸

Ihmiselle toimijana olisi siis hankalaa unohtaa kaikki omat etunsa tai piirteensä ja tehdä päätöksiä täysin alkuperäisessä tilassa tietämättömyyden verhon takana. Entäpä koneelle? Haluan palata hetkeksi Päivänsalon määritelmään D2: ”Järkevä henkilö on omaa laajan sympaattisen tiedon ihmisten intresseistä, laajan tiedon maailman tosiasioista, kypsät päättelytaidot ja laajasti tunnustetun näkemyksen moraaliin.” Kyseiset ominaisuudet vaikuttavat ohjelmoitavissa olevilta. Jos syrjäytämme käsitteen ”henkilö” (person) käsitteellä ”toimija” (agent), olemme määritelmällisesti hieman lähempänä mahdollista oraakkelia, myös keinoälyllä varustettua²²⁹.

Vallalla on siis lukuisia määriä eri näkemyksiä kehollisen keinoälyn kyvystä moraaliseen toimintaan. Vastaus kysymykseen on sidoksissa moraalisen toimijan määrittelyyn. Mikäli toiminnallinen moraalit määritellään riittäväksi, moraalisen robotin kehittämiseen ei liene estettä. Tällöinkin on huomioitava, ettei kyky moraaliseen päättelyyn välttämättä tuo valmiuksia moraaliseen päätöksentekoon. Esimerkiksi niin kutsutun orgaanisen näkökulman kannattajien mielestä vain aidoilla organismeilla voi olla moraalinen status. Tämäkään näkökulma ei kuitenkaan kiellä, etteikö keinoäly voisi saavuttaa näennäismoraalista asemaa ja toimia hyvin yhteisössä esimerkiksi neuvonantajana tai näennäispersoonallisen vuorovaikutuksen tehtävissä.

3.4 Kehollisen keinoälyn vapaa tahto

3.4.1 Järki, autonomia ja vapaa tahto

Tässä vaiheessa näen mielekkääksi jälleen palata hetkeksi kysymyksen: ”mitä on vapaa tahto” ääreen. Mistä tämä vapaa tahto on peräisin ja miten se voidaan määritellä? Pohdiskelin vapaan tahdon käsitettä buddhalaisesta perspektiivistä jo tutkimukseni ensimmäisessä pääluvussa Käsittelem kysymystä seuraavassa

²²⁸ Päivänsalo 2007, 15, 19–20.

²²⁹ Boström 2014, 145–148.

Immanuel Kantin pohdintojen valossa. Keskeinen kysymykseni on, voiko vapaa tahto (mitä se sitten ikinä onkaan) ilmetä kehollisessa keinoälyssä.

Onora O'Neill kertoo Immanuel Kantin kuvanneen tahtoa ”elollisille olennoille ominaiseksi kausaalisuudeksi niin pitkälle, kun ne ovat rationaalisia²³⁰”. Kant erottelee selvästi kausaalisuuden, joka toimii luonnollisesta tarpeellisuudesta tai ulkopuolisten syiden takia, ja kausaalisuuden, joka toimii vapaasti. Vapauden Kant määrittelee alun perin negatiivisesti ominaisuudeksi, jonka avulla tahto voi toimia riippumatta ulkopuolisista syistä. Kant kuvaa seuraavaksi positiivista vapauden käsitettä negatiivisen vapauden käsitteen rinnalla.

Vapaan tahdon (jos tällaista ilmenee) täytyy (kyetä) olla vapaa sekä positiivisessa että negatiivisessa mielessä ja sillä täytyy olla kyky *itsemääräämisoikeuteen* tai *autonomiaan*.²³¹

Positiivinen vapauskäsitte on itsemääräämisoikeutta tai autonomiaa ja yhdenmukaista kategorisen imperatiivin kanssa. O'Neill esittää ajatuksen olevan, että sikäli, kun on olemassa vapaaseen valintaan kykeneviä olentoja, näiden täytyy pystyä soveltamaan toimiinsa lainomaisuutta ja toimimaan yleistettävien ohjeiden (universalizable maxims) mukaisesti.²³²

O'Neill toteaa, että Kant on väittänyt vain näyttäneensä, että

sikäli, kun on vapaita järkeviä olentoja, nämä eivät ole vain negatiivisessa mielessä vapaita, vaan kykeneviä autonomiaan, ja sikäli kun ne ovat kykeneviä autonomiaan, ne ovat moraaliperiaatteen alaisia, toisin sanoen, käytännöllisen järjen ylimmän periaatteen²³³.

Negatiivinen vapaus, riippumattomuus ulkopuolisista syistä on riippumattomuutta sattumanvaraisuudesta, myös toimijan itsensä sisäisistä sattumanvaraisuuksista. Autonomisen toiminnan täytyy olla yleistettävää eikä se voi riippua toimijan sattumanvaraisista ominaisuuksistaan. Auktoriteetti voi juontaa vain järjen edellyttämistä tekijöistä. O'Neill toteaa, ettei Kant tekstissään kerro ”mitkä järjen periaatteet ovat ja miksi niillä on auktoriteettia²³⁴.”²³⁵

²³⁰ “...a kind of causality belonging to living beings so far as they are rational.” O'Neill 1989, 52, käänös oma.

²³¹ “Free will (if it occurs) must be (capable of being) positively as well as negatively free; it must be a capacity of *self-determination* or *autonomy*.” O'Neill 1989, 53, käänös oma. Sulkeet ja kursiivit kuten alkuperäisessä tekstissä.

²³² O'Neill 1989, 52–53.

²³³ “...if rational beings are free, they are not simply negatively free, but (capable of) autonomy, and that if (capable of) autonomy, they are subject to the principle of morality, that is, to the supreme principle of practical reason.” O'Neill 1989, 53, käänös oma. Sulkeet ja kursiivit kuten alkuperäisessä tekstissä.

²³⁴ “[we do not learn] ... what the principle(s) of reason are or why they have authority.” O'Neill 1989, 54, käänös oma. Sulkeet ja kursiivit kuten alkuperäisessä tekstissä.

²³⁵ O'Neill 1989, 54.

Ihmisen järkevyydestä ja vapaudesta voidaan johtaa vain ehdollinen johtopäätös:

*Jos olemme järkeviä edellytetyllä tavalla, olemme myös vapaita, siten kykeneviä autonomiaa, ja moraalिसääntöjen alaisia. Jos taas emme ole järkeviä edellytetyllä tavalla, vaan jollakin muulla (kuten vain instrumentaalisella) tavalla, silloin järkevyytemme ja vapautemme väliin jää railo, ja kantilainen käsitys autonomiasta on meille merkityksetön.*²³⁶

O'Neillin mukaan Kant on hyvin tietoinen tällaisen railon olemassaolosta ja näkee, että vapaus on järkevien olentojen tahdon ominaisuus. Kant ei voikaan todistaa, vaan vain väittää, että jokaiselle vapaan tahdon omaavalle järkevälle olennoille pitää myöntää vapauden periaate ainoaksi, jonka mukaan tämä voi toimia. Jos on mahdollista toimia vain vapauden periaatteen mukaisesti, toimijan on käsitettävä itsensä vapaaksi riippumatta siitä, voidaanko vapaus osoittaa. O'Neill kuitenkin muistuttaa myös Kantin itsensä muistuttaneen, että oletetut ihmisen kokemukset eivät ole mikään tae oikeasta vapaudesta, eikä ihmisillä niin ikään ole minkäänlaista oivallusta, miksi moraalilait todellisesti olisivat sitovia.²³⁷

O'Neill etenee ratkomaan yllämainittua ongelmaa Kantin käytännöllisen järjen kritiikin avulla. Keino murtautua ulos päättelyn kehästä, jonka mahdollinen illuusio vapaudesta aiheuttaa, edellyttää uutta näkökulmaa ongelmaan. Puhtaan järjen kritiikin kantavana ajatuksena on löytää ajattelun soveltuvia mittapuita pohtimalla, miten voimme ja miten meidän täytyy säädellä ajatteluamme. Tämä säätely ei kuitenkaan ole ulkopuolelta tulevaa, vaan se on ajattelun itsesäätelyä tai autonomiaa. O'Neill kertoo Kantin todenneen: ”Tuntuu oudolta, että järki, jonka tehtävä on säädellä kaikkia pyrkimyksiä, itse tarvitsisi tällaista säätelyä²³⁸.” Joskin O'Neillin mukaan, juuri tämä tarve on avain Kantin kuvaukseen järjen auktoriteetista. Ajattelu tarvitsee ”...säätelyä, jotta voitaisiin hillitä sen taipumus pyrkiä laajentumaan mahdollisen kokemuksen kapea-alaisuuden ulkopuolelle, ja näin suojella sitä kohtuuttomuudelta ja virheilä²³⁹.”²⁴⁰

O'Neill toteaa, että Kantin mukaan järkeä voi vain, ja täytyy, itsesäädellä. Järjen säateleminen ei voi olla ulkopuolista, vaan sen täytyy olla autonomista. Näin,

²³⁶ “If we are rational in the required sense, then we are also free and so capable of autonomy and bound by morality. If we are not rational in the required sense, but only in some other (e.g., purely instrumental) sense, then there will be a gap between our rationality and our freedom, and the Kantian conception of freedom will be irrelevant to us.” O'Neill 1989, 54. Käännös oma. Kursiivit kuten alkuperäisessä tekstissä.

²³⁷ O'Neill 1989, 54–55.

²³⁸ “...that reason, whose proper duty it is to prescribe a discipline for all other endeavours, should itself stand in need of such a discipline may indeed seem strange.” O'Neill 1989, 57, käännös oma.

²³⁹ “...a discipline, to restrain its tendency towards extension beyond the narrow limits of possible experience and to guard it against extravagance and error.” O'Neill 1989, 57, käännös oma.

²⁴⁰ O'Neill 1989, 55–57.

O’Neill toteaa, railo järjen ja autonomian välillä voidaan sulkea. Tämä ei tapahdu tunnustamalla ihmisolenon olevan järkevä, ja siten vapaa ja autonominen, vaan tunnustamalla vain autonomisten olennot voivan toimia järjen periaatteiden (principles of reason) mukaan. O’Neillin mukaan järjellä itsellään ei ole yliluonnollista auktoriteettia. Järki voidaan todistaa oikeaksi vain kritiikin avulla ja kritiikki itsessään on ajattelun autonomiaa. Autonomia ei näin edellytä järkeä, vaan järki edellyttää autonomiaa.²⁴¹

Näin siis olemme päätyneet ajatukseen, ettei puhdas järki takaa autonomiaa, vaan autonomia on edellytyksenä järkevälle toiminnalle, sillä ilman autonomiaa toimija ei pysty kritisoimaan (järkeään) ajatteluaan. Järki tai rationaalisuus on siis vapauden ehto, mutta järkeä on kyettävä säätelemään autonomisesti.

Käsittelen seuraavassa kirkkoisä Augustinuksen ja Descartesin näkemyksiä vapaasta tahdosta. Christopher Gilbert kirjoittaa Augustinuksen erottavan kaksi erilaista vapautta, vapaan valinnan (free choice, *liberum arbitrium*), joka on kaikkien ihmisten ominaisuus syntiinlankeemuksen jälkeen, ja vapauden (*liberty, libertas*) joka vallitsee taivaassa. Gilbert kuitenkin näkee Augustinuksen vapauskäsitteessä nyansseja, joita Augustinus ei kuitenkaan suoraviivaisesti esittele kirjoituksissaan. Augustinus toteaa *De libero arbitrio* -teoksessaan ihmisen omaavan tarpeellisen vapauden tehdä pahoja tekoja. Myöhemmissä teoksissa painotetaan ihmisen kykenemättömyyttä tehdä hyvää ilman Jumalan armoa. Gilbertin mukaan, yhdistettäessä Augustinuksen ajatukset vapaata valintaa koskien, voidaan johtaa teoria neljästä erillisestä vapauden asteesta: (1) ihmisille luomisessa annettu vapaus, (2) lankeemuksen jälkeinen niiden ihmisten vapaus, jotka eivät toimi yhdessä armon kanssa, (3) lankeemuksen jälkeinen niiden ihmisten vapaus, jotka toimivat yhdessä armon kanssa ja (4) enkelien ja siunattujen vapaus taivaassa. Koska Augustinus Gilbertin mukaan näkee tahdon mielen kykynä itsehallintaan, mikään muu ei voi järkyttää mielen hallintaa paitsi tahto. Näin ollen, tehdessään syntiä, ihminen ”tekee näin tahtonsa vapaasta valinnasta²⁴²”, synnin ollessa ”muuttuvien ja epävarmojen asioiden tavoittelua... ...ikuisesti pysyvien jumalallisten asioiden” sijaan²⁴³.²⁴⁴

²⁴¹ O’Neill 1989, 55–57.

²⁴² “through the free choice of our will”, Gilbert 2005, 203, käänös oma.

²⁴³ “‘things which are changing and insecure’ instead of ‘godly things which are truly lasting’”, Gilbert 2005, 203, käänös oma.

²⁴⁴ Gilbert 2005, 202–203.

Gilbert toteaa ihmisillä olevan näin Augustinuksen mukaan rajallinen vapaa tahto: he voivat vapaasti tahtoa tehdä syntiä. Siinä kaikki. Jopa synninteosta pidättäytyminen vaatii armon väliintuloa. Taivaassa vallitseva täydellinen vapaa tahto on tähän verrattuna toinen ääripää. Augustinuksen mukaan todellinen vapaus ei perustu vain kykyyn tehdä hyvää, vaan kykenemättömyyteen tehdä syntiä. Gilbert siteeraa Augustinusta *De correptione e gratia* -teoksesta: ”Mikä olisi vapaampaa kuin vapaa tahto, kun sillä ei ole voimaa palvella syntiä²⁴⁵?” Gilbert toteaa ajatuksen- ja toiminnanvapauden kannattajien kieltävän jyrkästi tällaisen vapauden käsitteen, mutta huomauttaa, että näin tehdessään he ohittavat keskeisen aatteen Augustinuksen vapauskäsitteessä. Käsitteeseen vapauden määrän kasvusta mahdollisuuden lankeamiseen pienetessä versoi ajatuksesta, että vapaus tehdä väärin ei ollut olennaista vapaudessa. Synti nähtiin Jumalan antaman vapauden väärinkäytöksenä ja näin vapautta vähentävänä.²⁴⁶

Augustinus siis erotti ihmisen vapaan valinnan taivaassa vallitsevasta täydellisestä vapaudesta. Tämän lisäksi hänen kirjoituksistaan on Gilbertin mukaan mahdollista erottaa myös muunlainen vapauden tila; vapaus, joka ihmisellä oli, kun hänet luotiin. Gilbert kutsuu tätä termillä ”alkuperäinen vapaus” (original freedom). Tämä ei ollut Augustinuksen mukaan täydellistä vapautta, mutta ihmisellä oli tuolloin kyky tahtoa tehdä hyvää, samoin kun hänellä oli kyky tahtoa tehdä paha. Tämä vapauden tila jää ikään kuin taivaassa vallitsevan täydellisen vapauden ja langenneen ihmisen vapaan tahdon väliin. Ensimmäisellä ihmisellä olisi Augustinuksen mukaan ollut kyky valita oikein. Syntiinlankeemus kuitenkin tapahtui samalla turmellen ihmisen vapaan tahdon siten, ettei se enää omin voimin pystynyt muuhun kuin syntiin. Jotta langennut ihminen pystyisi tahtomaan ja toteuttamaan hyviä tekoja, tarvitaan armon väliintuloa. Gilbertin mukaan Augustinus kuitenkin kieltää, että tämä armon väliintulo ja vaikutus ihmisessä mitenkään tekisi mahdottomaksi ihmisen vapautta. Armo näet ei pakota ihmisen tahtoa vaan tekee yhteistyötä sen kanssa, näin auttaen tahtoa toimimaan.²⁴⁷

Gilbert näkee laadullisen eron armon hyväksyneen langenneen ihmisen ja armon kieltävän langenneen ihmisen tahdon vapaudessa. Armosta kieltänyt ihminen on ”synnin orja” (slave to sin). Hän on vapaa valitsemaan synnin, mutta ei

²⁴⁵ “What will be more free than the free will, when it shall have no power to serve sin?” Gilbert 2005, 204, käännös oma.

²⁴⁶ Gilbert 2005, 204–205.

²⁴⁷ Gilbert 2005, 205–207.

mitään muuta. Tämän hän tulkitsee Augustinuksen näkemyksissä vapauden alhaisimmaksi asteeksi. Armon hyväksynyt ihminen on hieman vapaampi. Hän kykenee yhteistyössä armon kanssa valitsemaan ja toimimaan oikein, joskin hänellä on yhä kyky milloin vain kieltää armo. Vaikka tämä vapauden aste on toiminnallisesti samanlainen alkuperäisen vapauden kanssa, näiden välillä on ero. Ensimmäinen ihminen, Aatami, oli luonnostaan kykenevä valitsemaan oikein, kun taas langennut ihminen ei voi toimia näin hyväksymättä yhteistyötä armon kanssa. Korkein vapauden aste vallitsee taivaassa, jossa siunatut pystyvät valitsemaan vain oikein.²⁴⁸

Myös Descartes pohtii vapaata valintaa ja väärin toimimista. Miksi ihminen toimii väärin hänellä kuitenkin ollessa Jumalan suoma järki? Vastuksena tähän Descartesilla on tahto. Descartes erottaa havaitsemisen kyseisen havainnon arvioinnista, siis todentamisesta ja kieltämisestä. Hän toteaa havaitsemisen olevan järjen passiivista toimintaa siinä missä arviot aktiivista tahdonalaista toimintaa. Havainnot eivät itsessään voi olla vääriä, jos jokin vaikuttaa näyttävän joltakin, se todennäköisesti vaikuttaa tältä. Gilbertin mukaan Descartes totesi silti ihmisen voivan erehtyä arvioinnissaan tätä havaintoa koskien. Virheitä tapahtuu, kun ihminen arvostelee asioita, joista hän ei ole varma. Gilbert erottelee kolme pääkohtaa:

- (1) Syyn niin älyllisiin kuin moraalisiin virheisiin on löytynyt.
- (2) Syy näihin on vapaan tahdon väärinkäyttö
- (3) Kykymme tehdä väärin ja syntiä linkittyy välinpitämättömyyden tuomaan vapauteen (freedom of indifference).

Gilbert toteaa Descartesin määritelleen välinpitämättömyyden ”tahdon tilana, jolloin tahtoa ei liikuta mikään havainto totuudesta tai hyvyydestä²⁴⁹”. Toisaalta Descartes huomaa, että välinpitämättömyyttä voi ilmetä myös silloin, kun syyt valita eri tavoin ovat olemassa. Tällöin välinpitämättömyys juontaa kyvystä valita yksi vaihtoehto, hyvistä syistä valita toisin huolimatta.²⁵⁰

Sälyttäessään välinpitämättömyyden ihmisen tahdolle Descartes näin osoittaa väärintekemisen olevan ihmisen oma puute, ei jumalallisen järjen vika. Jos tarpeellista tietoa arviointiin ei ole, ainoa oikea toimintatapa Descartes’in mukaan on olla valitsematta ja näin välttää virhe. Älylliset ja moraaliset virheet ovat tulosta vapaan tahdon väärinkäytöstä. Gilbert toteaa välinpitämättömyyden olevan

²⁴⁸ Gilbert 2005, 207–209.

²⁴⁹ “...that state of the will when it is not impelled one way rather than another by any perception of truth or goodness.” Gilbert 2005, 211, käännös oma.

²⁵⁰ Gilbert 2005, 209–211.

Descartesin näkemyksissä vapauden alin aste ja näkee tässä samankaltaisuutta Augustinuksen alimman vapauden asteen kanssa. Kummankin tuloksena on vain kyky toimia väärin. Siinä missä virheitä tapahtuu asioita ilman parempaa tietoa tai välinpitämättömästi arvostellessa, virheet voi Descartesin mukaan välttää jos hyväksytään vain selvästi havaitut asiat. Tämä on välttämättömyyden tuomaa vapautta (freedom of spontaneity). Gilbert toteaaakin Descartesin usein määrittelevän välttämättömyyden tuoman vapauden ihmisen kykenemättömyytenä valitsemaan toisin, kuin minkä hän selkeästi havaitsee oikeaksi ja hyväksi. Vapaa tahto on sitä enemmän välttämättömyyden sanelema, ja siten vapaampi, mitä enemmän siihen vaikuttavat selkeät, varmat (järjen) havainnot. Gilbert näkeekin järjen edustavan Descartesin ajattelussa samaa kuin armo Augustinuksen ajattelussa.²⁵¹ Toimija on sitä vapaampi, mitä enemmän armo/järki tekee yhteistyötä tahdon kanssa. Vapainta toiminta on silloin, kun Descartes'in mukaan oikein järjen ehdoilla toimiminen on välttämätöntä tai Augustinuksen mukaan väärin tekeminen ei enää ole edes mahdollista, koska todellinen vapaa tahto tahtoo vain hyvää.

3.4.2 Täydellinen vapaus

Yksi kysymyksistäni tässä tutkimuksessa on ollut, voiko vapaa tahto ilmetä kehollisessa keinoälyssä. Voitaneen arvella vapaan tahdon linkittyvän verrattain vahvasti kokemuserustaiseen tietoisuuteen. Mutta mitä on vapaus? Onko tahto sitä vapaampi, mitä enemmän siihen vaikuttaa – Descartes'in mukaan – järki tai armo, kuten Augustinus kertoo. Tulisiko keholliseen keinoälyyn implementoida ihmisen tahdon vapaus vai ”enkeliä ja siunattujen vapaus taivaassa”

Millaisina enkelit on ymmärretty? Reijo Työrinoja on muun muassa kirjoittanut katsauksessaan ”Enkelit keskiajan taivaalla, Enkelioppiin liittyviä filosofisia kysymyksiä” Jumalan jo enkelit luodessaan ohjelmoineen enkeleihin kaiken tarvittavan tiedon, todellisuuden lajimuodot ja niiden suhteet. Enkelit onkin tietonsa alueella kaikkietävä, enkeli on virheetön ajattelija (pure intellect)²⁵², eikä enkelille ole tarpeen tukeutua aistitietoon. Työrinoja kertookin joidenkin kirjoittajien kommentoineen enkelien tiedon ohjelmoituna muistuttavan keinoälyä.²⁵³

²⁵¹ Gilbert 2005, 201–213, 218.

²⁵² Lenz 2008, 153.

²⁵³ Työrinoja 2004, 2721–2722.

Työrinoja kirjoittaa 1200-luvun olleen enkeliopin kukoistusaikaa. Enkelioppi, angelologia, oli keskeisessä asemassa niin dominikaani- kuin fransiskaaneljeskunnassakin. Sekä dominikaaneja edustanut Tuomas Akvinolainen että fransiskaani Bonaventura opettivat tuolloin samaan aikaan Pariisin yliopistossa. Työrinoja toteaa enkeliopin olleen erityisesti fransiskaaneille sekä teologisesti että ideologisesti merkittävä. Goff, Hellman ja Hammond kirjoittavat Bonaventuran – joka Pariisin yliopistossa tunnettiin myös arvonimellä serafitohtori (doctor seraphicus)²⁵⁴ – todenneen opissaan enkelien tiedosta ja persoonallisuudesta

Meidän on tiedettävä, että heti luomisensa hetkellä enkeleille suotiin täydellisyys neljällä tapaa: olemuksen vaatimattomuus, luonteen yksilöllisyys, järkevyyt (. . .) ja vapaus valita (. . .). Nämä neljä pääasiallisen ominaisuuden lisäksi on neljä muuta: toiminnan mestarillisuus, palveluksen antaumuksellisuus, ymmärryksen terävyys ja muuttumattomuus hyvän ja pahan valinnassa.²⁵⁵

Enkelit luotiin siis järkeviksi ja niillä oli vapaus valita. Mikä on huomionarvoista, on se, että enkelit valitsivat vain kerran ja enkelien valinta oli hyvän ja pahan välillä. Tämä ensimmäinen vapaa valinta määritteli enkelin luonteen ikuisesti, eikä tätä valintaa voinut muuttaa. Goff, Hellman ja Hammond toteavatkin, ettei enkeliä voi pelastaa. Pahan valinneet enkelit putosivat taivaasta. Jumalan armo täydellisti hyvän valinneiden enkelien vapaan tahdon ja ne siirtyivät synnittömästä olemuksesta täydellisyyteen.²⁵⁶

Gilbert kirjoittaa, että Enkelit nauttivat taivaassa Augustinuksen mukaan täydellisestä vapaudesta, *libertas*. Augustinus vakuuttaa, että enkeliolennolla on ”ikuinen onnellisuus nauttiessaan Luojansa läsnäolosta ikuisesti, ansaiten tämän ainaisella tahdolla ylläpitää oikeutta²⁵⁷”. Mutta kuten jo aiemmin kirjoitin, Augustinukseksi mukaan täydellinen vapaus ei kuitenkaan ole vain kyvyssä tehdä hyvää, vaan kykenemättömyydessä tehdä syntiä. Kuten Gilbert toteaa, *libertas*

²⁵⁴ Työrinoja 2004, 2720. Myös Tuomas akvinolaisella oli Pariisin yliopistossa arvonimi, enkeliopettaja (doctor angelicus).

²⁵⁵ “We must know that, at the very instant of their creation, the angels were endowed with four perfections: simplicity of essence; individuality of person; rationality (. . .) and freedom of choice (. . .). These four main attributes are accompanied by four others: virtuosity in action, dedication in service, acuteness in understanding, and immutability in the choice of either good or evil.” Goff, Hellmann & Hammond 2013, 312, käännös oma. Sulkeet ja pisteet kuten alkuperäisessä tekstissä.

²⁵⁶ Goff, Hellmann & Hammond 2013, 306, Työrinoja 2004, 2720.

²⁵⁷ “...has everlasting happiness, enjoying forever its Creator, and deserving this by its constant will to uphold justice.” Gilbert 2005, 204, käännös oma.

tarjoaa vain ohjaavan hallinnan (guidance control), kyvyn hallita tekojaan. *Libertas* ei tarjoa säätelevää hallintaa (regulative control), kykyä toimia toisin, kuin toimii.²⁵⁸

Gilbertin mukaan Aatamilla oli nämä molemmat kyvyt. Ensimmäinen ihminen, joka sieraimiin Jumala puhalsi elämän henkäyksen, oli kykenevä sekä tekemään syntiä että tekemään hyvää. Aatamilla oli ohjauksen hallinta samoin kuin täysi säätelevä hallintakin.²⁵⁹ Mutta Steven B. Smith kirjoittaa kirjassaan *Political Philosophy*, että ihminen luotiin kuitenkin ilman tietoa hyvästä ja pahasta eikä Jumala tahtonut tätä tietoa ihmiselle paljastaakaan. Jumala kieltää Aatamia syömästä hyvän ja pahaan tiedon puusta. Aatami ja Eeva kuitenkin syövät tullakseen Jumalan kaltaisiksi, ja samalla oppivat hyvän ja pahan eron. Smith kirjoittaa, että he omaksuvat moraalin ja samalla he tulevat ihmisiksi.²⁶⁰

Myös Immanuel Kant erottaa pyhän tahdon (holy will) ihmisten tahdosta. Robert Stern kirjoittaa artikkelissaan ”Kant, Moral Obligation and The Holy Will” moraalin muodostuvat ihmisille Kantin mukaan imperatiivina, koska ihmisillä on taipumus toimia väärin syntiinlankeemuksen vuoksi. Pyhälle tahdolle tällaista imperatiivisia ei ole, koska pyhä tahto toimii aina oikein.²⁶¹ Kant erotteleekin myöhemmässä tuotannossaan käsitteet vapaa valinta (Willkür) ja vapaa valinta järjen sanelemana (Wille). Kant näkee näistä vain jälkimmäisen moraalisenä ja autonomisena.²⁶²

Onko siis ihmisten tahto vapaa, kun on mahdollista valita myös vastoin moraalista vaatimusta vai onko ihmisten tahto vapaa vasta, kun se pystyy valitsemaan hyvän omasta halustaan? Olihan myös kirkkoisä Augustinuksen näkemys lopulta se, että Jumalan tahto ei ollut ristiriidassa ihmisen vapaan tahdon kanssa.²⁶³ Jos pohdimme hetken buddhalaista näkemystä vapaasta tahdosta, tahdon nähdään olevan vapaa, kun se on vapaa *mieltä haittaavista tekijöistä*²⁶⁴. Risto Saarinen toteaa, että ihmisen kykyä toimia oikein häiritsevät Augustinuksen kirjoituksissa kaksi epätäydellistä tahtoa. Juurtuneet tavat estävät kohti oikeaa pyrkimisen. Jumalan armoon heittäytyminen parantaa ihmisen tahdon, jolloin ihminen *vapautuu haitallisista, häiritsevistä mielihaluista*.²⁶⁵

²⁵⁸ Gilbert 2005, 203–204.

²⁵⁹ Gilbert 2005, 206.

²⁶⁰ Smith 2012, 93.

²⁶¹ Stern 2013, 126–127.

²⁶² O’Neill 1989, 67.

²⁶³ Saarinen 2011, 21–22.

²⁶⁴ Federman 2010, 9–10.

²⁶⁵ Saarinen 2011, 22.

Läntisen vapaan tahdon käsitteen keksijän näkemys vapaudesta ei siis lopulta vaikuttaisi eroavan määritelmällisesti buddhalaisesta vapaan tahdon käsitteestä, vaikka näkemykset tämän vapauden saavuttamisesta eroavatkin toisistaan. Tämä on siis se tahdon vapaus, johon ihmisen tahrautuneine mielineen pitäisi pyrkiä. Kysymys on, jos emme itse osaa tai kykene saavuttamaan tahdon vapautta, pystymmekö implementoimaan sitä koneeseen?

Johtopäätökset

Tämän tutkimuksen tarkoitus oli selvittää miten tietoisuus on ymmärretty theravāda-buddhalaisuudessa ja millaisia taustaoletuksia tähän tietoisuuskäsitykseen liittyy erityisesti suhteessa moraaliin ja vapaaseen tahtoon, ja voivatko tähän tapaan ymmärretyt tietoisuus, moraali ja vapaa tahto ilmetä kehollisessa keinoälyssä. Tutkimusmetodini oli systemaattinen analyysi.

Kuten johdannossa totesin, vaikka theravāda-buddhalaisen opin ymmärrystä tietoisuudesta ja sen suhdetta moraaliin ja tahdonvapauteen on tutkittu, tutkimukseni toisesta osa-alueesta – voivatko tietoisuus, moraali ja vapaa tahto, siten kun ne ymmärretään theravāda-buddhalaisuudessa ilmetä kehollisessa keinoälyssä – ei aiheeni valinnan inspiraationa toimineen David Hughes'in artikkelin "Compassionate AI and Selfless Robots" lisäksi ole kirjoitettu edeltävää tutkimusta. Aihe herättää monenlaisia kysymyksiä. Rajasin omat tutkimuskysymykseni seuraavasti: Mitkä ovat tietoisuuden, yksilön, ilmenemisen ehdot theravāda-buddhalaisen opin mukaan ja mikä on tietoisuuden ja moraalisen toimijuuden ja vapaan tahdon suhde? Miten tietoisuuden käsittä tulkitaan niin tieteessä kuin theravāda-buddhalaisen perinteen näkökulmasta? Voivatko tietoisuus, moraali ja vapaa tahto ilmetä kehollisessa keinoälyssä viimeaikainen tutkimuksen valossa?

Rajasin buddhalaisuuden laajana ajatusmaailmana theravāda-buddhalaisuuteen, läntisellä tieteellä viittaaan kognitiiviseen neurotieteeseen ja tämän paradigmaan, että tietoisuus on aivojen tuottamaa. Kehollisella keinoälyllä viittaaan keinoälyyn, joka on fyysisen rakenteensa perusteella selkeästi ihmisen rakentama kone. Tietoisuudesta, moraalisesta toiminnasta ja tahdonvapaudesta on useita määritelmiä, eikä tarkoitukseni ole kyseenalaistaa tässä tutkimuksessa käsityksiä niiden moninaisista olemuksista. Tässä tutkimuksessa käsittelen tietoisuuden käsitettä fenomenalisena, subjektiivisena läsnäolon tunteena,

moraalin käsitettä moraaliseen tietoon pohjautuvana tekemisenä ja tahdonvapauden käsitettä ihmisen vapautena toimia myös vastoin sitä, mikä on oikein.

Hahmotin kokonaiskuvaa aiheestani tutustumalla theravāda-buddhalaisen opin perusteisiin yksilön tietoisuudesta, moraalista ja tahdonvapaudesta. Lähteitáni olivat Rupert Gethinin *The Foundations of Buddhism*, Steven Collinsin *Selfless persons. Imagery and thought in Theravāda Buddhism*, Damien Keown'n *The Nature of Buddhist Ethics* sekä artikkelit "What Kind of Free Will did the Buddha Teach?", "A Buddhist View of Free Will: Beyond Determinism and Indeterminism" ja "Buddhism and the Freedom of Will: Pali and Mahayanist Responses.

Teorioita tietoisuuden luonteesta etsin Ted Honderichin teoksesta *Actual Consciousness*. Lähteenäni käytin myös B. Alan Wallacen *Contemplative Science: Where Buddhism and Neuroscience Converge*. Keinotekoisen moraalien kehitystä viimeaikaisen tutkimuksen valossa olen tutkinut lähteenäni Wendell Wallachin ja Collin Allenin yhteistyössä kirjoittamaan *Moral Machines: Teaching Robots Right from Wrong* –teoksen valossa. Tahdonvapautta käsitellessäni käytin lähteenä Onora O'Neillin *Constructions of Reason, Explorations of Kant's Practical Philosophy*. Christopher Gilbertin artikkelista "Grades of Freedom". Selvitti minulle tahdonvapautta kirkkoisä Augustinuksen ja René Descartesin aatteiden valossa. Lähteenäni on ollut myös buddhalaisuudesta, neurotieteestä sekä tietoisuudesta ja vapaasta tahdosta kirjoitettuja tutkimusartikkeleita.

Kysyin tutkielmani ensimmäisessä pääluvussa, mitkä ovat tietoisuuden, yksilön, ilmenemisen ehdot theravāda-buddhalaisen opin mukaan ja mikä on tietoisuuden ja moraalisen toimijuuden ja vapaan tahdon suhde? Buddhalaisen opin mukaan tietoisuus ilmenee toimijassa yhtenä toimijan viidestä osatekijästä. Muut neljä ovat siis (1) muoto, (2) ajatukset ja tunteet, (3) käsitteen muodostus sekä (4) halut ja toiveet. Buddhalaisuuden keskeinen kulmakivi on oppi ei-itsestä (anātman/anattā). Oppi ei siis tunnusta pysyvää itseä, muuttumatonta sielua. Toimijalla voi olla käsitys itsestä. Käsitys itsestä nähdään jopa merkityksellisenä, sillä vailla kahletta ei olisi mitään, mistä vapautua. Itseys on vain opillisesti väärin.

Tietoisuuden keskeistä roolia ajattelussa alleviivaa se, että (Theravāda-) buddhalaisuuden lainatessa bramiiniuskosta käsitteet uudelleensyntymisestä, tekojen vaikutuksesta kohtaloon, ja vapautumisesta, tekoja merkityksellisemmäksi kohosi aikomus tekojen takana. Seuraavaa olomotoa ei määritellyt teko, vaan

aikomus. Valaistuminen saavutettiin tietoisesti suoritetun itsen havainnoinnin avulla.

Asiat ja tapahtumat koetaan theravāda-buddhalaisen opin mukaan kausaaliseksi. Paitsi yksi yksilön osatekijöistä, tietoisuus on myös yksi ehdonvaraisen seuraamuksen kahdestatoista lenkistä. Täten tietoisuus johtaa kausaalisesti asioihin ja tekoihin, joko moraalisesti oikeisiin tai väärin. Kysymys tahdonvapaudesta ei ollut keskeinen ongelma buddhalaisuudessa. Toimijalla on olosuhteiden niin salliessa mahdollisuus, vapaus, valita oikea teko näin välttää ehdonvaraisen seuraamuksen epäsuotuisasti valaistumiseen vaikuttavat tekijät. Oikean teon valitessaan toimija seuraa Jaloa Kahdeksanosaista Polkua kohti valaistumista. Samalla toimija vähentää omaa ja toisten kärsimystä. Näin toimija harjoittaa Buddhan opetuksen mukaista myötätuntoa, jonka mukaan kaikkiin olentoihin tulee suhtautua rakastavalla ystävällisyydellä, myötätunnolla, sympaattisella ilolla ja mielentyyneydellä. Yllämainittuun perustuen analyysini on tukenut olettaa, että theravāda-buddhalaisen opin ymmärtämä tietoisuus on edellytys vapaalle tahdolle ja siten moraaliselle toiminnalle.

Tutkielmani toisessa pääluvussa kysyin, miten tietoisuuden käsittä tulkitaan niin tieteessä kuin theravāda-buddhalaisen perinteen näkökulmasta. Vaikka (ihmisten) tietoisuuden luonnetta ei ole tieteellisesti määritelty, voitaneen todeta edelleen Pentti Haikosta mukailleen tietoisuuden olevan ”immateriaalinen läsnäolon tunne²⁶⁶”. Tietoisuuteen linkittyvät tietoisuuden teorioiden mukaan (1) joltakin vaikuttavuus, qualia, ja (2) joltakin tuntuminen. Edelleen tietoisuuden voidaan olettaa olevan (3) subjektiivinen kokemus johon (4) sisältyy erilaisia intentioita. Nämä kaikki neljä sisältyvät myös teoriaan tietoisuudesta fenomenalisena ilmiönä (5). Lähteittäni mukaan vallitsevin teoria tietoisuudesta on viimeaikaisessa tutkimuksessa ollut funktionalistinen teoria. Tämä teoria selittäisi, mistä mielen sisäiset, eivät vain tietoiset, tosiasiat rakentuvat; miten mieli toimii. Tutkimukseni valossa tiede vaikuttaisi siis tutkivan tietoisuuden ilmiötä joko vahvasti tietojenkäsittelynä, neurotieteenä tai biologisena ilmiönä. Näin lienee perusteltua todeta, että tietoisuus on viimeaikaisessa tutkimuksen valossa aivojen toimintaan linkittynyt, toiminnallinen ilmiö. Tosin, tiede ei ole vielä pystynyt selittämään, miten aivot synnyttävät tietoisuuden. Koska todisteet puuttuvat, tätä tieteen paradigmaa tietoisuudesta aivojen tuottamana on myös haastettu. Haastava teoria

²⁶⁶ Haikonen 2007, 2.

on, että tietoisuus on jotain, jota aivot välittävät ja johon kokemukset vaikuttavat. Kysymys on, jos tietoisuus on jotain, jota aivot välittävät, miten tätä voidaan mitata? Introspektion pätevydestä toimijan tapana tarkastella omaa tietoisuuttaan, vallitsee ilmeinen ristiriita.

Wallacen esittelemä kognitiivisen neurotieteen paradigman haastava näkökulma tietoisuuteen mukailee buddhalaisen ajatusmaailman näkökulmaa. Buddhalaisen opin mukaan tietoisuus on toimijasta erillinen toiminnallinen ilmiö, joka yhtenä viidestä osatekijästä varmistaa toimijan olemisen, mutta myös jatkuvuuden läpi elämän kierron. Tietoisuus ei synny toimijassa samalla kuin toimija (vrt. yksilö) syntyy eikä lopu toimijan menehtyessä. Tietoisuuden virta jatkaa kulkuaan niin kauan, kun se saa ”ravintoa”, jota ovat fyysinen ravinto, aistivaikutelmat, henkinen tahtominen ja tietoisuus. Mielihalu rakentaa edelleen uutta tietoisuutta ja johtaa tietoisuuden uudelleen syntymiseen. On huomionarvoista, että buddhalainen näkemys tosin erottaa tietoisesta tilan (citta-vīthi) joka ei jatku. Tiedostamaton tila (vīthi-mutta tai bhavaṅga) jälleen on luonteeltaan jatkuva. On kuitenkin pidettävä mielessä, että vaikka tämä tiedostamaton tila jatkuu, se ei ole pysyvä ja ikuinen, vaan sen ominaisuuteen kuuluu olla hetkellinen, ja syntyä, olla ja kuihtua pois.

Tutkielmani kolmannessa pääluvussa kysyin, voivatko tietoisuus, moraalit ja vapaa tahto ilmetä kehollisessa keinoälyssä viimeaikainen tutkimuksen valossa. Tavoitteesta tekee haasteellisen tietoisuuden ilmiön tuntemattomuus. Koneiden mahdollinen tietoisuus lienee vuosien päässä. Edelleen on myös hyvä pohtia millaista tämä tietoisuus olisi. Niin kauan, kuin tietoisuuden olemusta ei ymmärretä, kehollisen keinoälyn tietoisuus voinee parhaimmillaankin olla toiminnallista. Robotit voitaneen saada toimimaan ikään kuin ne olisivat tietoisia. Toisin sanoen, ei ole syytä, miksi ihmisen rakentamaan koneeseen ei voisi sisällyttää kaikkia tunnettuja tietoisuuden mekanismeja. Tietoisuuden sisältö asettaa haasteen, koska tätä ei tunneta. Entä mitä voidaan sanoa robottien tietoisuudesta buddhalaisen opin valossa? Tahdon tähän liittyen jälleen viitata Dalai Laman lauseeseen:

Jos tietokoneen fyysinen olomuoto saavuttaa valmiuden toimia olomuotona tietoisuudelle, tietoisuuden virta ehkä todella saattaisi asettua koneeseen..²⁶⁷

²⁶⁷ “If the physical basis of the computer acquires the potential to serve as a basis for a continuum of consciousness ... a stream of consciousness might actually enter into a computer” Hughes 2012, 69, käännös oma.

Mikä on edelleen huomionarvoista, on se, että buddhalaisen opin mukaan tietoisuus tekee kodin neljään muuhun osatekijään, skandhaan, ja näin tavallaan ankkuroituu näihin.²⁶⁸ Pohtikaamme, nämä neljä muuta osatekijää ovat (1) muoto, vaikutelma kehosta ja aisteista, ja ympäröivästä maailmasta, (2) fyysisen maailman vaikutuksesta nousevat ajatukset ja tunteet, (3) kyky luokitella asiat, käsitteen muodostus sekä (4) halut ja toiveet.

Vaikuttaa siltä, että niin kauan, kuin kehollinen keinoäly on fyysisen rakenteensa perusteella selkeästi kone, ei organismi, on verrattain haastavaa rakentaa tästä koneesta sellaista, että sen fyysinen olomuoto mahdollistaisi tietoisuuden virran asettumisen. Viittaa edelleen James Hughesin ajatukseen, että buddhalainen ajatusmaailma olisi verrattain avoin robotissa ilmenevälle tietoisuudelle, kirjaimellisesti ”buddhalaiset ovat avoimempia ajatukselle tietoisuuden ilmenemisestä koneissa.²⁶⁹”. Analyysini valossa minun on todettava, että vaikka tämä ajatus ei liene kirjaimellisesti epätosi, koneen tietoisuudelle soveltuvaa olomotoa voi olla haasteellista toteuttaa. Voinemme epäillä, voiko ihmisen rakentama kone olla olomuoto tietoisuudelle niin kauan, kun emme tiedä, mitä tietoisuus on. On kuitenkin huomioitava, että mikäli kehollisen keinoälyn käsitettä laajennetaan siten, että se käsittää näennäisbiologiset entiteetit, myös kysymystä tietoisuuden ilmenemisestä on tarkasteltava uudelleen.

Voiko robotti olla moraalinen toimija? Tähän kysymykseen voidaan tarjota erilaisia vastuksia riippuen siitä, miten moraalinen toimija määritellään ja miten kriteerit moraaliselle toiminnalle nähdään. Mikäli ikään-kuin, toiminnallinen moraalinen toiminta riittää, koneen rajattu moraalinen toimintatapa lienee varsin saavutettavissa. Ei liene lainkaan mahdotonta, etteikö robotista voitaisi rakentaa James Moorin kategorioiden mukaista selkeää moraalista toimijaa. Mikä on olennaista, on se, että vaikka robotti ei kykenisikään olemaan eettinen päätöksentekijä, tämä ei estä sitä etteikö se voisi olla eettinen päättelijä. Kuten jo kerroin, muun muassa Drew McDermottin toteaa, että eettinen päätöksenteko edellyttää vapaata tahtoa ja toimijalla on mahdollisuus olla aito eettinen päätöksentekijä vain, jos sillä olisi tarpeeksi vapaata tahtoa toimia myös vastoin moraalista vaatimusta²⁷⁰.

²⁶⁸ Collins 1982, 216.

²⁶⁹ Unlike faiths that posit some uniqueness to the human form that would make artificial minds impossible, Buddhists are more open to the possibility of consciousness instantiated in machines. Hughes 2012, 69, käännös oma.

²⁷⁰ McDermott 2008, 2, 7.

McDermottin toteamus on kiinnostava ja sen tiimoilta on mielekästä jatkaa pohtimaan kysymystä tahdonvapaudesta. Voiko vapaa tahto ilmetä kehollisessa keinoälyssä. Kysymystä hankaloittaa jälleen käsitteen määrittely. Miten vapaan tahdon käsite määritellään? Kuten todettu, Immanuel Kantin mukaan vapaus on järkevien olentojen tahdon ominaisuus. Kirkkoisä Augustinuksen mukaan todellinen vapaus taas perustuu kykenemättömyyteen tehdä syntiä. Näin ollen, ne, jotka nauttivat enkelien ja siunattujen vapaudesta taivaassa, eivät voi toimia vastoin moraalista vaatimusta. Kuten jo todettu, myös Descartes'in mukaan älylliset ja moraaliset virheet ovat tulosta vapaan tahdon väärinkäytöstä.

Näin vaikuttaa siis siltä, että vapaa tahto käsitteenä on jotain erilaista ihmisten vapaasta tahdosta, joka mahdollistaa myös vastoin moraaliperiaatetta toimimisen. Ihmisillä on siis niin kutsuttu säätelevä hallinta (regulative control) toimintaansa.²⁷¹ Olisiko mielekästä varustaa robotti ihmisen vapaalla tahdolla. Viitaten Gilbertiin ensimmäisellä ihmisellä, Aatamilla, oli nämä molemmat kyvyt. Hänellä oli ohjauksen hallinta samoin kuin täysi säätelevä hallintakin.²⁷² Ensimmäinen ihminen siis pystyi halutessaan tekemään vastoin Jumalan tahtoa. Näin hän toimikin, ja – Smithiä jälleen lainatakseni – söi hyvän ja pahan tiedon puusta, oppi hyvän ja pahan eron ja tuli näin ihmiseksi. Voitaneen näin pohtia, että mikäli koneen tahdottaisiin ymmärtävän hyvän ja pahan ero ihmisen lailla, olisi koneeseen ohjelmoitava perisynti. Ajatus on kiinnostava, joskaan ei kiehtova. Lisäksi voitaneen pohtia, että jos ihmisten vapaa tahto on perisyntin turmelema, miksi tätä virhettä tulisi monistaa. Wallach ja Allen eivät liene olleen täysin väärässä todetessaan, ettei konemoraalin kehityksen päämäärä monistaa jokaista ihmisen piirrettä, vaan kehittää piirteet, jotka mahdollistavat eettisen toiminnan²⁷³. Ehkä roboteille tulisi siis rakentaa enkelien ja siunattujen vapaus taivaassa, kykenemättömyys tehdä syntiä?

Lienee kuitenkin vielä mielekästä pohtia vapaan tahdon kysymystä theravāda-buddhalaisuuden näkökulmasta. Läntisten ajattelijoiden, esimerkiksi kirkkoisä Augustinuksen ja Immanuel Kantin, mielestä ihminen ei voinut olla täysin rationaalinen tai täysin vapaa syntiinlankeemuksen vuoksi. Buddhalaisesta näkökulmasta jälleen toimija pystyi täydellistymään elämän kierron aikana ensin tiedostamalla että saṃsāra on ongelma ja nirvāṇa (nibbāna) on vastaus, toisin

²⁷¹ Gilbert 2005, 204.

²⁷² Gilbert 2005, 206.

²⁷³ Wallach&Allen 2009, 178.

sanoen Neljä Jaloa Totuutta, ja tämän jälkeen seuraamalla Jaloa Kahdeksanosaista polkua. Valaistuminen saavutettiin täydellistymällä moraalissa ja viisaudessa, ja yksilö oli kykeneväinen tähän itse. Buddhalaisessa ajattelussa ”perisyntin” paikalla on yksilön oma mielihalu, joka tuottaa edelleen kärsimystä. Vapautumalla tästä mielihalusta voi vapautua myös kärsimyksestä ja saavuttaa valaistumisen. Näin ollen voitaneekin verrata ihmisten vapaata tahtoa buddhalaiseen valaistumattoman tahtoon ja enkelien ja siunattujen vapaata tahtoa taivaassa buddhalaisen valaistuneen vapaaseen tahtoon.

On silti kysyttävä, jos tietoisuus ei ilmene robotissa, miten tällaisen valaistuneen vapaan tahdon implementointi robottiin olisi mahdollista, kun vapaan tahdon (toiminnan) edellytyksenä on buddhalaisen opin mukaan omien tilojen tiedostettu havainnointi.

Lisäksi kysyisin, jos tietoisuus ilmenee robotissa, olisiko tämä robottiin asettuminen kuitenkin lopulta rangaistus? Valaistuneellahan erotetaan kaksi olemisen tasoa. Collinsin mukaan näistä ensimmäinen on nirvāṇa elämässä, jolloin valaistuneella on nimi ja muoto, mutta kärsimys ja mielihalu ovat kaikonneet. Lopullinen nirvāṇa saavutetaan vasta elämän kaikotessa.²⁷⁴ On kuitenkin hyvä huomioida, että robotti ei määritelmällisesti ole kuolevainen. Lisäksi kiinnittäisin huomiota, että vaikka tietoisuus olisi valaistunut, se on elämässä yhä edeltävän karmansa vaikutuksen alainen. Onnettomuus voi kohdata myös valaistunutta edeltävien pahojen tekojen vuoksi. Voisi miltei arvella robottiin asettuneen tietoisuuden antavan uuden määritelmän opille kiirastulesta. On pohdittava, pystyisikö robottiin asettunut tietoisuus milloinkaan saavuttamaan vapautumista elämän kierrosta, joka kuitenkin on buddhalaisen ajattelun päämäärä.

²⁷⁴ Collins 1982, 206–207.

Lähde- ja kirjallisuusluettelo

Lähteet

- Collins, Steven
1982 Selfless Persons. Imagery and thought in Theravāda Buddhism. Cambridge: Cambridge university press.
- Honderich, Ted
2014 Actual Consciousness. Oxford: Oxford University Press.
- Federman, Asaf
2010 "What Kind of Free Will did the Buddha Teach?" Philosophy East & West 60: 1, 1-19. Academic Search Complete, EBSCOhost.
<http://search.ebscohost.com.libproxy.helsinki.fi/login.aspx?direct=true&db=a9h&AN=46794090&site=ehost-live&scope=site>.
Viitattu 20.11.2015.
- Gethin, Rupert
1998 The Foundations of Buddhism. Oxford; New York: Oxford University Press.
- Gier, Nicholas F. & Kjellberg, Paul
2004 Buddhism and the Freedom of Will: Pali and Mahayanist Responses. Freedom and Determinism. Cambridge: A Bradford Book.
<http://search.ebscohost.com.libproxy.helsinki.fi/login.aspx?direct=true&db=nlebk&AN=122498&site=ehost-live&scope=site>.
Viitattu 20.11.2015.
- Gilbert, Christopher
2005 Grades of Freedom: Augustine and Descartes. Pacific Philosophical Quarterly 86: 2, 201-224. Academic Search Complete, EBSCOhost.
<http://onlinelibrary.wiley.com.libproxy.helsinki.fi/doi/10.1111/j.1468-0114.2005.00223.x/epdf>. Viitattu 5.3.2016.
- Keown, Damien
1998 The Nature of Buddhist Ethics. London: Macmillan.
- Wallach, Wendell & Allen, Colin
2009 Moral Machines: Teaching Robots Right from Wrong. New York: Oxford University Press.
- Wallace, B. Alan,
2011 A Buddhist View of Free Will: Beyond Determinism and Indeterminism. Journal of Consciousness Studies 18: 3-4, 217-233.
<http://www.alanwallace.org/buddhistviewoffreewill.pdf>
- Wallace, B. Alan
2007 Contemplative Science: Where Buddhism and Neuroscience Converge. The Columbia Series in Science and Religion. New York: Columbia University Press. Viitattu 20.11.2015.

Kirjallisuus

Abney, Keith

- 2012 Robotics, Ethical theory, and Metaethics: A Guide for the Perplexed. Robot Ethics. The Ethical and Social Implications of Robotics. Cambridge, London: The MIT Press.

Black, Max

- 1952 Critical Thinking: An Introduction to Logic and Scientific Method. Englewood Cliffs: Prentice-Hall.

Churchland, Paul M.

- 1993 A Neurocomputational Perspective: The Nature of Mind and the Structure of Science. 3 ed. Cambridge, London: The MIT Press.
http://web.b.ebscohost.com.libproxy.helsinki.fi/ehost/ebookviewer/ebook/bmxlYmtfXzE3NzdfX0FO0?sid=30574258-8899-4be6-831f-9262c6900428@sessionmgr106&vid=0&format=EB&lpid=lp_COVER-0&rid=0. Viitattu 27.7.2016.

De Silva, Padmasiri.

- 2005 An Introduction to Buddhist Psychology, 4th ed. [Online]
<http://www.palgraveconnect.com/pc/doi/10.1057/9780230509450>. Viitattu 17.9.2015.

Gips, James

- 1991 Towards the Ethical Robot. The Second International Workshop on and Machine Cognition: Android Epistemology Perdido Key, Florida.
<http://www.cs.bc.edu/~gips/EthicalRobot.pdf>. Viitattu 27.7.2016.

Goff, Jared, Hellmann, J.A. Wayne & Hammond, Jay M.

- 2013 A Companion to Bonaventure. Boston: Brill, 2013. eBook Collection (EBSCOhost).
http://web.b.ebscohost.com.libproxy.helsinki.fi/ehost/ebookviewer/ebook/bmxlYmtfXzY3MjM2N19fQU41?sid=808396ec-18a6-45c6-8762-a5fbb47f629d@sessionmgr106&vid=0&format=EB&lpid=lp_289&rid=0. Viitattu 15.8.2016.

Gunkel, David

- 2012 The Machine Question. Cambridge: MIT Press.

Haikonen, Pentti O.

- 2007 Robot Brains : Circuits and Systems for Conscious Machines. Chichester: Wiley-Interscience. eBook Collection (EBSCOhost), EBSCOhost.
<http://web.a.ebscohost.com.libproxy.helsinki.fi/ehost/ebookviewer/ebook/bmxlYmtfXzIwODM3NF9fQU41?sid=bea62382-ee2d-4378-8d7a-30d59ab2d254@sessionmgr4010&vid=0&format=EB&rid=1>. Viitattu 28.7.2016.

- Himma, Kenneth E.
 2009 Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent? *Ethics and Information Technology*, Volume 11(1): 19–29 doi:10.1007/s10676-008-9167-5.
<http://link.springer.com.libproxy.helsinki.fi/article/10.1007/s10676-008-9167-5>. Viitattu 29.7.2016.
- Hughes, James
 2012 *Compassionate AI and Selfless Robots: A Buddhist Approach. Robot Ethics. The Ethical and Social Implications of Robotics*. Cambridge, London: The MIT Press.
- Hume, David
 1739-40 *A treatise of Human Nature*. <http://www.davidhume.org/texts/thn.html>. Viitattu 27.7.2016.
- James, William
 2008 *The Works of William James. Electronic Edition. Volume 11: Essays in Religion and Morality*. Charlottesville, Virginia, USA: InteLex Corporation.
<http://pm.nlx.com.libproxy.helsinki.fi/xtf/view?docId=jamesw/jamesw.1.1.xml;chunk.id=william.james.v11.d2000;toc.depth=1;toc.id=william.james.v11.d2000;brand=default>. Viitattu 31.12.2015.
- Keown, Damien
 2004 *A Dictionary of Buddhism*. Oxford ; New York: Oxford University Press.
<http://www.oxfordreference.com.libproxy.helsinki.fi/view/10.1093/acref/9780198605607.001.0001/acref-9780198605607>. Viitattu 9.2.2016.
- Koch, Christof & Tononi, Giulio
 2008 Can Machines Be Conscious? *IEEE Spectrum* 45: 6, 55-59. Academic Search Complete, EBSCOhost.
<http://spectrum.ieee.org.libproxy.helsinki.fi/biomedical/imaging/can-machines-be-conscious>. (<http://www.klab.caltech.edu/news/koch-tononi-08.pdf>) Viitattu 5.3.2016.
- Lenz, Martin.
 2008 *Ashgate Studies in Medieval Philosophy: Angels in Medieval Philosophical Inquiry: Their Function and Significance*. Abingdon: Routledge. ProQuest ebrary. Web.
<http://site.ebrary.com.libproxy.helsinki.fi/lib/helsinki/reader.action?docID=10226947&ppg=144#>. Viitattu 2.8.2016.
- McDermott, Drew V.
 2008 Why Ethics is a High Hurdle for AI. Presented at North American Conference on Computers and Philosophy (NA-CAP) Bloomington, Indiana, July, 2008.
<http://www.cs.yale.edu/homes/dvm/papers/ethical-machine.pdf>. Viitattu 1.8.2016.

- Nanamoli, Bhikkhu & Bodhi, Bhikkhu
 1995 Teachings of the Buddha: Middle Length Discourses of the Buddha: A Translation of the Majjhima Nikaya. Somerville: Wisdom Publications. ProQuest ebrary. Web.
<http://site.ebrary.com.libproxy.helsinki.fi/lib/helsinki/reader.action?docID=10408024>. Viitattu 9.2.2016.
- Derris, Karen.
 2004 Dhyāna (Trance State). Encyclopedia of Buddhism. New York: Macmillan. *Gale Virtual Reference Library*.
http://go.galegroup.com.libproxy.helsinki.fi/ps/retrieve.do?inPS=true&prodId=GVRL&userGroupName=uhelsink&resultListType=RELATED_DOCUMENT&contentSegment=9780028659107&isBOBIndex=true&docId=GALE|CX3402600132#226. Viitattu 17.9.2015.
- O'Meara, Richard M.
 2012 Contemporary Governance Architecture Regarding Robotics Technologies: An Assessment. Robot Ethics. The Ethical and Social Implications of Robotics. Cambridge: The MIT Press.
- O'Neill, Onora
 1989 Constructions of reason: explorations of Kant's practical philosophy. Cambridge: Cambridge University Press.
- Platon
 - Phaedo. South Bend, IN, USA: Infomotions, Inc., 2001. ProQuest ebrary.
<http://site.ebrary.com.libproxy.helsinki.fi/lib/helsinki/reader.action?docID=5000813>. Viitattu 8.3.2016
- Päivänsalo, Ville
 2007 Balancing Reasonable Justice: John Rawls and Crucial Steps Beyond. Aldershot: Ashgate.
- Remes, Pauliina
 2004 Riippumattomuus ja kuinka se saavutetaan. Stoalaisuus. Tiedon, tunteiden ja hyvän elämän filosofia. Helsinki: Gaudeamus.
- Rennison, Nick.
 2001 Sigmund Freud. Harpenden, GBR: Pocket Essentials. ProQuest ebrary. Web.
<http://site.ebrary.com.libproxy.helsinki.fi/lib/helsinki/reader.action?docID=10005854&ppg=1> Viitattu 31.12.2015.
- Saarinen, Risto.
 2011 Weakness of will in Renaissance and Reformation thought. Oxford: Oxford University Press.

Searle, John R.

- 2014 Introduction: Addressing the hard problem. *Journal of Integrative Neuroscience* 13: 2, vii-xi.
<http://web.a.ebscohost.com.libproxy.helsinki.fi/ehost/pdfviewer/pdfviewer?sid=a94f6c22-0556-4fc8-81e9-1e53970423d4%40sessionmgr4009&vid=1&hid=4106>.
Viitattu 8.3.2016.

Smith, Steven B.

- 2012 *Political Philosophy*. New Haven: Yale University Press.

Stern, Robert.

- 2013 *Kant, Moral Obligation and The Holy Will. Kant on Practical Justification: Interpretive Essays*. Oxford: Oxford University Press.

Sullins, John P.

- 2006 When Is a Robot a Moral Agent? *International Review of Information Ethics* 6: 12, 23–30.
http://www.i-r-i-e.net/inhalt/006/006_Sullins.pdf. Viitattu 29.7.2016.

Torrance, Steve

- 2008 Ethics and consciousness in artificial agents. *AI & Society* 22:4, 495–521.
DOI:10.1007/s00146-007-0091-8.
<http://link.springer.com.libproxy.helsinki.fi/article/10.1007/s00146-007-0091-8>.
Viitattu 28.7.2016.

Ward, Jamie

- 2015 *The Student's Guide to Cognitive Neuroscience*, 3rd ed. London: Psychology Press. ProQuest ebrary.
<http://site.ebrary.com.libproxy.helsinki.fi/lib/helsinki/reader.action?docID=11021798&ppg=14#>. Viitattu 24.8.2016.

Whitby, Blay

- 2012 *Do You Want a Robot Lover? The Ethics of Caring Technologies. Robot Ethics. The Ethical and Social Implications of Robotics*. Cambridge: The MIT Press.

Internet-lähteet

Cyborg Buddha

- <http://ieet.org/index.php/IEET/cyborgbuddha>. Viitattu 30.10.2015

EC, robotics

- <https://ec.europa.eu/digital-single-market/en/robotics>. Viitattu 8.3.2016

EC, DSM

- <https://ec.europa.eu/digital-single-market/node/77880>. Viitattu 8.3.2016

EC, PPP

- Strateginen tutkimusagenda, Draft 0v42 11/10/2013.
<https://ec.europa.eu/digital-single-market/en/robotics-public-private-partnership-horizon-2020>. Viitattu 8.3.2016.

EC, Robotics 2020

- Robotics 2020, Strategic Research Agenda for Robotics in Europe.
http://www.eu-robotics.net/cms/upload/PDF/SRA2020_0v42b_Printable_.pdf
Viitattu 29.10.2015.

EC, strategy

- <https://ec.europa.eu/digital-single-market/en/the-strategy-dsm>.
Viitattu 8.3.2016

euRobotics AISBL

- <https://eu-robotics.net/about/about-eurobotics-aisbl/>. Viitattu 8.3.2016.

fr.oldict.com

- <http://fr.oldict.com/robot/>. Viitattu 13.8.2016.

Stanford Encyclopedia of Philosophy

- Eliminative Materialism. <http://plato.stanford.edu/entries/materialism-eliminative/>. Viitattu 26.8.2016,

The Buddha and His Teachings

- <http://www.bps.lk/olib/bp/bp102s-files/OEBPS/Text/10Main03.html>.
Viitattu 13.1.2016